Edinburgh Research Explorer

Structural brain imaging correlates of general intelligence in UK Biobank
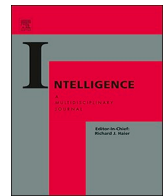
Structural brain imaging correlates of general intelligence in UK Biobank

Cox S.R.[a,b,c,*], Ritchie S.J.[d], Fawns-Ritchie C.[a,b], Tucker-Drob E.M.[e], Deary I.J.[a,b]

[a] *Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, UK*
[b] *Department of Psychology, The University of Edinburgh, UK*
[c] *Scottish Imaging Network, A Platform for Scientific Excellence (SINAPSE) Collaboration, Edinburgh, UK*
[d] *Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK*
[e] *Department of Psychology, University of Texas, Austin, TX, USA*

ARTICLE INFO

ABSTRACT

The associations between indices of brain structure and measured intelligence are unclear. This is partly because the evidence to-date comes from mostly small and heterogeneous studies. Here, we report brain structure-intelligence associations on a large sample from the UK Biobank study. The overall $N = 29,004$, with $N = 18,426$ participants providing both brain MRI and at least one cognitive test, and a complete four-test battery with MRI data available in a minimum $N = 7201$, depending upon the MRI measure. Participants' age range was 44–81 years (M = 63.13, SD = 7.48). A general factor of intelligence (g) was derived from four varied cognitive tests, accounting for one third of the variance in the cognitive test scores. The association between (age- and sex-corrected) total brain volume and a latent factor of general intelligence is $r = 0.276$, 95% C.I. = [0.252, 0.300]. A model that incorporated multiple global measures of grey and white matter macro- and microstructure accounted for more than double the g variance in older participants compared to those in middle-age (13.6% and 5.4%, respectively). There were no sex differences in the magnitude of associations between g and total brain volume or other global aspects of brain structure. The largest brain regional correlates of g were volumes of the insula, frontal, anterior/superior and medial temporal, posterior and paracingulate, lateral occipital cortices, thalamic volume, and the white matter microstructure of thalamic and association fibres, and of the forceps minor. Many of these regions exhibited unique contributions to intelligence, and showed highly stable out of sample prediction.

1. Introduction

The association between brain volume and intelligence has been one of the most regularly-studied—though still controversial—questions in cognitive neuroscience research. The conclusion of multiple previous meta-analyses is that the relation between these two quantities is positive and highly replicable, though modest (Gignac & Bates, 2017; McDaniel, 2005; Pietschnig, Penke, Wicherts, Zeiler, & Voracek, 2015), yet its magnitude remains the subject of debate. The most recent meta-analysis, which included a total sample size of 8036 participants with measures of both brain volume and intelligence, estimated the correlation at $r = 0.24$ (Pietschnig et al., 2015). A more recent re-analysis of the meta-analytic data, only including healthy adult samples ($N = 1758$), found a correlation of $r = 0.31$ (Gignac & Bates, 2017). Furthermore, the correlation increased as a function of intelligence measurement quality: studies with better-quality intelligence tests—for instance, those including multiple measures and a longer testing time—tended to produce even higher correlations with brain volume (up to 0.39). In a meta-analysis, issues of cross-cohort heterogeneity might have an important bearing on the magnitude of the correlation.

Here, we report an analysis of data from a large, single sample with high-quality MRI measurements and four diverse cognitive tests. We use latent variable modelling to create a general intelligence ('g') factor from the cognitive test and estimate its association with both total brain volume and several more fine-grain imaging-derived indices of brain structure. We judge that the large N, study homogeneity, and diversity of cognitive tests relative to previous large scale analyses provides important new evidence on the size of the brain structure-intelligence correlation. By investigating the relations between general intelligence and characteristics of many specific regions and subregions of the brain in this large single sample, we substantially exceed the scope of previous meta-analytic work in this area.

There is considerable debate about what the association between brain size and general intelligence means. It is unclear, for example, whether brain size is a direct proxy for neuron number (discussed in Pietschnig et al., 2015). There is also an apparent paradox that there are substantial sex differences in total brain volume (on the order of 1.41 standard deviations; Ritchie et al., 2018) but litte-to-no sex differences in mean intelligence (Deary, Irwing, Der, & Bates, 2007; Johnson, Carothers, & Deary, 2008; Lakin & Gambrell, 2014; Ritchie et al., 2018). More recent work indicates that multiple brain properties might be required to better explain individual differences in general intelligence, and some of these might be compensatory for differences in overall brain size (Deary, Penke, & Johnson, 2010; Kievit et al., 2012; Kievit et al., 2014; Luders et al., 2004; Ritchie et al., 2015). For example, having a more sparsely- and better-organised dendritic arbor (as measured by cortical neurite density and orientation dispersion) may predict higher intelligence beyond simple measures of grey matter volume (Genç et al., 2018). Furthermore, in an older cohort with a narrow age range ($N = 672$), Ritchie et al. (2015) found that incorporating multiple global, but tissue-specific, brain MRI measures (including tissue volumes, measures of white matter microstructure, and hallmarks of brain ageing) accounted for up to 21% of the variance in general intelligence, which was substantially higher than could be accounted for by total brain size alone (~12%).

Such results, combined with indications that there is regional heterogeneity in the magnitude of intelligence associations across both grey and white matter (Jung & Haier, 2007; Deary et al., 2010; Basten, Hilger, & Fiebach, 2015; Karama et al., 2011; Cox et al., 2018; Ryman et al., 2016), extend the focus beyond a single, well-replicated proxy (total brain volume) and toward tissue- (and region-) specific associations with general intelligence. One of the most influential accounts of the neurobiological underpinnings of general intelligence (also known as general cognitive ability, or "g") has been the Parieto-Frontal Integration Theory (P-FIT; Jung & Haier, 2007). The P-FIT was initially based on a synthesis of disparate structural and functional brain imaging results. However, none of the Brodmann regions implicated in intelligence were supported by >60% of the studies reviewed, which the authors pointed out might be considered a relatively weak consensus.

The P-FIT model implicates the following regions as being associated with intelligence differences: the lateral frontal, superior temporal, medial temporal, parietal and extrastriate (lateral occipital) regions, along with the white matter tracts that connect them. Specific reference was originally made to the arcuate fasciculus; this pathway is variously described as being just adjacent to the superior longitudinal fasciculus, or as one of the components thereof (Dick & Tremblay, 2012; Kamali, Flanders, Brody, Hunter, & Hasan, 2014). Together, these form a 'dorsal stream' of anterior-posterior cortical connectivity. Alongside other fibres such as the inferior longitudinal, inferior fronto-occipital, uncinate, and cingulum fasciculi, these 'association' fibres—along with the genu of the corpus callosum (forceps minor)—facilitate connectivity across the distal cortical regions highlighted by the P-FIT model. The model has generally received support from subsequent work (Deary et al., 2010; Basten et al., 2015; Karama et al., 2011; Cox et al., 2018; Ryman et al., 2016).

As with the meta-analyses on brain volume and intelligence described above (Gignac & Bates, 2017; Pietschnig et al., 2015), the broad heterogeneity of studies on the P-FIT might produce a less precise picture of the brain basis of cognitive abilities. Further evaluation of the model would greatly benefit from large-sample research that investigates the grey and white matter components of this putative intelligence framework, together in the same analysis. We conduct that analysis in the present study.

We capitalize on data from the UK Biobank study, a large-scale biomedical study of health and wellbeing, which includes brain MRI and various measures of cognitive ability. The UK Biobank participants have completed various cognitive measures; originally, they were administered a battery of bespoke tests with relatively poor reliability (Lyall et al., 2016). Using an earlier data release (Ritchie et al., 2018), we previously estimated the correlation between brain size and one of those tests, "Fluid Intelligence" (which we refer to as Verbal-Numerical Reasoning) to be $r = 0.177$. We found that the correlation did not differ by sex. Another study using an earlier release of UK Biobank imaging data examined the association between Verbal-Numerical Reasoning and brain size, reporting a correlation of $r = 0.19$ ($N = 13,608$; Nave, Jung, Linnér, Kable, & Koellinger, 2019). In addition, analyses of regional white and 10 grey matter measures have been reported with respect to Verbal-Numerical Reasoning in an earlier UK Biobank release; however, the authors of that study cited several reasons to doubt that this test, in isolation, is a valid indicator of fluid cognitive ability (Kievit, Fuhrmann, Borgeest, Simpson-Kent, & Henson, 2018; see also Hagenaars et al., 2016).

In this pre-registered study, we use a newer subset of UK Biobank participants who have completed an enhanced cognitive assessment battery at their brain imaging assessment. The overlap of the complete cognitive battery and the various MRI measures ranges from $N = 8165$ to 7318 following exclusions that are described below. Their data have only recently been released, and have not previously been analysed by our team. The enhanced cognitive battery includes three new measures based on standardised cognitive tests: Symbol-Digit Substitution, Matrix Reasoning, and Trail-Making. These three tests, combined with the previous Verbal-Numerical Reasoning measure, allows the estimation of brain imaging associations with a latent factor of general intelligence (g), that arguably gives coverage of the cognitive domains of reasoning, processing speed, working memory, and executive function. In a large sample size, the current study design thus: results in a better-quality cognitive measure than was previously possible in the UK Biobank data; mitigates variability in the administration and measurement of cognitive and brain imaging constructs (potentially allowing for stronger brain-intelligence correlations; Gignac & Bates, 2017); and, given the detailed brain imaging measures available, facilitates a detailed estimate of the global and regional brain correlates of latent general intelligence.

Our analyses followed a preregistered protocol and 4 hypotheses (https://osf.io/w7evd/). First, we tested whether the four cognitive tests were correlated moderately-highly ($r > 0.40$), and formed a latent general factor that explains 40% or higher of the variance across tests. Next, we examined the association between this cognitive factor and total brain volume, and we hypothesised that there would be no significant sex difference in the size of the brain-cognitive correlation for any of the models. We then hypothesised that different global measures of grey and white matter would each account for significant unique variance in g. We then aimed to test associations between general intelligence and brain grey and white matter regional measures, hypothesising that the strongest associations would concur with regions implicated by the Parieto-Frontal Integration Theory of intelligence (P-FIT; Jung & Haier, 2007).

2. Methods

The UK Biobank study is a large-scale biomedical study of the determinants of the diseases of middle and older age, which includes brain MRI and measures of cognitive function (Sudlow et al., 2015). Cognitive tests and brain imaging data were acquired on the same assessment day. The tests used here were administered at the UK Biobank brain imaging assessment. The imaging assessment took place at 3 different assessment centres. The majority were in Manchester, with more recent appointments now also taking place in Newcastle, and most recently in Reading. Cognitive tests were administered to participants working independently on a touchscreen computer with no tester observing. MRI data was acquired using the same hardware and software. The current data release from UK Biobank initially included 30,316 participants who attended the scanning appointment, i.e. they had a record

Table 1
Participant characteristics split out by assessment centre, and for the full sample.

	Manchester		Reading		Newcastle		Full Sample	
	M(SD)[a]	N	M(SD)	N	M(SD)	N	M(SD)	N
Age (years)	62.818 (7.484)	22,037	65.202 (7.299)	866	63.962 (7.379)	6101	63.130 (7.480)	29,004
Sex (F:M)	11,315: 10,722	22,037	453: 413	866	3256: 2845	6101	15,024: 13,980	29,004
Matrix Reasoning	8.047 (2.113)	8994	8.277 (2.045)	850	7.904 (2.124)	5671	8.007 (2.115)	15,515
Symbol-Digit	19.212 (5.255)	9019	18.753 (5.152)	858	18.714 (5.235)	5681	19.005 (5.247)	15,558
VNR	6.778 (2.070)	20,455	6.800 (2.058)	860	6.466 (2.001)	5668	6.713 (2.061)	26,983
TMTb[a]	496 (236)	9103	512 (230)	866	503(236)	5710	499 (235)	15,679
TBV	1,170,062 (110900)	17,223	–	0	1,154,425 (110376)	3104	1,167,674 (110960)	20,327
GM	618,392 (55474)	17,226	–	0	611,240 (54923)	3104	617,300 (55448)	20,330
NAWM	547,110 (61736)	16,146	–	0	537,269 (61090)	3062	545,541 (61737)	19,208
WMH [a]	2500 (3641)	16,146	–	0	3290(4506)	3062	2622 (3816)	19,208
gFA	0.025 (0.824)	15,448	–	0	−0.131 (0.834)	2989	0.00 (0.827)	18,437
gMD	0.022 (0.918)	15,448	–	0	−0.113 (0.907)	2989	0.00 (0.918)	18,437

Note. Means and standard deviations (SD) reported, except for [a]median and interquartile ranges are given. VNR: verbal numerical reasoning, TMTb: Trail Making Test Part b, TBV: total brain volume, GM: grey matter volume, NAWM: normal-appearing white matter volume, WMH: white matter hyperintensity volume, gFA: general factor of white matter fractional anisotropy, gMD: general factor of white matter mean diffusivity.
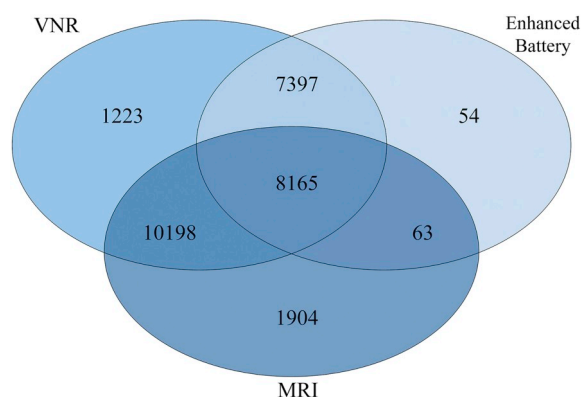

Fig. 1. Indicative overlap between initial cognitive measures in the imaging visit (VNR; verbal numerical reasoning), MRI measures, and the Enhanced Cognitive Battery (Matrix Reasoning, Symbol-Digit and Trail-Making Part B) among the 29,004 participants included in the current analysis. For ease of illustration, the MRI numbers are based on grey matter volume (highest N among global MRI measures), and the Enhanced Battery numbers are based upon Trail Making Part B (highest N among the Enhanced Battery). A total of 18,426 have MRI and at least one cognitive test. There are slight variations in missingness among Enhanced Battery and MRI measures (see Table 1).

for age at scanning. Following exclusions, the total N = 29,004. The minimal N with complete cognitive-MRI overlap was N = 7318; further information is provided in Statistical Analysis, Table 1 and Figs. 1 and S1. UK Biobank Field IDs are listed in Table S1. Most of them had scores for the Verbal-Numerical Reasoning test (always administered at the MRI appointment). Many fewer had the more recently-introduced cognitive tests (Matrix Pattern Completion, Symbol-Digit and Trail Making Test Part B). Missing MRI data was due to the lag between MRI acquisition and its subsequent processing for release by the UK Biobank. The UK Biobank study was conducted under generic approval from the NHS National Research Ethics Service (approval letter dated 17th June 2011, ref. 11/NW/0382), under approved project 10279.All data and materials are available via UK Biobank (http://www.ukbiobank.ac.uk).

2.1. Cognitive tests

The four cognitive tests used in the current study were: Matrix Reasoning, Symbol-Digit Substitution, Verbal-Numerical Reasoning, and Trail-Making Test. Specifically, for the Trail-Making Test, we used part B, since this test includes both elements of speed and executive functioning (Salthouse, 2011).

2.1.1. Matrix pattern completion

The UK Biobank Matrix Reasoning test is an adapted version of the Matrices test in the COGNITO battery (Ritchie et al., 2014). This test of non-verbal fluid reasoning requires participants to inspect a grid pattern with a piece missing in the lower right-hand corner and select which of the multiple choice options at the bottom of the screen completes the pattern both horizontally and vertically. This 15-item test assesses participants' ability to problem solve using novel and abstract materials. The score is the number of correctly answered questions in three minutes.

2.1.2. Symbol-digit substitution

Symbol-Digit Substitution was used as a measure of processing speed. It is similar in format to the Symbol Digit Modalities Test (Smith, 1991), which is a well-validated measure of processing speed. At the top of the screen, participants were shown a key pairing shapes with numbers. Beneath the key were rows of shapes with an empty box under each shape. Using the key, participants had 60 s to enter the number in the empty boxes that are paired with the shapes. Participants were instructed to work as quickly and as accurately as possible. The score is the number of correct symbol-digit matches made in 60 s.

2.1.3. Verbal-numerical reasoning (VNR)

Referred to as 'Fluid Intelligence' in UK Biobank, this test of verbal and numerical reasoning required participants to answer 13 multiple-choice questions assessing verbal (e.g., "Bud is to flower as child is to?" Possible answers: Grow/Develop/Improve/Adult/Old) and numerical (e.g. "150…137…125…114…104… What comes next?" Possible answers: 96/95/94/93/92) abilities. Each question appeared at the top of computer screen, and 3–5 possible answers were provided underneath. Participants were to select which of the answers they thought was correct, or select "Do not know", or "Prefer not to answer". The score is the number of questions answered correctly in 2 min.

2.1.4. Trail-making test part B

This test is a computerised version of the Halstead-Reitan Trail-Making Test (Reitan & Wolfson, 1985). It is often said to be an assessment of executive function. Though not considered a classical test of intelligence, Trails B performance is genetically and phenotypically strongly related to general intelligence as well as the cognitive domain of processing speed (Hagenaars, Cox, Hill, Davies, & Liewald, 2018; MacPherson et al., 2017; MacPherson, Allerhand, Cox, & Deary, 2019; Salthouse, 2011). In part B, participants were presented with the numbers 1–13, and the letters A-L arranged quasi-randomly on a computer screen. The participants were instructed to switch between

touching the numbers in sequential order, and the letters in alphabetical order (e.g., 1-A-2-B-3-C) as quickly as possible. In this computerised version, when the participant touches the correct number or letter, the number or letter is highlighted and a line appears connecting the correct answer to the previously pressed response. Each time the participant pushes an incorrect number or letter, the number or letter flashes red to indicate to the participant a mistake has been made. The participant must select the correct number or letter in the sequence to move on. UK Biobank records both the time taken and the number of errors made to complete part B. In this analysis, we do not examine errors because few participants made errors. The median number of errors was 0. The time taken to complete part B accounts for errors as making errors will increase the amount of time taken to complete the task. The score used here is the time (in deci-seconds) taken to successfully complete the test. Those with a score coded as 0 (denoting "Trail not completed") had their score set to missing.

2.2. Brain imaging acquisition and analysis

All brain MRI data were acquired on a Siemens Skyra 3 T scanner with a standard Siemens 32-channel head coil, in accordance with the open-access protocol (http://www.fmrib.ox.ac.uk/ukbiobank/protocol/V4_23092014.pdf), documentation (http://biobank.ctsu.ox.ac.uk/crystal/docs/brain_mri.pdf), and publication (Alfaro-Almagro et al., 2018). T_1-weighted MPRAGE data was acquired in the sagittal plane at 1 mm isotropic resolution; the T_2-weighted FLAIR acquisition at $1.05 \times 1 \times 1$ mm resolution, was also acquired in the sagittal plane. The diffusion MRI (dMRI) data was acquired using a spin-echo echo-planar sequence with 10 T_2-weighted (b $\approx$ 0 s mm^2) baseline volumes, 50 b = 1000 s mm^{-2} and 50 b = 2000 s mm^{-2} diffusion-weighted volumes, with 100 distinct diffusion-encoding directions and 2 mm isotropic voxels. We used global and regional brain Imaging Derived Phenotypes (IDPs) provided by the UK Biobank brain imaging team: total brain volume (TBV, which is the sum of grey and white matter and excludes cerebrospinal fluid), grey matter volume (GM), and white matter volume (WM) from FSL FAST (Zhang, Brady, & Smith, 2001), 14 subcortical volumes using FSL FIRST (Patenaude, Smith, Kennedy, & Jenkinson, 2011) and white matter hyperintensity volume (WMH) using BIANCA (Griffanti et al., 2016), which uses both T1-weighted and T2-weighted volumes. We estimated normal-appearing white matter volume (NAWM) as the difference between total WM and WMH. Regional brain information was also available as UK Biobank IDPs in the form of tract-averaged fractional anisotropy and mean diffusivity for each of 27 white matter tracts using AutoPtx (de Groot et al., 2013), and as individual grey matter cortical segmentations according to the Harvard-Oxford Atlas (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases). The T1-W volume was non-linearly warped to MNI152 space using FNIRT; the Harvard-Oxford cortical atlas is also defined in MNI152 space, and the previously-estimated warp field is inverted and applied to the ROIs to derive a version of the ROIs in native space for masking on the FAST grey matter segmentation. These white matter tracts and cortical regions are shown in Fig. 2. Head positioning coordinates were derived from the NIFTI imaging data: X and Z head positioning coordinates used the Center of Gravity of the brain mask, which was converted into real-world coordinates using the qform matrix of the image. The Y coordinate was obtained by linearly registering the brain mask to a reference brain mask and taking the smallest Y coordinate.

2.3. Statistical analyses

All individuals who reported any of the neurological or neurodegenerative health conditions listed in Supplementary Material (Table S1) were removed prior to analysis. Outliers (+/-4SDs) were removed from brain and cognitive measures. This was <1% of data for all variables; full numbers for each cognitive and MRI measure are provided as Supplementary Data. All participants were of White European ancestry. Of an initial 30,316 UK Biobank participants who had a record of age at attending the imaging assessment, 29,004 provided data on at least one of the primary variables of interest (global brain imaging or cognitive) following exclusions, and 18,426 had at least one cognitive test and MRI data. All structural equation modelling (SEM) analyses were conducted using Full Information Maximum Likelihood (FIML) estimation within R, using the *lavaan* package for SEM (Rosseel, 2012). FIML takes advantage of all available data, including data from participants who are missing data on some of the dependent variables. We provide information on the number of participants with complete data for illustrative purposes in the Results section. Throughout, unless explicitly stated, indicators were corrected for age and sex, with the MRI variables also corrected for scanner head positioning confounds (see Section 2.2; X, Y and Z coordinates provided by the UK Biobank team: UKB IDs: 25756, 25,757, 25,758). Specifically, we do not correct variables for age when running multi-group models where age is already the grouping term (Sections 2.3.4 and 3.4), and we do not correct variables for sex when sex is the grouping term (Sections 2.3.3 and 3.5). In contrast to our pre-registration, covariates were applied to manifest variables within SEMs (i.e. we did not need to residualise data outside the model to enable model convergence and fit). Model fits were assessed with a chi-squared test, the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA), and the Standardised Root Mean Square Residual (SRMR).

2.3.1. Estimating a latent general factor of general intelligence, 'g'

We performed a confirmatory factor analysis (CFA) of the 4 cognitive tests: Symbol-Digit Substitution, Matrix Reasoning, Trail-Making Test Part B, and Verbal-Numerical Reasoning. We hypothesised that the four tests would correlate moderately-highly (with intercorrelations of $r > 0.40$), and would form a single latent general factor explaining ~40% of the variance across the 4 tests, with good fit to the data (CFI and TLI > 0.95, SRMR and RMSEA < 0.05). We ran a version without, and then with age and sex correction at the manifest level. Since principal components analyses (PCAs) are commonly also used in intelligence research (e.g. Nave et al., 2019), but do not separate common and test-specific variance, we also provide a PCA estimate of *g* (tests not corrected for age and sex) using the first unrotated principal component, for comparison with the CFA.

2.3.2. Associations between general intelligence (g) and global brain MRI measures

Next, we estimated the association of the latent general intelligence factor ('*g*') with total brain volume, and then 6 global measures of grey and white matter: grey matter, normal-appearing white matter and white matter hyperintensity volumes (TBV, GM, NAWM, WMH), and general factors of fractional anisotropy and mean diffusivity (*g*FA and *g*MD). The general factors of white matter microstructure were formed from diffusion indices for white matter pathways of interest, which were extracted from confirmatory factor analysis, as previously described (Cox et al., 2016). We tested each individual brain-*g* association, i.e. we fitted a separate SEM for each brain MRI measure. We then fitted a single SEM in which all indicators contributed to *g* variance; a so-called Multiple Indicators Multiple Causes (MIMIC; Muthén, 1989) model. We excluded TBV from this, to avoid model fit and theoretical part-whole issues. Significant correlated residual paths among the imaging variables estimated from modification indices were included. False-Discovery Rate (FDR; Benjamini & Hochberg, 1995) correction of the *p*-values (implemented using the p.adjust function in R) was applied across the six bivariate associations of interest, and then across each of the path estimates in the multivariate SEM. Manifest variables were corrected as described above.

We then conducted an additional—non-pre-registered—analysis, to investigate whether the substantially lower proportion of *g* variance accounted for by multiple global MRI measures in this sample, when
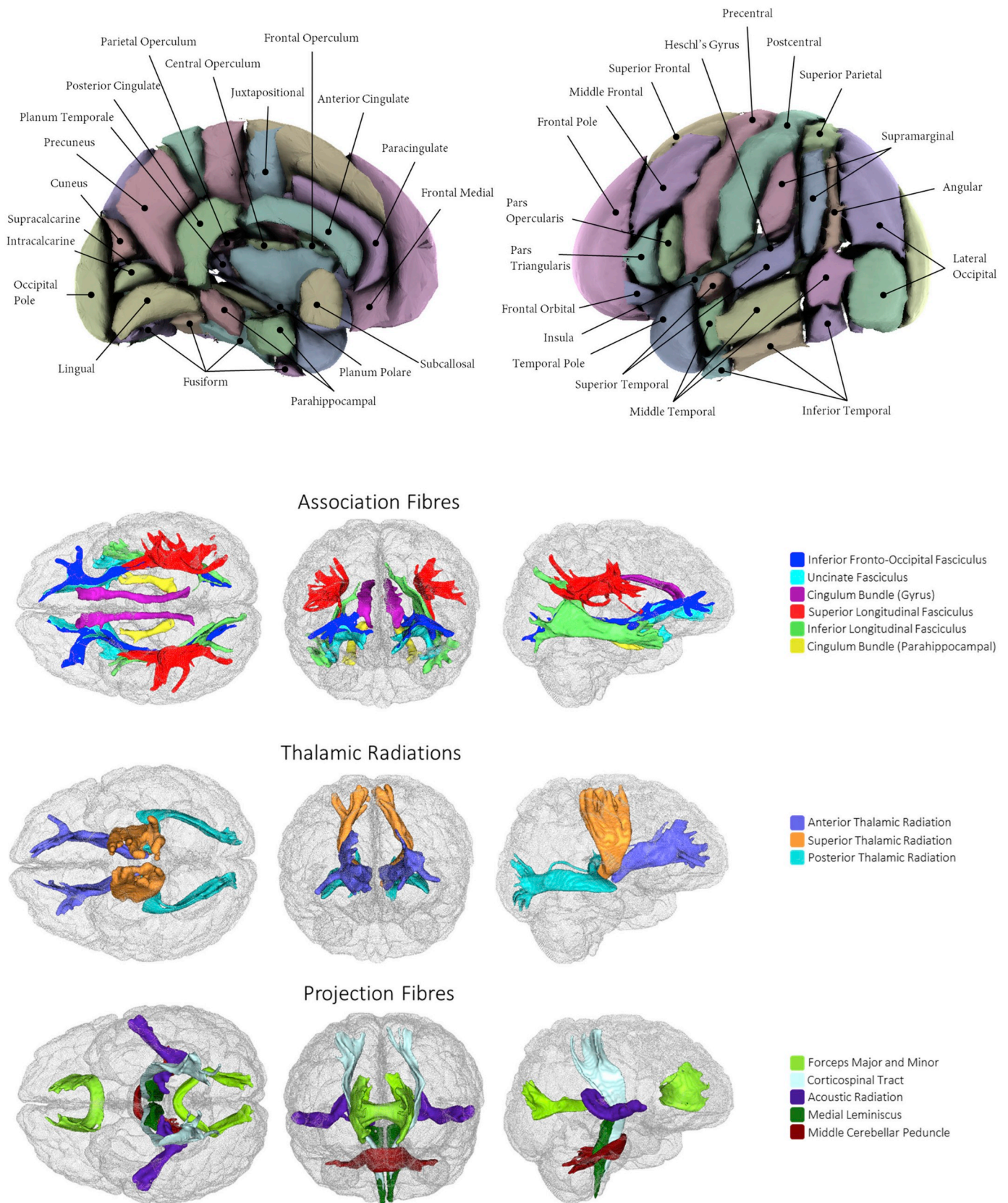
Fig. 2. Brain imaging regions of interest according to the Harvard-Oxford Atlas (**top: cortical regions**) and AutoPtx (**bottom: white matter tracts**; adapted from Cox et al., 2019).

compared to our prior work in an older cohort (Ritchie et al., 2015), was due to moderating effects of age. We split the sample by age into groups with equal-sized cognitive-MRI overlap (middle age $N = 10,164$; older age N = 10,166) to ensure no imbalance in statistical power (above and below age 63.29 years). Initially, we tested for measurement invariance of g between the two age groups. Specifically, we were interested in weak factorial invariance (equal factor loadings) rather than strong (equal loadings and intercepts; as defined by Widaman, Ferrer, & Conger, 2010), given the expected difference in cognitive performance between middle and older participants. We did this by comparing two multi-group SEMs; in the first, the cognitive test loadings on g were freely estimated, whereas in the second they were constrained to be equal in the middle and older-aged groups. We used a chi-squared test, the Akaike Information Criterion (AIC), and the sample-adjusted Bayesian Information Criterion (saBIC), and an additional check of factor congruence (coefficient of factor congruence; Lorenzo-Seva & ten Berge, 2006) using the 'psych' package, to test whether there was a difference between these sub-models. Congruence coefficients index the similarity between factor solutions; a congruence coefficient > 0.90 indicates an extremely high level of similarity of the factor solutions. Next, we used a different set of two multi-group models to test whether the associations between g and the MRI measures differed as a function of age. In the first, the g-brain associations were freely estimated, and in the second, they were constrained to be equal between the two age groups. All measures were corrected for sex, and the MRI measures were also corrected for MRI head position. The group g loadings were set to equality. We ran this test for both a single g-TBV association, and then where multiple global MRI measures (GM, NAWM, WMH, gFA, gMD) predicted g. Differences between the two multi-group SEMs were assessed with a chi-squared test, the AIC, and the saBIC.

2.3.3. Sex differences in g-brain MRI associations

We then investigated sex differences in the size of the brain-cognitive associations. Before doing so, we tested for measurement invariance between the sexes by creating a multi-group SEM including just the cognitive tests, with sex as the grouping variable, and tested for strong measurement invariance (as defined by Widaman et al., 2010). If strong measurement invariance was found (i.e., the model with strong invariance does not fit significantly more poorly, by a chi-squared test, the AIC, and the saBIC, than one where factor loadings and intercepts are freely-estimated), we aimed to test a set of models where the brain-cognitive associations was fixed to equality across the two sub-models grouped by sex, and one where it was freely-estimated. We used a chi-squared test, the AIC, and the saBIC to test whether there was a difference between these sub-models (thus indicating that there is a sex difference). For these analyses, the variables were adjusted for all the above-mentioned covariates except sex (because this was the grouping variable).

2.3.4. Associations between g and regional brain MRI measures

Finally, we examined associations between g and regional brain measures: i) the fractional anisotropy and mean diffusivity in 27 white matter pathways, ii) cortical volumes of 48 regions according to the Harvard-Oxford cortical atlas segmentations, and iii) 14 subcortical volumes (bilateral nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen, thalamus). We applied FDR correction within each family of tests: across all 96 cortical tests, and separately across the 27 tests of WM tracts for FA, and then for MD, and across all 14 subcortical tests. We hypothesised that the associations between general intelligence and brain volumes across the cortex would be consistent with the Parieto-Frontal Integration Theory (P-FIT; Jung & Haier, 2007), and be strongest in lateral frontal, superior parietal and temporal regions. Likewise, we hypothesised that thalamic and association fibres, plus forceps minor will show the statistically largest associations with general intelligence. The additional use of subcortical

volumes was an addition to our pre-registered plan; subcortical structure did not figure largely in the P-FIT (Jung & Haier, 2007), though more recent work has reported associations between intelligence and overall subcortical volume (Ritchie et al., 2015), caudate (Basten et al., 2015; Grazioplene et al., 2015; Rhein et al., 2014), hippocampal (Valdés Hernández et al., 2017) and thalamic volume (Bohlken et al., 2014).

2.3.5. Out-of-sample prediction

Following peer review, we also included a series of analyses in which we tested the ability to predict g from multivariate brain structural parameters. To do this, we treated the larger Manchester sample as a training set, and the Newcastle sample as the test set. We did this separately for cortical volumes, FA in white matter tracts, MD in white matter tracts, and subcortical volumes. Training was performed by initially fitting a MIMIC SEM in which multiple brain MRI measures predicted g (covariates were age and sex for all manifest variables, and also head coordinates for all manifest imaging variables). Modification indices were consulted to identify residual correlations among MRI indicators until model fit was within our pre-registered criteria. We did not apply regularised regression methods here, given recent evidence that this approach does not outperform standard SEM methods in MIMIC models using larger samples (Jacobucci, Brandmaier, & Kievit, 2019). Given the highly similar left/right bivariate associations with g in our foregoing analyses, averages of left and right were computed for white matter, cortical and subcortical regions in the interests of model parsimony. We then created weighted a composite score in the test data (and also in the training data, for comparison), according to the beta weights observed in the training sample, and ran a new MIMIC SEM (fixing g loadings and the residual correlation across samples). Standardised associations between the weighted composite MRI score and g were then compared between training and test samples. Finally, in response to peer review, we conducted a supplementary analysis to ascertain the degree to which the associations between regional volumetric measures (cortical and subcortical volumes) and g were independent of total brain volume. We did this in two different ways: first, by adjusting each ROI in the training MIMIC model for TBV, from which the beta weights were then estimated and applied into the test set, and second by using the beta weights from the initial MIMIC models (i.e. uncorrected for TBV), and then including total brain volume as a covariate in the regression between g and the composite weighted scores.

3. Results

3.1. Estimating a latent general factor of general intelligence, 'g'

Participant characteristics are shown in Table 1. The cognitive tests were all correlated with medium effect sizes according to Cohen (1992): the Pearson's r range was $|0.300$ to $0.405|$. A first principal component (without age and sex correction) accounted for 55% of the variance, with loadings ranging from $|0.71$ to $0.80|$ (Table S2). The confirmatory factor analysis, which was informed by $N = 27,100$ (complete cognitive data on $N = 15,029$), in which each indicator was not corrected for age and sex, had two fit indices (TLI and RMSEA) outside our pre-registered criteria (CFI = 0.973, TLI = 0.918, RMSEA = 0.078, SRMR = 0.024). Modification indices suggested the addition of a residual correlation between Verbal-Numerical Reasoning and Matrix Reasoning ($r = 0.170$), following which model fit was above our pre-registered threshold across all fit indices (CFI = 0.995, TLI = 0.969, RMSEA = 0.048, SRMR = 0.010). The general factor of cognitive ability accounted for 40% of the cognitive test score variance (standardised loadings were Matrix Reasoning = 0.550, Symbol-Digit = 0.626, Verbal-Numerical Reasoning = 0.532, Trail-Making part B = -0.794). When we corrected each cognitive test within the SEM for age and sex, keeping the abovementioned residual correlation, the

Table 2
Associations between *g* and global MRI measures across the whole sample.

Model	Individual			Simultaneous		
	Std. Est.	*SE*	*p*	*Std. Est.*	*SE*	*p*
TBV	0.276	0.019	<0.001	–		–
GM	0.281	0.020	<0.001	0.201	0.019	<0.001
NAWM	0.246	0.013	<0.001	0.102	0.020	<0.001
WMH	−0.106	0.013	<0.001	−0.097	0.013	<0.001
*g*FA	0.090	0.018	<0.001	−0.003	0.018	0.865
*g*MD	−0.066	0.012	<0.001	−0.037	0.019	0.049

Note. Standardised estimates (Std. Est.) and standard errors (SE) reported. TBV: total brain volume, GM: grey matter volume, NAWM: normal-appearing white matter volume, WMH: white matter hyperintensity volume, *g*FA: general factor of white matter fractional anisotropy, *g*MD: general factor of white matter mean diffusivity. Manifest variables are corrected for age and sex; brain measures also corrected for head positioning confounds.

model fit the data well (CFI = 0.998, TLI = 0.978, RMSEA = 0.030, SRMR = 0.004). The general factor of cognitive ability accounted for 32% of the cognitive test score variance (standardised loadings were Matrix Reasoning = 0.505, Symbol-Digit = 0.479, Verbal-Numerical Reasoning = 0.592, Trail-Making part B = −0.666).

3.2. Associations between general intelligence (g) and brain MRI measures

Results between *g* and global brain MRI measures are shown in Table 2 and Fig. 3. SEM fit statistics are reported in Table S3, and residual correlations among the global brain tissue measures from the MIMIC model are reported in Table S4. In all models, the cognitive and MRI indicators were adjusted for age and sex, and the MRI indicators also adjusted for head positioning confounds. The latent factor of general intelligence was associated with TBV at $\beta = 0.276$, $p < .001$. This model estimate was informed by N = 18,426 (complete data $N = 8092$). Associations with GM ($\beta = 0.281$) and NAWM ($\beta = 0.246$) were significantly larger than the other three tissue-specific measures (i.e., WMH, *g*FA, *g*MD; *p* for comparisons <0.001). The associations of *g* with white matter hyperintensities, *g*FA, and *g*MD were $\beta = −0.106$, 0.090 and − 0.066, respectively.

In a multivariate SEM (MIMIC model; complete cases = 7494), we found that all MRI measures (GM, NAWM, WMH, *g*FA, *g*MD) accounted

for 6.16% of the variance in *g*. The unique contributions to this variance were largest for grey matter volume ($\beta = 0.201$) with equivalent contributions of NAWM and WMH volumes ($\beta = 0.102$ and $\beta = −0.097$, respectively). Neither measure of white matter microstructure made significant unique contributions beyond this, following FDR correction (*g*FA $\beta = −0.003$, $p = .865$; and *g*MD $\beta = −0.037$, $p = .049$).

Ritchie et al. (2015) previously reported that the total variance in *g* explained by multiple MRI markers in an older cohort (all aged approximately 73 years) was 18–21%. To determine whether the difference between that estimate and the one reported here may be attributable to differences in the age range of the samples, we conducted a post-hoc supplementary test of differences in the proportion of variance explained by age. Initially, we tested whether *g* exhibited weak measurement invariance across the two age groups. Both models had excellent fit to the data, and were highly similar: saBIC indicated that weak invariance was preferred, contradicting the AIC results and the small but significant difference detected by the chi-squared test ($\Delta\chi^2(3) = 27.617$, $p \leq .001$, ΔAIC = 22, ΔsaBIC = −6.32). Comparing the magnitude and rank order of the freely estimated factor loadings between middle-aged (MR = 0.506, SDS = 0.539, VNR = 0.614, TMTb = −0.719) and older (MR = 0.522, SDS = 0.492, VNR = 0.569, TMTb = −0.701) participants also suggested that *g* exhibited weak factorial invariance between groups (coefficient of factor congruence = 1.00).

We then investigated whether the proportion of *g* variance accounted for by MRI measures was substantially greater in older than middle-aged participants. Results are shown in Table 3. A model with unconstrained *g*-MRI associations fitted the data significantly better than when the associations were constrained to equality between age groups ($\Delta\chi^2(7) = 183.22$, $p \leq .001$, ΔAIC = 169, ΔsaBIC = 134). In the model in which *g*-MRI associations were allowed to differ by age group, GM, NAWM, WMH, *g*FA and *g*MD together explained a total of 5.4% of the variance in *g* among the middle aged group, compared to 13.6% in older age. *g*-brain association magnitudes were all stronger in older age for GM (0.159 versus 0.298), WMH (−0.092 versus −0.132), *g*FA (0.051 versus −0.096) and *g*MD (0.012 versus −0.131); the one exception was NAWM volume (which showed stronger *g* associations in middle than older age; 0.134 versus 0.054). Moreover, we did not observe this significant age difference in variance explained in *g* by TBV alone, or by each individual MRI measure in isolation (all *p*-values for chi-squared tests were non-significant following FDR correction, with
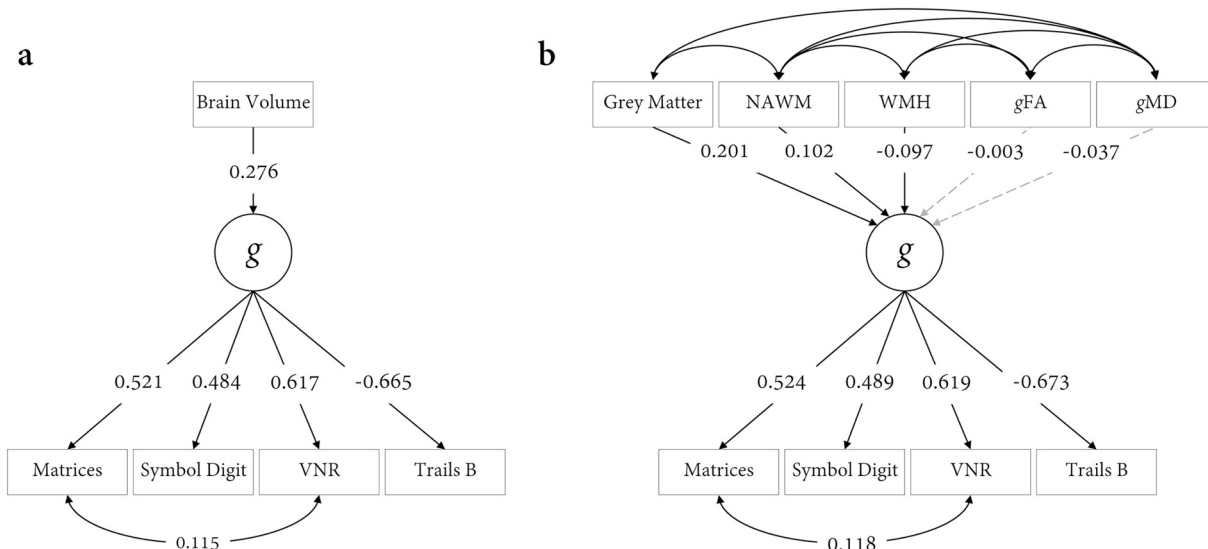

Fig. 3. Associations between global brain MRI measures and *g*. Panel **a)** shows associations with total brain volume, and panel **b)** shows tissue-specific brain MRI measures accounting for 6.54% of the variance in *g*. Standardised estimates are reported; grey dashed paths are non-significant. Indicators are all corrected for age, sex, with imaging data also corrected for scanner head position coordinates. MRI residual correlations are shown in Table S4.

Table 3
Individual and unique contributions to *g* from global MRI measures across middle and older age groups.

Model	Middle (≤ 63.29 yrs)				Older (> 63.29 yrs)			
	Individual		Simultaneous		Individual		Simultaneous	
	Std. Est.	*p*	*Std. Est.*	*p*	*Std. Est.*	*p*	*Std. Est.*	*p*
TBV	0.287	< 0.001	–	–	0.308	< 0.001	–	–
R²	*0.082*		–	–	*0.095*		–	–
GM	0.276[a]	< 0.001	0.159	< 0.001	0.328[a]	< 0.001	0.298	< 0.001
NAWM	0.257	< 0.001	0.134	< 0.001	0.268	< 0.001	0.054	< 0.001
WMH	−0.115[a]	< 0.001	−0.092	< 0.001	−0.165[a]	< 0.001	−0.132	0.007
gFA	0.113	< 0.001	0.051	0.029	0.108	< 0.001	−0.096	< 0.001
gMD	−0.065		0.012	0.628	−0.126		−0.131	< 0.001
R²			*0.054*				*0.136*	

Note. Std. Est: standardised estimate. Groups split at 63.29 years. [a]Magnitudes were significantly different by age, according to a χ^2 test (FDR *q* < 0.05). Models are corrected for sex; brain measures also corrected for head positioning confounds. Associations between *g* and TBV were not significantly different between middle and older ages: $\Delta\chi^2(1) = 3.874$, *p* = .049, ΔAIC = 2, ΔsaBIC = −3.223. However, the magnitude of *g* associations with multiple global measures (simultaneously modelled) were significantly different between age groups: $\Delta\chi^2(7) = 183.22$, *p* ≤ .001, ΔAIC = 169, ΔsaBIC = 134. TBV: total brain volume, GM: grey matter volume, WM: white matter volume, WMH: white matter hyperintensity volume, FA: fractional anisotropy, MD: mean diffusivity.

the exception of GM and WMH, which was more strongly associated with *g* in the older group. (GM: 0.276 versus 0.328; $\Delta\chi^2(1) = 10.827$, *p* < .001; WMH: −0.115 versus −0.165; $\Delta\chi^2(1) = 6.262$, *p* = .012).

In summary, whereas the individual associations between brain measures and intelligence were of relatively similar magnitudes in middle and older age, when entered together they explained more than double the variance in *g* in older age. Inspection of the age differences in the correlational structure among these imaging markers (Fig. S2) indicated that the source of increased variance explained is not likely to be attributable to their diverging collinearity (i.e. they overlap less in older age, and thus convey more unique information), given that the only notable differences were in associations between WMH and both gFA and gMD, which were stronger in older than younger age.

3.3. Sex differences in g-brain MRI associations

Before testing for sex differences in the size of the associations between *g* and brain MRI measures, we tested for measurement invariance between the sexes. We found that the model of strong factorial invariance of *g* did not fit more poorly than the model in which factor loadings and intercepts were freely-estimated (ΔAIC = 0, ΔsaBIC = −31, $\Delta\chi^2 (6) = 11.821$, *p* = .066). These results are reported in Table S5. We then tested for sex differences in the magnitude of associations between *g* and MRI measures. We did so by comparing two group models; one in which the brain-*g* association is fixed to equality between sexes, and the other in which it is freely estimated. We found that there were no significant sex differences in the magnitude of the association between *g* and total brain volume (females $\beta = 0.260$, males $\beta = 0.214$ when freely estimated; both $\beta = 0.276$ when constrained to be equal) or for any global brain MRI measure (all *p* ≥ .117) except for NAWM (*p* = .008, standardised estimate for females = 0.236, males = 0.177), which remained significant following FDR correction.

3.4. Associations between g and white matter microstructure

SEMs testing associations between *g* and the FA and MD of each white matter tract fitted the data well (all CFI ≥ 0.995, TLI ≥ 0.987, RMSEA ≤0.015, SRMR ≤0.009); results are shown in Fig. 4, and Tables S6 and S7. Associations with *g* were in the expected direction, such that higher *g* was related to higher FA and lower MD. Only a few pathways had non-significant associations with *g* (FA and MD in the left acoustic radiation, FA in the middle cerebellar peduncle, and MD in the right parahippocampal cingulum, Forceps Major, and bilateral medial lemniscus). The effect sizes were not homogeneous across tracts (FA range = 0.012 to 0.110; MD range = −0.100 to 0.007). Consistent

with our hypothesis, the magnitude of associations with *g* were numerically largest within thalamic pathways (FA mean = 0.078, MD mean = −0.091), and in association fibres and Forceps Minor (FA mean = 0.062, MD mean = −0.049) than within projection fibres and Forceps Major (FA mean = 0.039, MD mean = 0.027).[1] However, it is also notable that both aspects of the cingulum bundle showed among the weakest *g* relationship among association fibres, and that more generally there was a considerable amount of overlap between these classes of tract (for example, the right corticospinal tract MD was associated with *g* at levels comparable with most association fibres).

3.5. Associations between g and cortical regions

Associations between *g* and cortical regional volumes were all positive and all significant following FDR correction. The results are reported in Fig. 5 and in Table S8; all models fitted the data well (CFI ≥ 0.995, TLI ≥ 0.985, RMSEA ≤0.018, SRMR ≤0.010). As with the white matter analyses above, there was regional heterogeneity in association magnitudes across the cortical surface. Substantial portions of the frontal lobe (frontal pole, frontal orbital, subcallosal) were among the numerically largest associations, bilaterally (range = 0.166 to 0.216), and these were significantly larger than other frontal regions (*p* < .001). Associations between the insula cortex and *g* (left = 0.194, right = 0.205) were also large compared to the average magnitude across all ROIs (M = 0.116, SD = 0.036). Notably, the temporal lobe (range = 0.152 to 0.062) exhibited a gradient of anterior > posterior for both lateral and medial portions, and the lateral surface also showed evidence of a superior > inferior gradient. Compared to the above-mentioned frontal, anterior temporal and insula volumes, parietal regions were consistently and significantly more weakly associated with *g* (range = 0.066 to 0.100, *p* < .001). With the exception of the lingual, precuneus, and lateral occipital cortex (range = 0.110 to 0.156), occipital volumes were among the most weakly associated with *g* (range = 0.065 to 0.093).

3.6. Associations between g and subcortical volumes

As with the cortical analyses, subcortical volumes were all positively associated with *g*, and all were significant following FDR correction. The results are reported in Table S9. All models fitted the data well (CFI ≥ 0.994, TLI ≥ 0.983, RMSEA ≤0.018, SRMR ≤0.011), and

---

[1] We note that the splitting of the commissural fibres in this way was a pre-registered hypothesis.
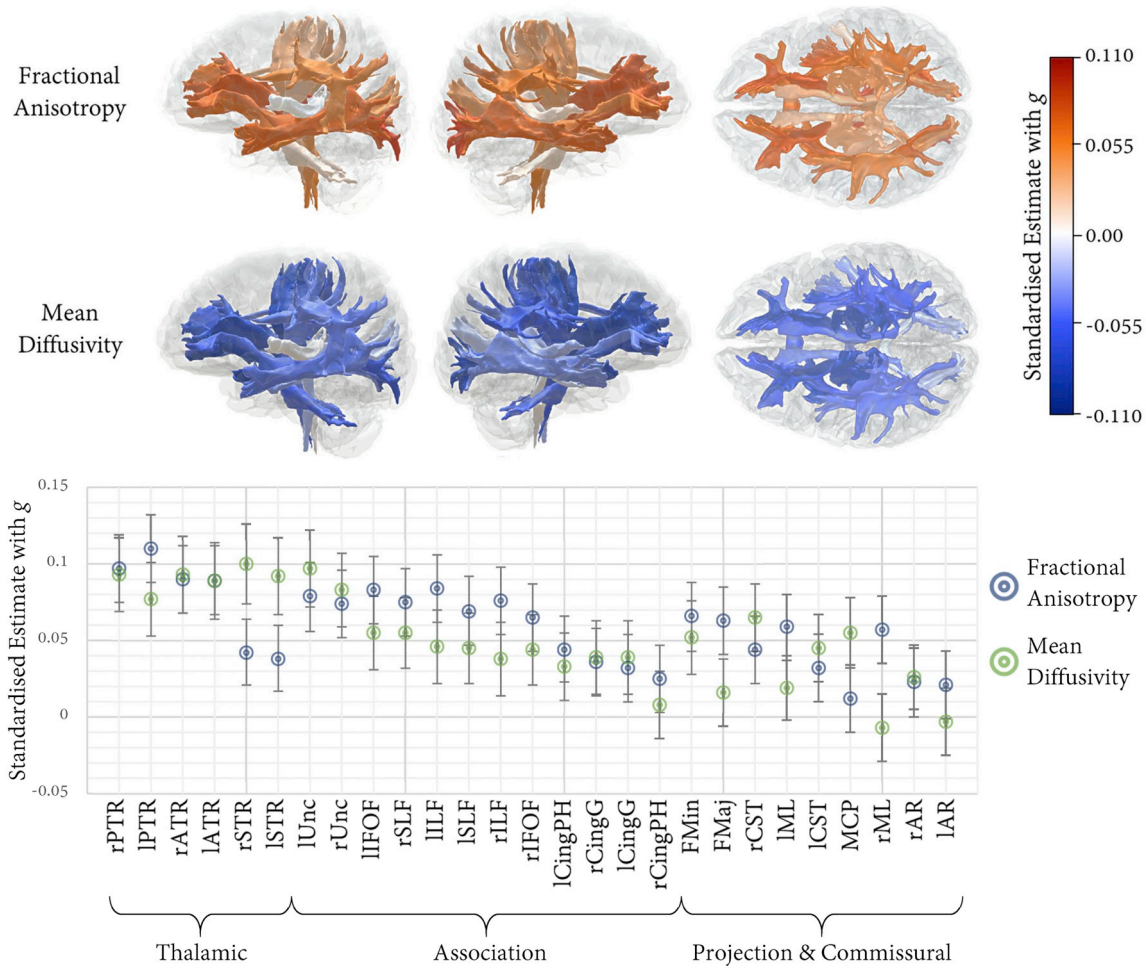
Fig. 4. Associations between white matter tract-specific microstructure and *g*. Tabulated results also reported in Tables S6 and S7. Top panel shows left and right lateral and superior views of the white matter tracts of interest, heatmapped according to association magnitude. Lower panel displays the association magnitudes sorted by tract class and then from strongest to weakest (based on the average of FA and MD), with 95% CIs; MD valences have been flipped to aid visual comparison.
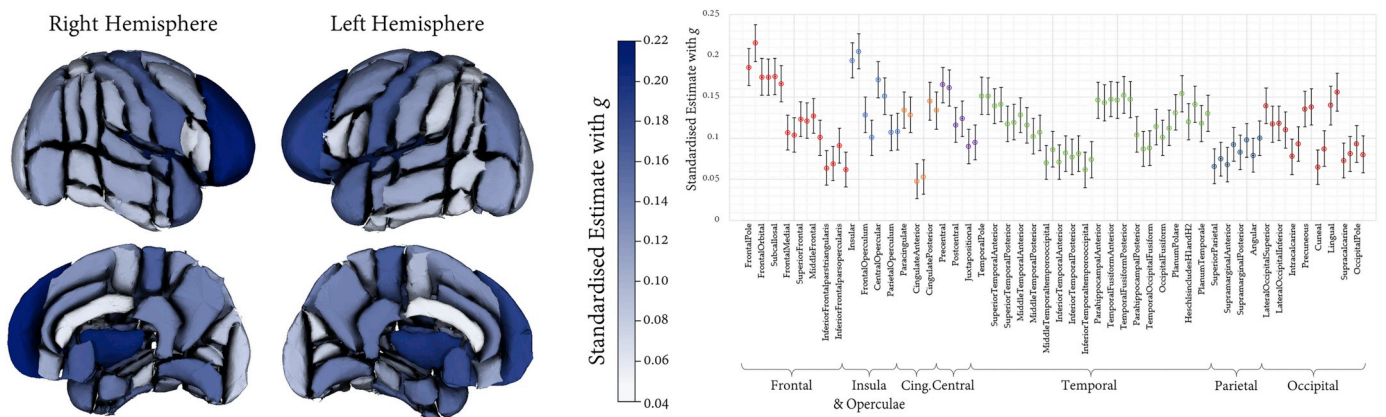

Fig. 5. Associations between regional cortical volumes and *g* with 95% CIs. Left and right associations are shown separately (left hand regions appear first). Association magnitudes are also reported in Table S5.

standardised estimates ranged from 0.062 to 0.256. Largest effect sizes were found for the Thalamus (left = 0.251, right = 0.256), which was significantly larger than for all other subcortical regions (*p* < .001). The Amygdala showed the weakest associations of all (left = 0.075, right = 0.062), whereas the remaining volumes (Accumbens, Caudate, Hippocampus, Pallidum and Putamen) showed comparable magnitudes (range = 0.105 to 0.165).

3.7. Out of sample prediction of g using multivariate MRI data

We investigated whether simultaneously modelling the regional MRI predictors of *g* would substantially alter the pattern of regional associations, when compared to the results of creating a single model for each region, as reported above. We tested the generalisability of these findings by splitting the sample by scanning site, assigning
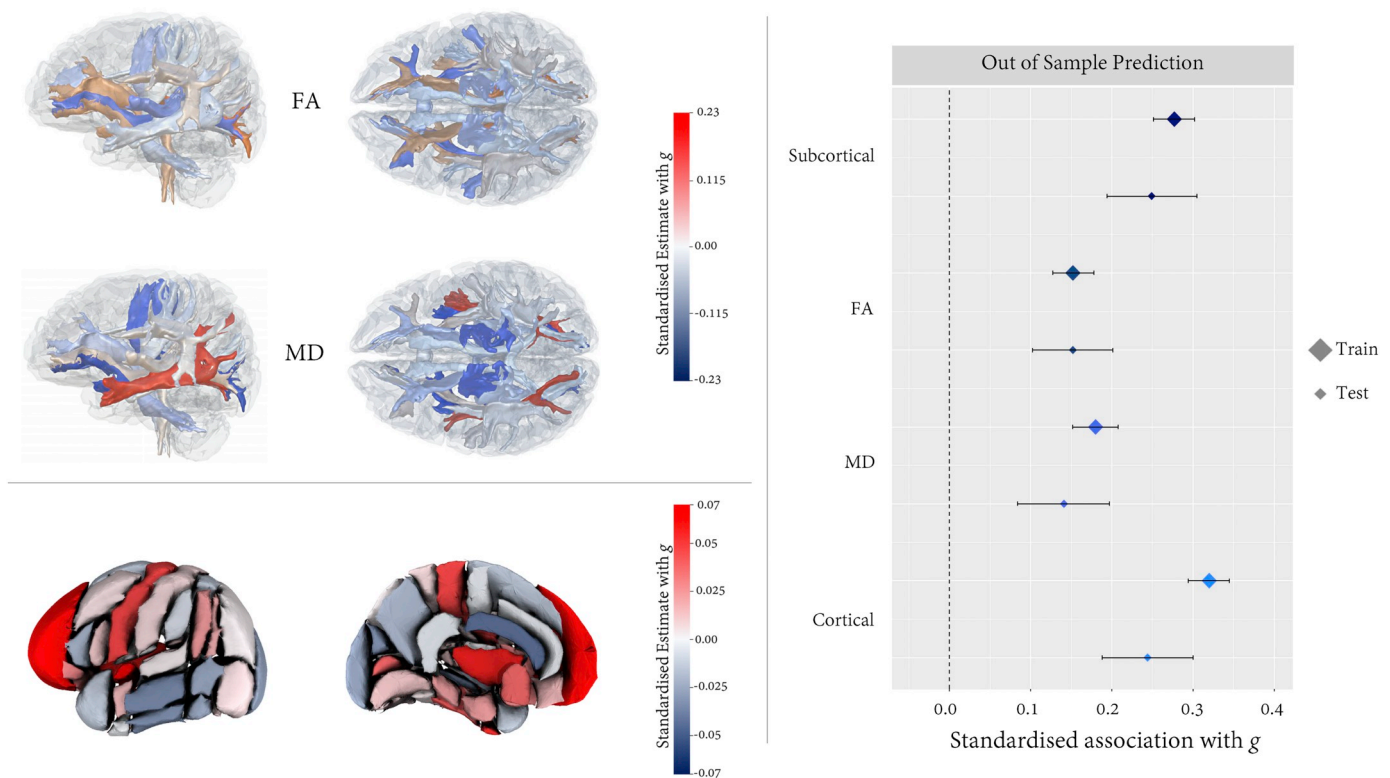
Fig. 6. Out of sample prediction (Manchester to Newcastle) of *g* from regional MRI data. Left panel shows the spatial distribution of standardised beta weights. Right panel shows the associations (standardised estimates and 95% confidence intervals) between *g* and the weighted composite scores (derived using those weights) in the training and test samples.

Table 4

Out of sample prediction (Manchester to Newcastle) of *g* from regional MRI data.

		Est.	SE	*p*	ci.lower	ci.upper	Complete N
FA	Train	0.152	0.013	< 0.001	0.127	0.178	4768
	Test	0.152	0.025	< 0.001	0.102	0.201	2510
MD	Train	0.180	0.014	0.003	0.152	0.208	4707
	Test	0.141	0.029	0.001	0.084	0.197	2494
Cortical	Train	0.320	0.013	< 0.001	0.294	0.345	5246
	Test	0.244	0.028	< 0.001	0.188	0.300	2589
Subcortical	Train	0.277	0.013	< 0.001	0.251	0.302	5253
	Test	0.249	0.028	< 0.001	0.194	0.305	2595

Note. Standardised estimates (Est.), standard errors (SE) and 95% CIs (ci.upper and ci.lower) are reported. FA: fractional anisotropy, MD: mean diffusivity.

Manchester as the training data, and Newcastle as the test data. The beta weights obtained from the MIMIC models in the training data are shown in Fig. 6 and Tables S10–13. The associations between *g* and the resultant weighted composite score are shown in Table 4 and Fig. 6).

In the training data, unique contributions to *g* variance were made by higher FA in the anterior and posterior thalamic radiations, the uncinate and middle cerebral peduncle (β range 0.053 to 0.128), but with lower FA in the acoustic radiations and inferior fronto-occipital fasciculus ($\beta = -0.060$, and $\beta = -0.082$, respectively). For MD, beta weights were highest in the expected direction in the posterior and superior thalamic radiations, uncinate and middle cerebellar peduncle (β range -0.030 to -0.177), but the association was positive in the inferior longitudinal fasciculus ($\beta = 0.225$). For cortical volumes, *g* was most strongly associated with larger fronto-polar, subcallosal, insula, precentral, anterior superior temporal, temporal fusiform, and intracalcarine cortices (β range 0.038 to 0.062), but also with a smaller anterior cingulate gyrus and cuneal volume ($\beta = -0.041$, and $\beta = -0.032$, respectively). Finally, larger thalamus ($\beta = 0.194$),

putamen ($\beta = 0.225$) and hippocampal ($\beta = 0.024$) volumes uniquely predicted *g*.

The weighted composite scores showed a highly stable out of sample prediction of *g*. In the case of FA, the point estimates were identical to three decimal places ($\beta = 0.152$), and *g* associations were also of comparable magnitude for MD (train $\beta = 0.180$; test $\beta = 0.141$), cortical (train $\beta = 0.320$; test $\beta = 0.244$) and subcortical (train $\beta = 0.277$; test $\beta = 0.249$) volumes. Correcting for TBV partly attenuated associations between *g* and composite weighted scores for cortical (range 22.5% to 40.9%) and subcortical (range 2.5% to 20.5%) volumes, but they remained significant predictors of *g*, whose magnitudes were stable across training and test data (Table S14).

4. Discussion

In this large sample of middle and older aged participants, we found that the association between total brain volume and a latent factor of general intelligence was $r = 0.276$. The current single-cohort analysis was not confounded by cross-cohort heterogeneity in the protocol for intelligence and brain size measurement which have affected recent meta-analyses of this association (McDaniel, 2005; Rushton & Ankney, 2009; Pietschnig et al., 2015; Gignac & Bates, 2017). This estimate is at the mid-point between the meta-analytic effect size estimate from Pietschnig et al., (2015; $r = 0.24$) and the quality-corrected estimate of Gignac and Bates (2017; $r = 0.31$). It is also considerably larger than previous estimates using a single cognitive indicator (verbal numerical reasoning) in an earlier UK Biobank release ($r = 0.19$, $N = 13,608$; Nave et al., 2018; $r = 0.177$, $N = 5216$; Ritchie et al., 2018), emphasising the utility of our latent variable approach, which was also informed by a larger sample. The fact that the association between *g* and TBV was not significantly different between sexes is in contrast the results reported by McDaniel (2005), but not with a larger, more recent meta-analysis (Pietschnig et al., 2015), and prior work in an earlier UK

Biobank release using the VNR only (Ritchie et al., 2018). Given that the mean differences in intelligence are generally extremely modest or null (Deary, Irwing, Der, & Bates, 2007; Johnson, Carothers, & Deary, 2008; Lakin & Gambrell, 2014; Ritchie et al., 2018), this adds weight to the hypothesis that more specific brain characteristics compensate for the relatively larger brain size difference between males and females.

We also ascertained that the g-TBV association belies important heterogeneity at both the global tissue level, and at the regional level across cortex, subcortex and white matter. GM and NAWM were the strongest global tissue correlates of g, with WMH and microstructural measures showing weaker but significant associations in separate models. However, when modelled simultaneously in a MIMIC model, unique contributions of WMH and NAWM were near identical (GM still largest), whereas information about white matter microstructure did not carrying any unique information about individual differences in g. Together, these measures explained only 6.16% of g variance. This was substantially lower than prior estimates using similar global structural brain metrics (Ritchie et al., 2015) which explained as much as 21% g variance. Given that their participants were from an older cohort – all participants were born in 1936 and were approximately 73 years old at scanning – we conducted a post-hoc analysis to ascertain whether the brain measures would account for more variance in older than in younger participants in our sample. We found that these measures accounted for more than double the g variance in older participants compared to those in middle-age (13.6% and 5.4%, respectively); thus while still smaller than that accounted for in Ritchie et al. (2015), this does support the notion that age moderates the relationship between general intelligence and multiple aspects of brain tissue structure. This age moderation pattern was only observed in a multi-predictor analysis that simultaneously included multiple MRI-based predictors and not in an analysis of total brain volume or individual MRI predictor alone.

These age differences stood in contrast to the apparent age invariance of the association between g and total brain volume (as found by Pietschnig et al., 2015). This supports the notion that total brain volume is a proxy for several other aspects of brain integrity whose variances are i) uniquely informative for cognitive function and ii) more informative than brain size alone. Moreover, total brain volume is likely to be an age-varying indicator of brain integrity, raising questions about the value of considering the brain size-intelligence relationship, in isolation, for furthering our mechanistic insight into the cerebral basis of intelligence. Overall, the age moderation pattern suggests that g may be less sensitive to the variance around 'healthier' / 'younger' averages (higher GM and NAWM volume, lower WMH volume, higher FA and lower MD).

In our analysis of regional brain correlates of intelligence, the cortical and grey matter associations were stronger than for the regional white matter microstructural parameters, on average, though association magnitudes were all of small effect size (Cohen, 1992). Nearly all regional measures were significant following FDR correction, and our findings of regional heterogeneity of g associations across the tissues of the brain were partly consistent with our hypotheses based on the P-FIT (Jung & Haier, 2007). Specifically, we hypothesised that g would be most strongly positively associated with lateral frontal, superior parietal and temporal cortical volumes, and show stronger (positive for FA, negative for MD) relationships with g in thalamic and association fibres, plus forceps minor. In accordance with this, we found relatively stronger associations in regions such as the frontal pole, dorsolateral frontal cortex, paracingulate, anterior aspects of both lateral and medial temporal lobes, and lateral occipital cortex. However, the comparatively weaker associations in inferior frontal, anterior cingulate, and superior parietal / angular / supramarginal areas were less consistent with P-FIT. Moreover, medial frontal regions (orbitofrontal and subcallosal), central and precentral gyri were among the strongest associations here, but were not explicitly implicated in prior reviews (Basten et al., 2015; Jung & Haier, 2007), and we also found associations with the insula and precuneus / posterior cingulate volumes which were only more recently implicated in general intelligence (Basten et al., 2015), and concurs with more recent insights into the dense and wide-ranging connectivity profile of the insula (Nomi, Schettini, Broce, Dicks, & Uddin, 2018). With reference to the white matter pathways, magnitudes were consistently smaller than for cortical regional volumes, but were strongest among thalamic and most association pathways, along with the forceps minor, which facilitate connectivity across many of the distal cortical regions highlighted by the P-FIT model.

Finally, we opted to include subcortical volumes in a post-hoc (non-pre-registered) analysis. Consistent with prior reports (Basten et al., 2015; Grazioplene et al., 2015; Rhein et al., 2014), we found significant bilateral associations with the caudate, though these were not significantly larger than the magnitudes found for the majority of subcortical structures. In fact, thalamic volume was substantially more strongly related to general intelligence (≥ 1.5 times as large) than any other subcortical structure (r for left and right = 0.255 and 0.251). This finding is in line with the highly complex connectivity profile of the thalamus, whose various nuclei share connections across much of the cortex (including prefrontal and hippocampal pathways; Behrens et al., 2003; Aggleton et al., 2010), its role in orchestrating cortical activity as well as an information relay (Rikhye, Wimmer, & Halassa, 2018), and a prior report of its phenotypic and genetic associations with intelligence (Bohlken et al., 2014). It is also consistent with previously-reported associations of the thalamus and its radiations with ageing, and to potential determinants thereof, such as vascular risk (Cox et al., 2016; Cox et al., 2019). However, it is also notable that the association between g and all subcortical structures, though not as large as for the Thalamus, were still comparable or larger than those exhibited by white matter microstructural measures.

Following peer review, we also included an analysis that considered the unique contributions of each regional MRI measure to intelligence, accounting for their correlational structure (and thus also their contribution beyond the more global, brain-wide metrics to which they all contribute). We also ascertained the robustness of these results by testing their ability to predict g out of sample, exploiting the fact that the data had been sampled from two testing sites (Manchester and Newcastle). The regions making the strongest unique contributions to intelligence were broadly in line with our bivariate results: frontal pole, subcallosal, insula, anterior lateral and medial temporal cortical volumes, microstructure of the thalamic radiations and uncinate fasciculus, alongside greater volumes of the thalamus, putamen and hippocampus, though the effect sizes were all substantially lower than when modelled individually (which is to be expected given their large degree of collinearity). Notably, the weighted composite score based on the analysis in the training set showed excellent out of sample performance, adding some weight to the generalisability of these findings.

Our supplementary analyses showed that the regional volumetric information accounted for significant g variance beyond total brain volume, though we note that interpreting this type of correction is complex, especially when correcting the individual regions for volume during the initial training model; the resultant weightings indicate unique regional volumetric contributions to cognitive ability beyond all other regions and also beyond TBV (for which all regional volumes are proxies, to some degree). As such, each weighting represents the importance of cortical or subcortical configural differences, were all brains the same size. Given the disproportionate regional composition of the cortex as a function of brain size (Essen, 2018), these results are therefore more difficult to directly reconcile with fundamental questions of 'where in the brain is bigger better?'. Finally, we observe that whereas our uncorrected out of sample predictions did not appear to explain substantially more g variance than total brain volume alone, increasing the regional specificity of that explanation might offer more tractable insights into the underlying cerebral basis of general cognitive ability.

The study has several limitations. The information reported here is correlational in nature, and though it describes what intelligent brains look like (insofar as these are some of the axes along which brains differ as a function of intelligence), it cannot directly differentiate between regions that are and are not *required* to support the cognitive processes subsumed

beneath the umbrella of *g*. Nevertheless, it continues to be of interest and value to robustly quantify how and where brain structure and intelligence are associated; along with longitudinal data, lesion studies and other methodologies, such studies will help to triangulate the contributions that brain regions play in giving rise to individual differences in *g*. The study sample is also range restricted in three respects. First, they are members of a voluntary research study (and UK Biobank is known to be range restricted in some ways compared to the general population; Fry et al., 2017), and participants in this study are also what is known as "WEIRD" (from Western, educated, industrialised, rich and democratic societies; Henrich, Heine, & Norenzayan, 2010), meaning that these results are obtained in a doubly-selective group. Second, we know that the brain imaging subset of UK Biobank participants tends to live in less deprived areas (Lyall et al., 2019); given the known associations between SES and cognitive function, the sample is also likely to therefore be range restricted with respect to brain and cognitive measures. Third, the age range of the sample is restricted to middle and older age, omitting important maturational periods of life where different global and regional metrics may be differentially relevant to intelligence differences. These sample limitations may affect the generalisability of our results with respect to the total brain, global tissue or regional results, and the pattern of age moderation; these whole-life-course patterns could more optimally be addressed in a large scale multi-cohort mega-analytic framework. The degree to which these findings apply to non-WEIRD participants would also clearly benefit from future work. Though the reliability of the more recent cognitive tests that we used here is not known, we note that the loadings and proportion of variance explained would be very unlikely to occur if they were unreliable tests, and that these compare favourably with the correlational structure of other UK Biobank cognitive tests where test-retest reliability is known to be low (Lyall et al., 2014). Moreover, the tests selected here were based on well-validated cognitive tests, and a paper covering their design and reporting results of a validation study of this enhanced cognitive battery in UK Biobank is the subject of ongoing work by the authors (CFR and IJD). We also did not correct our full-sample analyses for assessment site. The sites are described as "identical" in UK Biobank Brain Imaging Documentation. Correcting for variability in head positioning inside the scanner may allow any potential systematic differences in the implementation of the same UK Biobank sampling protocol across sites to be mitigated while preserving any small but potentially meaningful variability that might otherwise have been eliminated by a nominal covariate. The close replication of *g* predictions out-of-sample do not suggest any non-trivial confounding by scanner differences. Finally, it could be argued that the brain imaging methods might limit the fidelity with which we can measure the regional specificity of *g* associations across the brain. The 27 major pathways have the advantage of being well characterised and aid consistent identification across subjects, but they do not allow a direct measure of the WM connectivity between specific cortical or subcortical sites of the brain in native space, which would allow for a more precise and stringent test of *g* associations with the WM pathways underlying the P-FIT, as well as a less biased set of pathways (for example, the current dataset has more information on thalamic connectivity than on other subcortical pathways). Similarly, the cortical parcellation used here was one of convenience and does not correspond directly onto the Brodmann Areas used by Jung and Haier (2007), which makes mapping the current findings onto prior hypotheses opaque. For example, whereas the Harvard-Oxford atlas includes a paracingulate region (Brodmann Area 32), this additional cortical fold is not always present, and thus is perhaps more usefully referred to as "superior medial" cortex (e.g. see Cox et al., 2014). Importantly, the superior lateral occipital area likely incorporates cortical territories that other parcellation schemas would designate as parietal; otherwise the parietal areas in the present schema are relatively small, which may account for the relatively stronger association between *g* and superior lateral occipital cortex. Likewise, frontal pole region subsumes a large portion of the frontal lobe compared to that described by Brodmann and is likely to include a sizeable portion of anterior dorsolateral prefrontal areas (BA 9/46). Though these concordance issues are well-known,

and there is no straightforward solution (Bohland, Bokil, Allen, & Mitra, 2009; Cox et al., 2014), it is important to interpret the results with these limitations in mind.

In conclusion, this preregistered study provides a large single sample analysis of the global and regional brain correlates of a latent factor of general intelligence. Our study design avoids issues of publication bias and inconsistent cognitive measurement to which meta-analyses are susceptible, and also provides a latent measure of intelligence which compares favourably with previous single-indicator studies of this type. We estimate the correlation between total brain volume and intelligence to be *r* = 0.276, which applies to both males and females. Multiple global tissue measures account for around double the variance in *g* in older participants, relative to those in middle age. Finally, we find that associations with intelligence were strongest in frontal, insula, anterior and medial temporal, lateral occipital and paracingulate cortices, alongside subcortical volumes (especially the thalamus) and the microstructure of the thalamic radiations, association pathways and forceps minor.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2019.101376.

References

Aggleton, J. P., O'Mara, S. M., Vann, S. D., Wright, N. F., Tsanov, M., & Erichsen, J. T. (2010). Hippocampal–anterior thalamic pathways for memory: Uncovering a network of direct and indirect actions. *European Journal of Neuroscience, 31*, 2292–2307.

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., ... Smith, S. M. (2018). Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage, 166*, 400–424.

Basten, U., Hilger, K., & Fiebach, C. J. (2015). Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence, 51*, 10–27.

Behrens, T. E. J., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshot, C. A., Boulby, P. A., ... Matthews, P. M. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature Neuroscience, 6*, 750–757.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and

powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–300.

Bohland, J. W., Bokil, H., Allen, C. B., & Mitra, P. P. (2009). The brain atlas concordance problem: Quantitative comparison of anatomical parcellations. *PLoS One, 4*, e7200.

Bohlken, M. M., Brouwer, R. M., Mandl, R. C., van Haren, N. E., Brans, R. G., can Baal, G. C., ... Hulshoff Pol, H. E. (2014). Genes contributing to subcortical volumes and intellectual ability implicate the thalamus. *Human Brain Mapping, 35*, 2632–2642.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 119–155.

Cox, S. R., Bastin, M. E., Ritchie, S. J., Dickie, D. A., Liewald, D. C., Muñoz Maniega, S., ... Deary, I. J. (2018). Brain cortical characteristics of lifetime cognitive ageing. *Brain Structure and Function, 223*, 509–518.

Cox, S. R., Ferguson, K. J., Royle, N. A., Shenkin, S. D., MacPherson, S. E., MacLullich, A. M., & Deary, I. J. (2014). Wardlaw JM. A systematic review of brain frontal lobe parcellation techniques in magnetic resonance imaging. *Brain Structure and Function, 219*, 1–22.

Cox, S. R., Lyall, D. M., Ritchie, S. J., Bastin, M. E., Harris, M. A., Buchanan, C. R., ... Deary, I. J. (2019). Assocations between vascular risk factors and brain MRI indices in UK Biobank. *European Heart Journal.* https://doi.org/10.1093/eurheartj/ehz100.

Cox, S. R., Ritchie, S. J., Tucker-Drob, E. M., Liewald, D. C., Hagenaars, S. P., Davies, G., ... Deary, I. J. (2016). Aging and brain white matter structure in 3513 UK Biobank participants. *Nature Communications, 7*, 13629.

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother-sister differences in the *g* factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence, 35*, 451–456.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*, 201–211.

Dick, A. S., & Tremblay, P. (2012). Beyond the arcuate fasciculus: Consensus and controversy in the connectional anatomy of language. *Brain, 135*, 3529–3550.

Essen, V. (2018). Scaling of human brain size. *Science, 360*, 1184–1185.

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., ... Allen, N. E. (2017). Comparions of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *American Journal of Epidemiology, 186*, 1026–1034.

Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., Ling, J. M., Güntürkün, O., & Jung, R. E. (2018). Diffusion markers of dendritic density and aborization in gray matter predict differences in intelligence. *Nature Communications, 9*, 1905.

Gignac, G., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence, 64*, 18–29.

Grazioplene, R. G., Ryman, S. G., Gray, J. R., Rustichini, A., Jung, R. E., & De Young, C. G. (2015). Subcortical intelligence: Caudate volume predicts IQ in healthy adults. *Human Brain Mapping, 36*, 1407–1416.

Griffanti, L., Zamboni, G., Khan, A., Li, L., Bonifacio, G., Sundaresan, V., ... Jenkinson, M. (2016). BIANCA (brain intensity abnormality classification algorithm): A new tool for automated segmentation of white matter hyperintensities. *NeuroImage, 141*, 191–205.

de Groot, M., Vernooij, M. W., Klein, S., Ikram, M. A., Vos, F. M., Smith, S., ... Andersson, J. L. (2013). Improving alignment in tract-based spatial statistics: Evaluation and optimization of image registration. *NeuroImage, 76*, 400–411.

Hagenaars, S. P., Cox, S. R., Hill, W. D., Davies, G., Liewald, D. C. M., CHARGE consortium Cognitive Working Group, ... Deary, I. J. (2018). Genetic contributions to trail making test performance in UK Biobank. *Molecular Psychiatry, 23*, 1575–1583.

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C. M., Ritchie, S. J., ... Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112 151) and 24 GWAS consortia. *Molecular Psychiatry, 21*, 1624–1632.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466*, 29.

Jacobucci, R., Brandmaier, A. M., & Kievit, R. A. (2019). A practical guide to variable selection in structural equation modelling by using regularised multiple-indicators, multiple-causes models. *Advances in Methods and Practices in Psychological Science, 2*, 55–76.

Johnson, W., Carothers, A., & Deary, I. J. (2008). Sex differences in variability in general intelligence: A new look at the old question. *Perspectives in Psychological Science, 3*, 518.531.

Jung, R. E., & Haier, R. J. (2007). The Parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *The Behavioural and Brain Sciences, 30*, 135–154.

Kamali, A., Flanders, A. E., Brody, J., Hunter, J. V., & Hasan, K. M. (2014). Tracing superior longitudinal fasciculus connectivity in the human brain using higher resolution diffusion tensor tractography. *Brain Structure and Function, 219*, 269–281.

Karama, S., Colom, R., Johnson, W., Deary, I. J., Haier, R., Waber, D. P., Lepage, C., Ganjavi, H., Jung, R., & Evans, A. C. (2011). Brain Development Cooperative Group, Cortical thickness correlates of specific cognitive performance accounted for by the general factor of intelligence in healthy children aged 6 to 18. *NeuroImage, 55*, 1443–1453.

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., Team, C.-C. A. N. R., & Henson, R. N. A. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multi-tasking. *Nature Communications, 5*, 5658.

Kievit, R. A., Fuhrmann, D., Borgeest, G. S., Simpson-Kent, I. L., & Henson, R. N. A. (2018). The neural determinants of age-related changes in fluid intelligence: A preregistered, longitudinal analysis in UK Biobank. *Wellcome Open Research, 3*, 38.

Kievit, R. A., van Rooijen, H., Wicherts, J. M., Waldorp, L. J., Kan, K.-J., Scholte, H. S., & Borsboom, D. (2012). Intelligence and the brain: A model-based approach. *Cognitive Neuroscience, 3*, 89–97.

Lakin, J. M., & Gambrell, J. L. (2014). Sex differences in fluid reasoning: Manifest and latent estimates from the cognitive abilities test. *Journal of Intelligence, 2*, 36–55.

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*, 57–64.

Luders, E., Narr, K. L., Thompson, P. M., Rex, D. E., Jancke, L., Steinmetz, H., & Toga, A. W. (2004). Gender differences in cortical complexity. *Nature Neuroscience, 7*, 799–800.

Lyall, D. M., Cox, S. R., Lyall, L. M., Celis-Morales, C., Cullen, B., Mackay, D. F., ... Pell, J. P. (2019). Association between APOE e4 and white matter hyperintensity volume, but not total brain volume or white matter integrity. *Brain Imaging and Behavior.* https://doi.org/10.1007/s11682-019-00069-9.

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., ... Pell, J. P. (2016). Cognitive test scores in UK Biobank: Data reduction in 480,416 participants and longitudinal stability in 20, 346 participants. *PLoS One, 11*, e0154222.

MacPherson, S. E., Allerhand, M., Cox, S. R., & Deary, I. J. (2019). Individual differences in cognitive processes underlying trail making test-B performance in old age: The Lothian birth cohort 1936. *Intelligence, 75*, 23–32.

MacPherson, S. E., Cox, S. R., Dickie, D. A., Karama, S., Starr, J. M., Evans, A. C., ... Deary, I. J. (2017). Processing speed and the relationship between trail making test-B performance, cortical thinning and white matter microstructure in older adults. *Cortex, 95*, 92–103.

McDaniel, M. A. (2005). Big-brained people are smarter: A meta-analysis of the relationship between in vivo brain volume and intelligence. *Intelligence, 33*, 337–346.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.

Nave, G., Jung, W. H., Linnér, R. K., Kable, J. W., & Koellinger, P. D. (2019). Are bigger brains smarter? Evidence from a large-scale preregistered study. *Psychological Science, 30*, 43–54.

Nomi, J. S., Schettini, E., Broce, I., Dicks, A. S., & Uddin, L. Q. (2018). Structural connections of functionally defined human insular subdivisions. *Cerebral Cortex, 28*, 3445–3456.

Patenaude, B., Smith, S. M., Kennedy, D. N., & Jenkinson, A. (2011). Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage, 56*, 907–922.

Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience and Biobehavioural Reviews, 57*, 411–432.

Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan neuropsycholgical test battery: Therapy and clinical interpretation.* Tucson, AZ: Neuropsychological Press.

Rhein, C., Muhle, C., Richter-Schmidinger, T., Alexopoulos, P., Doerfler, A., & Kornhuber, J. (2014). Neuroanatomical correlated of intelligence in healthy young adults: The role of basal ganglia volume. *PLoS One, 9*, e93623.

Rikhye, R. V., Wimmer, R. D., & Halassa, M. M. (2018). Toward an integrative theory of thalamic function. *Annual Review of Neuroscience, 41*, 163–183.

Ritchie, K., de Roquefeuil, G., Ritchie, C. W., Besset, A., Poulin, V., Artero, S., & Ancelin, M. (2014). COGNITO: Computerised assessment of information processing. *Journal of Psychology and Psychotherapy, 4*, 2.

Ritchie, S. J., Booth, T., Valdés Hernández, M. C., Corley, J., Munoz Maniega, S., Gow, A. J., ... Deary, I. J. (2015). Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence, 51*, 47–56.

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M., Reus, L. M., Alloza, C., ... Deary, I. J. (2018). Sex differences in the adult human brain: Evidence from 5216 UK Biobank participants. *Cerebral Cortex, 28*, 2959–2975.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modelling. *Journal of Statistical Software, 48*, 1–36.

Rushton, J. P., & Ankney, C. D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience, 119*, 691–731.

Ryman, S. G., Yeo, R. A., Witkiewitz, K., Vakhtin, A. A., van den Heuvel, M., de Reus, M., ... Jung, R. E. (2016). Fronto-parietal gray matter and white matter efficiency differentially predict intelligence in males and females. *Human Brain Mapping, 3711*, 4006–4016.

Salthouse, T. A. (2011). What cognitive abilities are involved in trail-making performance? *Intelligence, 39*, 222–232.

Smith, A. (1991). *Symbol digit modalities test.* Los Angeles, CA: Western Psychological Services.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., ... Biobank, C. R. U. K. (2015). An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med. Public Library of Science, 12*, e1001779.

Valdés Hernández, M. C., Cox, S. R., Kim, J., Royle, N. A., Muñoz Maniega, S., Gow, A. J., ... Deary, I. J. (2017). Hippocampal morphology and cognitive functions in community-dwelling older people: The Lothian birth cohort 1936. *Neurobiology of Aging, 52*, 1–11.

Widaman, K. F., Ferrer, F., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*, 10–18.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging, 20*, 45–57.