Registered Report

A quantified comparison of cortical atlases on the basis of trait morphometricity

Anna E. Fürtjes [a,*], James H. Cole [b,c,d], Baptiste Couvy-Duchesne [e,f,1] and Stuart J. Ritchie [a,1]

[a] Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK
[b] Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
[c] Centre for Medical Image Computing, Department of Computer Science, University College London, London, UK
[d] Dementia Research Centre, Institute of Neurology, University College London, London, UK
[e] Paris Brain Institute (ICM), Inserm (U 1127), CNRS (UMR 7225), Sorbonne University, Inria Paris, Aramis Project-team, Paris, France
[f] Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia

ABSTRACT

Background: Many different brain atlases exist that subdivide the human cortex into dozens or hundreds of regions-of-interest (ROIs). Inconsistency across studies using one or another cortical atlas may contribute to the replication crisis across the neurosciences.

Methods: Here, we provide a quantitative comparison between seven popular cortical atlases (Yeo, Desikan-Killiany, Destrieux, Jülich-Brain, Gordon, Glasser, Schaefer) and vertex-wise measures (thickness, surface area, and volume), to determine which parcellation retains the most information in the analysis of behavioural traits (incl. age, sex, body mass index, and cognitive ability) in the UK Biobank sample (N~40,000). We use linear mixed models to compare whole-brain morphometricity; the proportion of trait variance accounted for when using a given atlas.

Results: Commonly-used atlases resulted in a considerable loss of information compared to vertex-wise representations of cortical structure. Morphometricity increased linearly as a function of the log-number of ROIs included in an atlas, indicating atlas-based analyses miss many true associations and yield limited prediction accuracy. Likelihood ratio tests revealed that low-dimensional atlases accounted for unique trait variance rather than variance common between atlases, suggesting that previous studies likely returned atlas-specific findings. Finally, we found that the commonly-used atlases yielded brain-behaviour associations on par with those obtained with random parcellations, where specific region boundaries were randomly generated.

Discussion: Our findings motivate future structural neuroimaging studies to favour vertex-wise cortical representations over coarser atlases, or to consider repeating analyses across

---

Sex

Alcohol consumption

Age

Cigarette smoking

multiple atlases, should the use of low-dimensional atlases be necessary. The insights uncovered here imply that cortical atlas choices likely contribute to the lack of reproducibility in ROI-based studies.

1. Introduction

To better understand neuronal correlates of human behaviour, studies typically investigate the association between brain structural measurements in regions-of-interest (ROI) and behavioural traits. Recent efforts to collect large-scale neuroimaging data enable unprecedented opportunities for powerful statistical analyses. One of the most used software for structural brain analysis is *FreeSurfer* (Fischl et al., 2002) which is used to determine regional brain measures from a participant's brain image. Briefly summarised, FreeSurfer models the cortex in a two-dimensional mesh, with an arbitrary resolution of ~150,000 vertices per hemisphere (known as "fsaverage standard space"). Measurements are extracted at each vertex, and include cortical thickness, surface area and volume. With reference to a *cortical atlas* (Fig. 1), the software can subdivide the cortical mesh into ROIs and calculates their structural characteristics, such that the surface area—or grey-matter volume or cortical thickness.

Cortical atlases aim to outline structurally homogeneous and meaningful regions that reflect the organisation of the cortex, but there is no ground-truth parcellation. Different organisational characteristics are inferred based on distinct brain modalities, including anatomical landmarks, cytoarchitecture, and patterns of functional coactivation. Here, we consider seven available atlases: Desikan-Killiany (DK; Desikan et al., 2006), Destrieux (Destrieux, Fischl, Dale, & Halgren, 2010), Glasser (Glasser et al., 2016), Gordon (Gordon et al., 2014), Schaefer (Schaefer et al., 2017), Yeo (Yeo et al., 2011), Jülich-Brain (Amunts, Mohlberg, Bludau, & Zilles, 2020). Atlases differ in their anatomical boundaries and number of ROIs and were generated using different methods and samples.

Desikan-Killiany (68 ROIs) and Destrieux (148 ROIs) are landmark-based atlases, which means that the cortex is divided in a manner consistent with the macroscopic anatomy of gyri and sulci. ROIs in Desikan-Killiany were manually labelled on structural MRI scans from a sample of 40 participants between 19 and 87 years (mean age 55.95 years; Desikan et al., 2006), who had been originally recruited with a range of atrophy levels by the Washington University Alzheimer's Disease Research Center (Fotenos, Snyder, Girton, Morris, & Buckner, 2005). The Destrieux atlas follows what the authors described as "widely accepted anatomical conventions" (page 2; Destrieux et al., 2010) and was derived from 12 healthy participants between 18 and 33 years (mean age 21.67).

Yeo (34 ROIs), Gordon (333 ROIs) and Schaefer (500 ROIs) were derived from resting-state functional MRI data. Gordon et al. (2014) quantify gradients of functional activation across the cortex and use abrupt changes in these gradients as indicators of regional borders. Gordon's atlas was derived from 120 healthy community-dwelling participants between the ages of 19–32 years (mean age 25). Schaefer et al. (2017) integrates this approach by maximising uniformity within regions while neighbouring voxels are only assigned to the same area if abrupt gradient changes do not separate them. Schaefer et al. used images of 1,489 brains from the Genomics Superstruct Project (participants aged between 18 and 35 years). Yeo et al. (2011) outline 17 macroscopic network configurations that were stably estimated using clustered functional connectivity data from 1,000 healthy participants (mean age 21.3 years).

Glasser et al. (2016, 360 ROIs) parcellate regions using a semi-automated approach where regional boundaries were defined from multiple indicators including cortical architecture, function, and connectivity measures. It was outlined from 200 participants between 22 and 35 years in the Human Connectome Project cohort. Finally, Amunts et al. (2020) present the 'Jülich-Brain', a microstructural parcellation of 137 cortical areas in each hemisphere, reflecting cytoarchitecture across the cortex, which was derived based on 23 postmortem brains (version 2.9). The mean age of this sample was 64 years (range 30–86).

Different atlases outline different regional boundaries with spatial discrepancies (Alexander-Bloch et al., 2018; Bohland, Bokil, Allen, & Mitra, 2009), which likely influences study results. A recent study showed that atlas choice affects estimates of network topology and functional brain connectivity (Revell et al., 2022). In another study, different atlases either induced or masked associations between ROIs and age, meaning that comparable regions were associated with age in one atlas but not in another (Yaakub et al., 2020). The fact that many different atlases are frequently used introduces uncertainty when comparing and attempting to reproduce study results; probably exacerbating the existing lack of consensus about associations between the brain and behaviour across neuroimaging studies (Kharabian Masouleh, Eickhoff, Hoffstaedter, Genon, & Alzheimer's Disease Neuroimaging Initiative, 2019; Marek et al., 2022). The optimal atlas may differ for any given study, and to our knowledge, there is no data-driven guideline about which atlas maximises brain-trait association and prediction of specific behavioural traits.

While variance captured by ROIs is reduced compared with vertex-wise measures, the main advantage of employing cortical atlases is to reduce dimensionality which facilitates interpretation. Recently, a novel statistical framework was presented (Sabuncu et al., 2016), allowing consideration of *all* brain vertices simultaneously to estimate whole-brain associations with behavioural traits. This approach accounts for correlations between vertices and negates the need to summarise vertex measures across ROIs. Using mixed linear models, this method quantifies the proportion of trait
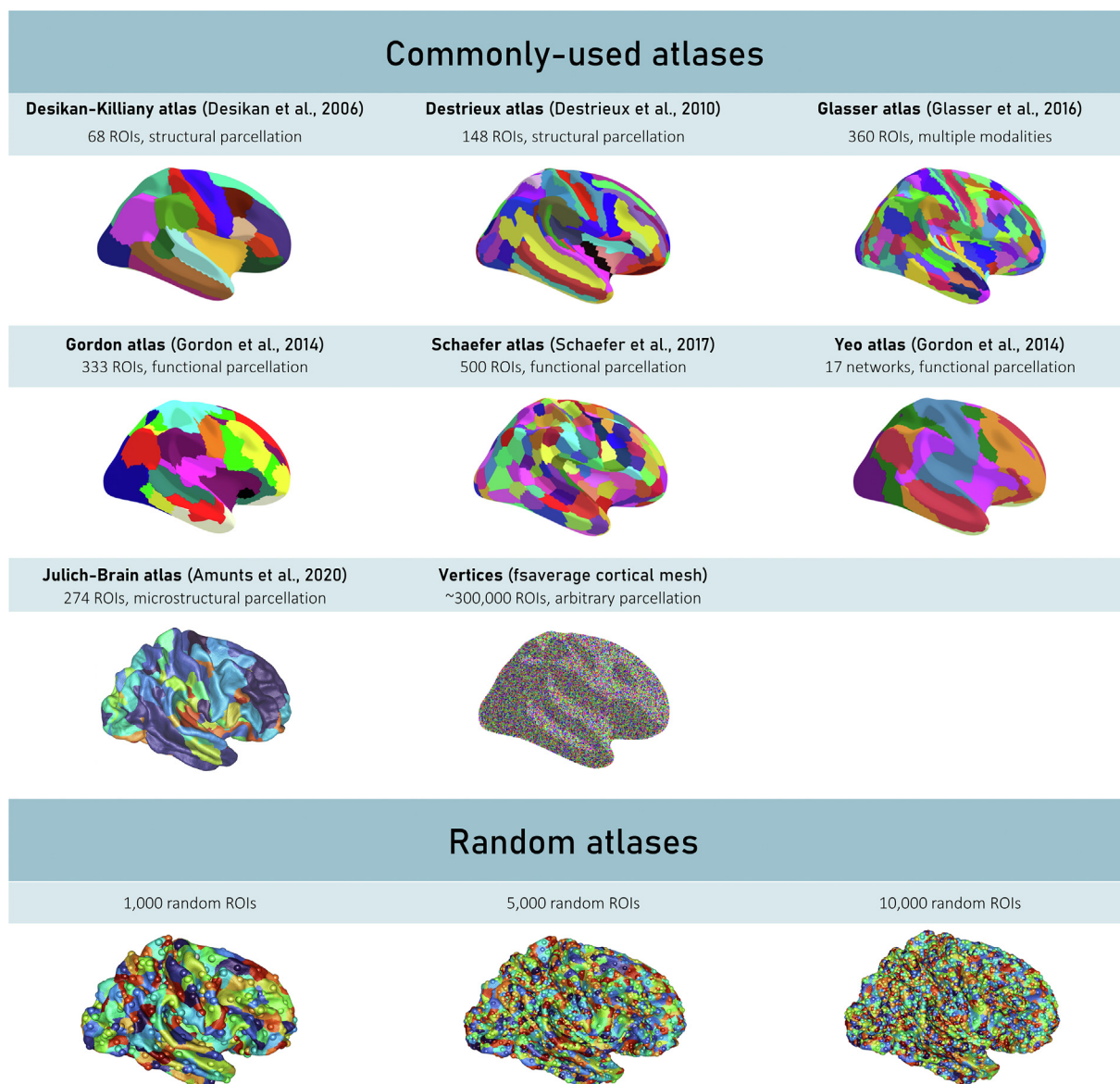
Fig. 1 − The surface-based cortical representations considered in this Registered Report (right hemisphere view). The first six commonly-used atlases were visualised with the ggseg package in R (Mowinckel & Vidal-Piñeiro, 2020). Note that the Yeo atlas annotation files dictate 17 networks in each hemisphere, resulting in 34 ROIs. Vertices were visualised based on 150,000 randomly simulated points. Some colours were generated with the *circulize* package in R (Gu, Gu, Eils, Schlesner, & Brors, 2014). The Julich-Brain atlas and the random parcellations were visualised from 3D coordinates using the *rgl* package in R (https://dmurdoch.github.io/rgl/). In random atlases, vertices with larger radius indicate the seeds from which random ROIs were grown. We also analysed a random atlas with 50,000 ROIs (i.e., 5 times the number of regions as displayed for the 10,000 ROIs), which is not shown in the figure.

variance (R^2) that can be attributed to brain morphology, or *morphometricity*. Conceptually, morphometricity is equivalent to *heritability* in genetics, which indicates the proportion of trait variance explained by many small polygenic effects.

A recent study estimated morphometricity for hundreds of traits, demonstrating that vertex-wise brain measures explained substantial proportions of trait variance (Couvy-Duchesne et al., 2020). In a supplementary analysis, DK ROI-summarised brain measures seemed to explain considerably less variance than the vertex-wise measures, though no statistical significance tests were performed to test this difference. This effect was most pronounced for age, for which the proportion of variance explained by vertex-wise cortical measures was $R^2{\sim}65\%$, but only $R^2{\sim}25\%$ when derived from DK measures. This suggests that averaging across ROIs masks informative inter-individual variance which is retained in vertex-wise cortical representations. Vertex-wise morphometricity has been shown to be robust across samples, and it explained considerably more variance in traits such as fluid intelligence ($R^2{\sim}13.2\%$) compared with recent predictions models using different methods to maximise brain-based predictive ability ($R^2{\sim}7\%$) of behavioural

traits (Gong, Beckmann, & Smith, 2021). This illustrates that morphometricity estimates could be promising indicators for how well different representations of the brain (vertices or different atlases) can account for individual differences. This could help formulate a recommendation for which brain representation is most promising to capture structural brain associations with behavioural traits.

This Registered Report (https://osf.io/dkw9t/) is the first study to compare commonly-used cortical atlases of varying dimensionality by quantifying and contrasting morphometricity estimates of behavioural variables. We hypothesised that more fine-grained cortical representations would yield larger morphometricity estimates compared with coarser atlases, because explanatory variance is lost when averaging across ROIs. We expected at least a two-fold increase in morphometricity between the coarsest and most fine-grained atlas, which was formulated based on preliminary results in Couvy-Duchesne et al. (2020). It is reasonable to expect that some atlases may be detailed enough to capture maximal morphometric variance. Such an atlas should outperform randomly-generated ROIs and vertex-wise representations, as it summarises anatomically coherent ROIs that should be unaffected by registration imprecision and anatomical variability. We expect to find trait-dependent optimal levels of atlas dimensionality, as per recent work assessing atlas performances based on functional MRI data (Dadi et al., 2020). There, the highest performing predictions in an age variable were achieved by ~150 ROIs, while ~300 ROIs were optimal for intelligence.

The UK Biobank study currently provides FreeSurfer data from which we extracted cortical vertex-wise measurements, as well as ROI-summarised cortical representations in seven commonly-used atlases: Yeo (34 ROIs), Desikan-Killiany (68 ROIs), Destrieux (148 ROIs), Jülich-Brain (274 ROIs), Gordon (333 ROIs), Glasser (360 ROIs), and Schaefer (500 ROIs). To better assess a relationship between morphometricity and dimensionality of the atlas, we randomly generated atlases with 1,000; 5,000; 10,000; and 50,000 ROIs. Morphometricity estimates based on surface area, grey-matter volume, and cortical thickness are reported for seven non-brain traits that are well-measured in UKB and known to be morphometric, that is, robustly associated with grey-matter structure: age (Cole et al., 2018), sex (Ritchie et al., 2018), cognitive abilities (Cox, Ritchie, Fawns-Ritchie, Tucker-Drob, & Deary, 2019), body mass index (Couvy-Duchesne et al., 2020), education (Sabuncu et al., 2016), cigarette smoking, and alcohol consumption (Couvy-Duchesne et al., 2020). We fitted mixed linear models to estimate the morphometricity (i.e. total association) between behavioural traits as dependent variables and either vertex-wise or ROIs measurements as independent variables. We tested for significant differences in morphometricity using likelihood ratio tests.

2. Materials and methods

2.1. *UKB neuroimaging data*

The UK Biobank (UKB) study is a population-based cohort for health-related information from ~500,000 individuals across the United Kingdom (Sudlow et al., 2015). Among baseline characteristics, physical, and cognitive assessments, it provides pre-processed MRI data from around 40,000 participants (Littlejohns et al., 2020). The Research Ethics Committee ethically approved the UKB study and participants signed informed consent.

Researchers have access to the FreeSurfer outputs (UKB field ID 20263; https://biobank.ndph.ox.ac.uk/showcase/field.cgi?id=20263) from the first neuroimaging visit that have been pre-processed and quality controlled on behalf of UKB (Alfaro-Almagro et al., 2018). Detailed documentation on T1-weighted + T2 FLAIR processing using FreeSurfer 6.0 can be found online (Smith, Alfaro-Almagro, & Miller, 2020). Combining two structural images in processing is considered more precise and associated with improved cortical segmentation (Lindroth et al., 2019) compared to processing T1w images only. The Freesurfer outputs already contained individual-level data with reference to Desikan-Killiany (Desikan et al., 2006), Destrieux, and unsmoothed vertex-wise data. The vertex-wise measurements considered here correspond to the "fsaverage" cortical mesh representation (~300,000 vertices both hemispheres). We derived individual-level cortical measurements with reference to Yeo et al. (2011), Destrieux et al., (2010), Desikan et al. (2006), Glasser et al. (2016), Gordon et al. (2014), Schaefer et al. (2017), and Jülich-Brain (Amunts et al., 2020). Most fsaverage atlas annotation files were obtained from MultiAtlasTT (https://github.com/faskowit/multiAtlasTT. Click or tap if you trust this link.">https://github.com/faskowit/multiAtlasTT), which were in *.gcs* format and were easily mapped using FreeSurfer 6.0 commands (mris_ca_label & mris_anatomical_stats). From this resource, Schaefer atlas annotation files were only available for a maximum of 500 ROIs. Annotation files for Yeo et al. (2011) were downloaded as part of the FreeSurfer 6.0 software. Jülich-Brain atlas (Amunts et al., 2020) annotation files (v2.9) were obtained from Mangin, Rivière, and Amunts (2021). Mapping atlases from *.annot* files required a purpose-built pipeline, which we share on GitHub (https://annafurtjes.github.io/Comparing_atlases/). Surface area and volume ROIs are summarised as the sum of vertices, and cortical thickness is an average measure.

2.2. *Random parcellations*

We generated contiguous random parcellations by randomly selecting and summarising across vertices in the cortical mesh. The number of parcellations matched the number of ROIs contained in the commonly-used atlases (Fig. 1). We also generated an additional four, more fine-grained parcellations with 1,000; 5,000; 10,000 and 50,000 random ROIs. To ensure that parcellations resulted in contiguous ROIs, random seeds were selected from which we grew random ROIs. We used the *vcgKDtree* R clustering algorithm (Rvcg package) to iteratively grow the ROIs, adding at each iteration six spatially proximal vertices (when possible). This yielded ROIs of roughly the same size. We hypothesised that if random atlases yielded larger morphometricity compared with commonly-used atlases, it would indicate that morphometricity depends on dimensionality and not the exact ROI-specific boundaries outlined in atlases. In this manuscript, we refer to the

commonly-used and random brain parcellations as well as vertex-wise measures using the notion *atlases* that are composed of *ROIs*.

2.3. UKB traits of interest

Morphometricity was estimated for non-brain traits to quantify the proportion of variance accounted for when using cortical measures represented by ROIs from a given atlas. We selected the following dependent variables, because they are highly morphometric, and well-phenotyped in the UKB sample. We considered age (field ID 21003), sex (field ID 31) and body mass index (field ID 21001). We used measures of cognitive ability to construct a "g" factor of general cognitive ability using the lavaan package in R (Rosseel, 2012). We used a confirmatory factor analysis with one factor, based on the following cognitive ability tests: Verbal Numerical Reasoning (field ID 20016, 20191), Trail Making − B (field ID 6350, 20157), Matrix Pattern Completion (field ID 6373), Tower Rearranging (field ID 21004), Symbol Digit Substitution (field ID 23324 & 20159), Pairs Matching (field ID 399), and Reaction Time (field ID 200032). The tests have been described in detail elsewhere, and as in previous work, we considered the first measured occasion for each participant (de la Fuente, Davies, Grotzinger, Tucker-Drob, & Deary, 2021). Compared with standard reference tests, the cognitive assessments exhibit good concurrent validity ($r = .83$) and test-retest reliability (mean Pearson $r = .55$) (Fawns-Ritchie & Deary, 2020). Usually, this general cognitive ability factor explains about 40% of the total variance in the included cognitive tests (Deary, Penke, & Johnson, 2010); here it accounts for 34%. According to commonly used heuristics (Hu & Bentler, 1998), the factor demonstrated good model fit (CFI = .97, RMSEA = .05; SFig. 1).

Additionally, we estimated morphometricity for educational qualification (field ID 6138), number of cigarettes smoked daily (field ID 2887) and frequency of drinking alcohol (field ID 20414; excluding former drinkers field ID 20406). Outliers that fall beyond 4 standard deviations from the mean were removed from the sample. This criterion was only applied to non-brain traits, as this criterion applied to brain measures created too many missing data entries for LMMs to reliably converge. Covariates included in all models consisted of UKB acquisition site (field ID 54), and head positioning in the MRI scanner (X, Y, Z coordinates, field IDs 25756, 25757, 25758).

2.4. Statistical analyses

Morphometricity. Morphometricity quantifies the proportion of inter-individual differences in a non-brain trait (dependent variable) accounted for by cortical measurements (independent variables). The dependent variables in this study are listed above and include age, sex, body mass index, and cognitive ability. Independent variables include the following ROI measures within cortical atlases: either ~300,000 vertex-wise measures, or atlas-wide brain ROIs, which include different amounts of ROIs depending on the atlas. Additionally, we considered four covariates (outlined above in UKB traits of interest).

To estimate morphometricity, we fitted linear mixed models (LMMs, Fig. 2). This approach recognises the

correlation structure between many independent variables, and was presented and validated elsewhere, using vertex-wise data (Couvy-Duchesne et al., 2020). Briefly, the LMM fits all cortical measurements as a vector of random effects that is constrained to a normal distribution and a structure of variance-covariance derived from the brain relatedness matrix (BRM, **B**). The BRM quantifies how similar participants are to one another based on cortical measures. We excluded 4 participants with outlying covariance (±8SD from mean brain-relatedness) from the analyses as this indicated oddly similar or dissimilar brains and could bias the LMM results. More detailed model definitions are outlined in the Supplementary Material. All covariates held constant, we used Restricted Maximum Likelihood (REML) implemented in the OSCA software (OmicS-data-based Complex trait Analysis; Zhang et al., 2019) to estimate the variance of the brain random effect (σ_b^2). σ_b^2 quantifies the total trait variance captured by all vertex-wise measurements, while σ_e^2 quantifies the residual variance accounted for by the error term. Morphometricity will be determined as follows (Couvy-Duchesne et al., 2020; Sabuncu et al., 2016):

$$R^2 = \sigma_b^2 / (\sigma_b^2 + \sigma_e^2)$$

As this is a Registered Report, we had introduced the detection of above-zero morphometricity as a "positive control", to test whether the proposed analyses allow for a fair test of the stated hypothesis. All morphometricity estimates presented in the results were significantly larger than zero, which confirms that there was enough systematic variance in the data to detect differences in atlas performance, should they exist. Though not pre-registered, we report in the Results that log-linear models described the relationship between morphometricity and atlas dimensionality well (it became obvious when plotting the data that there was a logarithmic relationship).

Comparison of atlases. To compare and test differences in morphometricity between two atlases (i.e., quantify the morphometricity specific to each atlas or shared between atlases), we extended the previous model by fitting two nested LMMs: the null model fits one variance component for all ROIs within one cortical atlas as random effects (non-brain trait ~ intercept + covariates + atlas1), and the alternative model includes an additional variance component for all ROIs within a second atlas (non-brain trait ~ intercept + covariates + atlas1 + atlas2). Likelihood ratio tests (LRT) allowed examination of whether morphometricity estimates (R^2) significantly differed between two commonly-used atlases (i.e., not the random atlases). The χ^2 distributed LRT statistic contrasts the improvement of fit from the null to the alternative model against the loss in degrees of freedom. To adjust for multiple testing, we considered Bonferroni corrected *p*-values below .05/(588) = 8.5×10^{-5} as significant. Refer to Fig. 2 for an illustration. In practice, we performed 588 pair-wise comparisons between commonly-used atlases across all measurement types and traits (7 non-brain traits x 3 measurement types x 28 pairs of atlases). Each comparison contrasted morphometricity within each trait-brain measurement type configuration.

In each LRT, the loss in degrees of freedom equals one, because we model *one* variance component per atlas. This
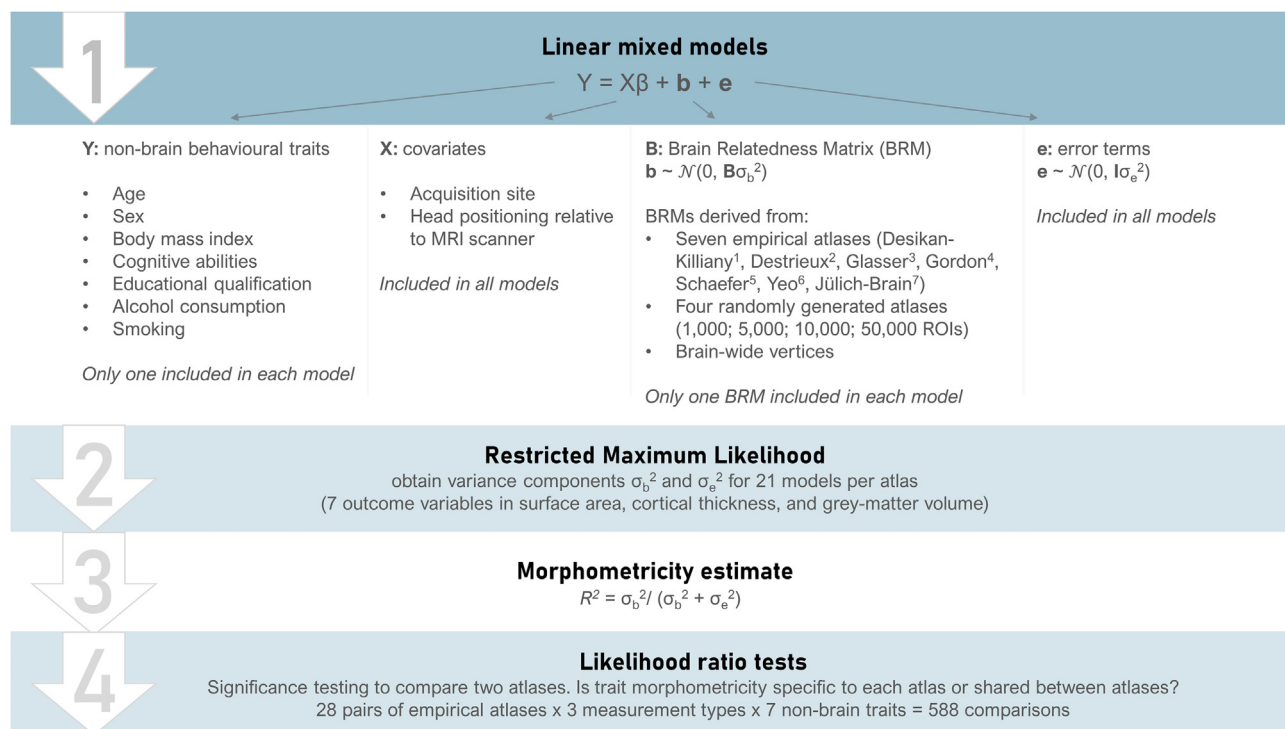
Fig. 2 − **Step-by-step illustration of the statistical analyses described in this Registered Report. The linear mixed model equation in step 1 is shown to illustrate the variables considered. This equation will not be solved *per se*, instead we will use Restricted Maximum Likelihood on the basis of model assumptions outlined in the methods to obtain variance components (step 2), that allow calculation of the morphometricity estimate (step 3). Likelihood ratio tests will only compare commonly-used atlases and vertex-wise measures against one another, but not random parcellations. [1]Desikan et al. (2006); [2]Destrieux et al. (2010); [3]Glasser et al. (2016); [4]Gordon et al. (2014); [5]Schaefer et al. (2017); [6]Yeo et al. (2011); [7]Amunts et al. (2020).**

means that testing does not depend on the numbers of cortical measurements contained in the different atlases, allowing a fair LRT comparison. To illustrate this lack of estimate inflation with more predictors, consider genome-wide complex trait analysis heritability estimates in which millions of genetic components are fit to a trait as random effects (Yang, Lee, Goddard, & Visscher, 2011). This is a well-established approach using LMMs equivalent to the one used here to estimate morphometricity. It is known that the heritability estimates do not alter when the number of genetic predictors considered in the analysis are increased or reduced, if they remain genome-wide representative. Refer to the supplementary material for pre-registered power calculations.

Best fitting model. In addition to the LRT, we calculated the Bayesian Information Criterion (BIC) (Schwarz, 1978) to compare and rank the models based on their penalised fit. The BIC is a well-established measure of fit that was calculated as the difference between model complexity ($\log(N_{participants})*p$; with p the number of parameters estimated in the model) and model fit ($2*LogLikelihood$), meaning that smaller values indicate better model fit.

Morphometricity of random parcellations. To understand whether commonly-used atlases explained more variance than expected using random atlas boundaries, we estimated the distributions of morphometricity estimates from 100 randomly generated atlases with the same number of ROIs as are included in the commonly-used atlases. For example, for comparison with Desikan-Killiany, we generated atlases with

68 random ROIs (across both hemispheres) and repeated the analysis 100 times, to obtain a distribution of morphometricity under random parcellations. We reported the percentage out of 147 estimates from commonly-used atlases (7 atlases x 7 traits x 3 measurement types) that were larger than 0%, 50%, and 99% of the 100 estimates from random parcellations.

2.5. Sensitivity analyses

Penalised regression. To test the robustness of the main results, we estimated how well grey-matter structure measures predicted non-brain traits by using LASSO models. LASSO is a "least absolute shrinkage and selection operator" (Tibshirani, 1996), that maximises predictive ability and can handle more predictors than observations. We estimated LASSO parameters on a training subset (random 80% sample subset, ~30,400 participants depending on number of total participants surviving data cleaning) with the big_spLinReg function in the bigstatsr package (Privé, Aschard, Ziyatdinov, & Blum, 2018). Using ten-fold cross-validation, the function separates the training set in ten folds, employs a Cross-Model Selection and Averaging procedure and outputs an averaged vector of coefficients.

We evaluated the prediction accuracy of the LASSO models in a validation sample (20%, ~7,600 participants) and report the prediction accuracy (i.e., R^2 between observed and predicted values), and mean absolute errors (i.e. mean absolute difference between observed and predicted measures of a
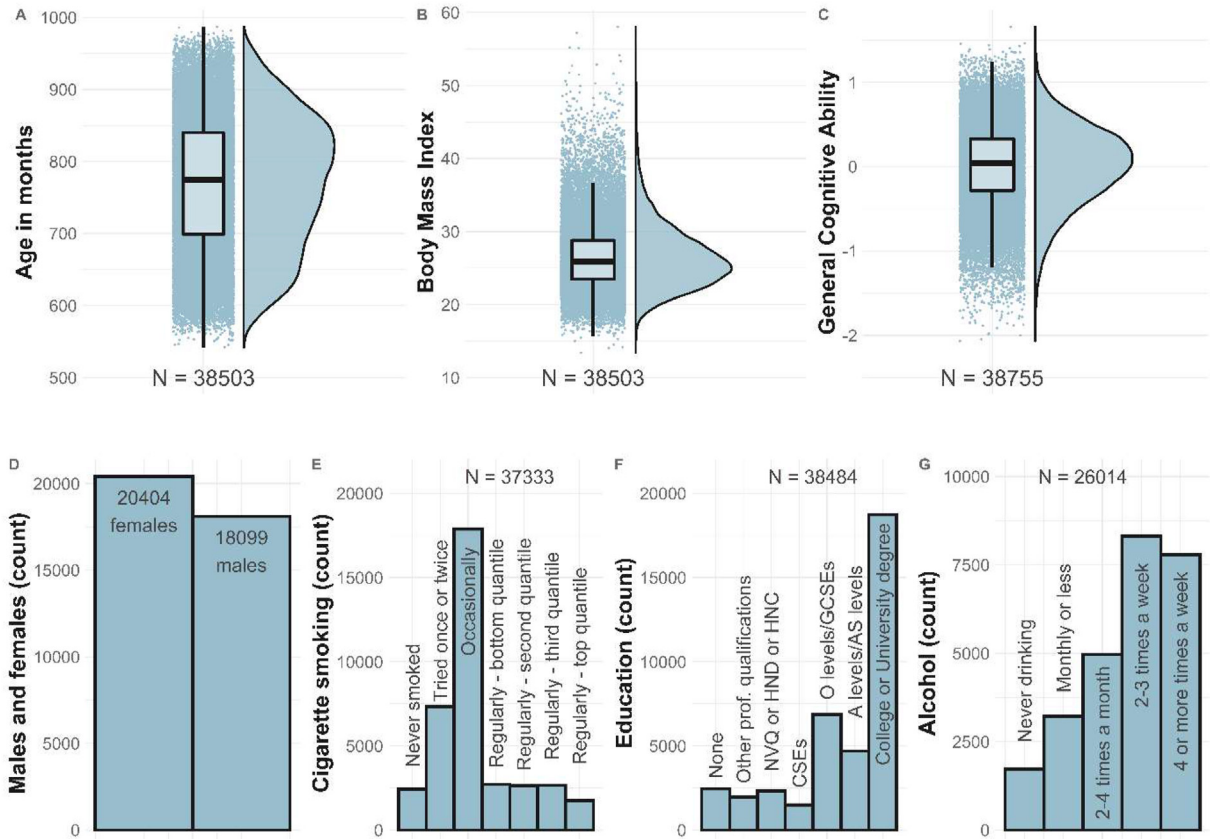
Fig. 3 − Descriptive statistics for each non-brain trait. Note the distribution for the education phenotype which is not representative of the general population but expected in the UKB sample (Fry et al., 2017).

non-brain trait). To statistically compare absolute errors between atlases, we contrasted the absolute errors obtained in models with different atlases using Wilcoxon's signed rank tests. To adjust for multiple testing, we considered Bonferroni corrected p-values below $.05/(588) = 8.5 \times 10^{-5}$ as significant.

3. Results

3.1. Descriptive statistics

Overall, we processed structural neuroimaging data for 42,957 participants (4 exclusions based on ±8SD BRM criterion). Participants who withdrew consent and those without T2-FLAIR measurements were removed from the phenotype data ($n_{removed} = 1,377$). Depending on data availability, differing numbers of participants were included in the morphometricity analysis of the seven non-brain traits. Final descriptive statistics are displayed by non-brain trait in Fig. 3.

3.2. Morphometricity estimates

First, we estimated the total association between each atlas and trait of interest, to assess if more complex atlases tend to retain more information. Morphometricity estimated from

the commonly-used atlases (Yeo, Desikan-Killiany, Destrieux, Glasser, Gordon, Jülich-Brain, Schaefer), random parcellations with 1,000, 5,000, 10,000 and 50,000 ROIs, as well as from vertex-wise measurements are displayed in Fig. 4. All traits were morphometric as indicated by significant LRTs examining whether the estimate was larger than zero (STable 1). Estimates differed between non-brain traits, atlases, and brain measurement types, as did the BIC quantifying model performances (SFig. 2). Overall, age and sex yielded the highest morphometricity out of the seven non-brain traits.

As hypothesised, we observed that atlases with more ROIs produced larger morphometricity across all seven traits and all brain measurement types. This was more pronounced for surface area than cortical thickness measures. The largest estimates were always yielded by vertex-wise measures. More generally, we observed substantial increases in morphometricity from high-dimensional compared to low-dimensional atlases (Fig. 4; STable 1). We found a median 3.41-fold increase in morphometricity (range increase = 1.49-fold−15.17-fold) obtained from vertex-wise measures (300,000 ROIs) compared with the Yeo atlas (34 ROIs). For example, we found 25% morphometricity estimated for sex from surface area measures represented by the Yeo atlas, compared with 88% morphometricity by vertex-wise measures (all raw estimates in STable 2).
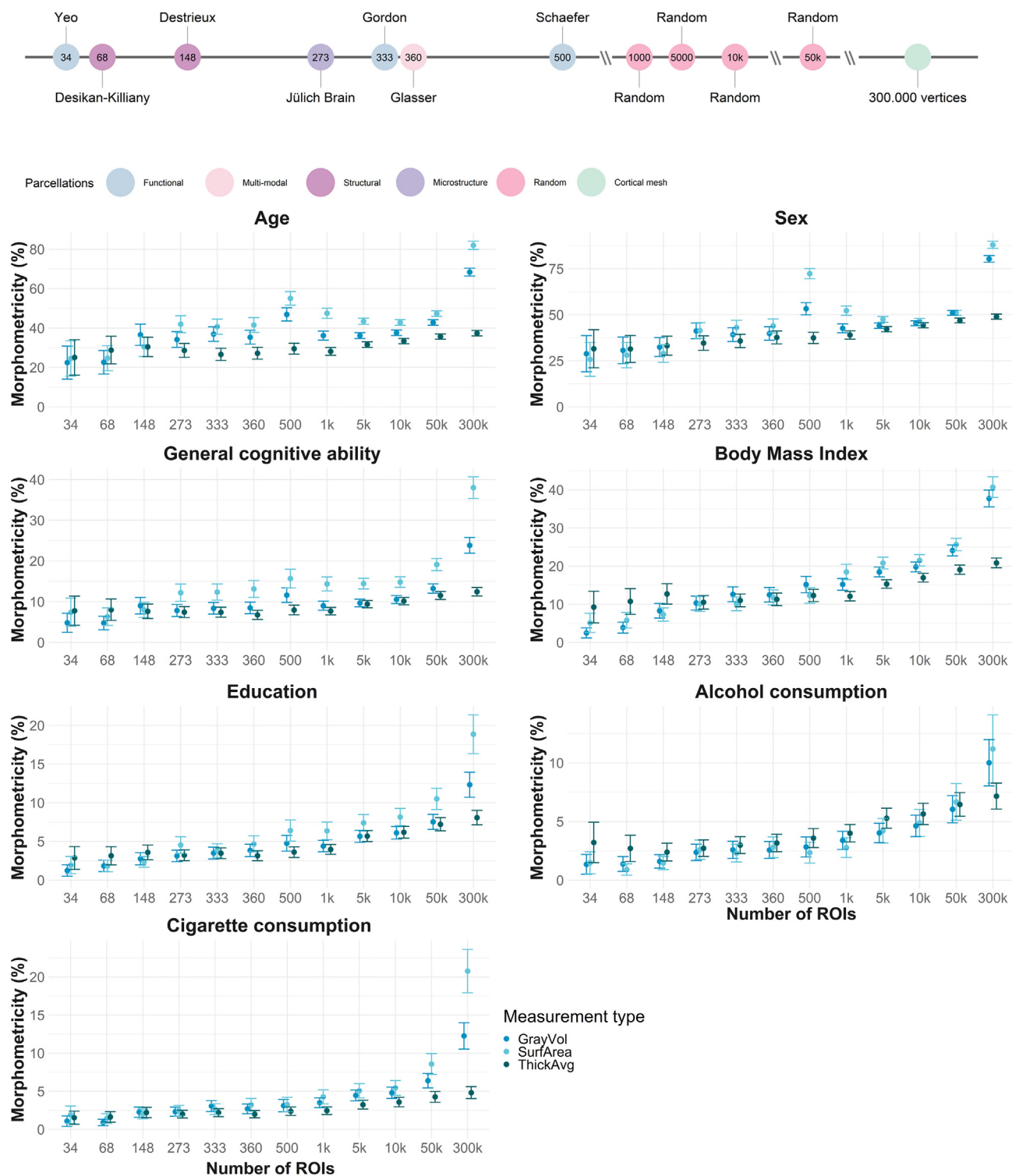
Fig. 4 — Morphometricity for seven behavioural traits estimated using different brain measurement types: GrayVol = cortical gray-matter volume, SurfArea = cortical surface area, ThickAvg = average cortical thickness.

3.3. Relationship between morphometricity and atlas dimensions

We observed strong linear-log relationships between morphometricity estimates and (logarithmically transformed) atlas dimensionality (SFig. 3), with R^2 estimates ranging between 56% and 98% (mean = 84.58%, SD = 11.89%). Table 1 reports intercepts (α) and coefficients (β) characterising this relationship for non-brain traits and brain measurement types separately. Based on information contained in Table 1, the expected morphometricity estimate (in %) may be inferred for each trait, as follows:

$$E(morphometricity) = \beta \times \log(number\ of\ ROIs) + \mathfrak{a}$$

With this formula, we can infer, for example, that the expected morphometricity for age using 200,000 surface area measurements is around 67% ($4.5 \times \log(200,000) + 12$). Table 1 shows that models of cortical thickness tended to yield smaller coefficients β than models of surface area, suggesting thickness-based morphometricity is less influenced by atlas dimensionality than surface area-based analyses.

3.4. Outlying estimates by Schaefer atlas

Morphometricity from the Schaefer atlas measures were unexpectedly large, they yielded worse model fit compared with other atlases (SFig. 2) and broke the continuity of increased estimates with larger atlas dimensionality (mainly associations with age and sex). According to our formula above, we expected the following morphometricity for an atlas with 500 ROIs: age$_{surface\ area}$ = 39.97%, age$_{volume}$ = 34.75%, sex$_{surface\ area}$ = 43.45%, sex$_{volume}$ = 40.73%. Instead, we found substantially larger estimates: age$_{surface\ area}$ = 55.01%, age$_{volume}$ = 46.90%, sex$_{surface\ area}$ = 72.30%, sex$_{volume}$ = 53.24%. Suspecting the Schaefer atlas measures may not be normally distributed, we re-estimated morphometricity using normalised Schaefer atlas data (Rank-based Inverse Normal Transformation) and obtained estimates as predicted by the formula: age$_{surface\ area}$ = 38.99% [35−42%], age$_{volume}$ = 36.46% [33−40%], sex$_{surface\ area}$ = 48.32% [45−52%], sex$_{volume}$ = 44.46% [41−48%]. Normalisation descriptively reduced morphometricity for all other traits too, but confidence intervals from adjusted and unadjusted estimates overlapped.

3.5. Atlases with random region boundaries

To understand whether commonly-used atlases explained more variance than expected using random atlas boundaries, we estimated the distributions of morphometricity estimates from 100 randomly generated atlases with the same number of ROIs as included in the commonly-used atlases. The distributions of estimates from random atlases are visualised in Fig. 5 alongside morphometricity estimated from the respective commonly-used atlas (red colours). In most cases, confidence intervals around the point morphometricity estimates from commonly-used atlases mapped well onto the spread in estimates from atlases with random parcellation (Fig. 5), suggesting that commonly-used atlases yielded estimates as expected (on average) by random parcellations. 69% of point estimates from commonly-used atlases (total of 147 estimates across seven atlases, seven traits and three brain measurement types) yielded morphometricity smaller than at least 50% of the point estimates from random atlases. 5% of commonly-used atlas estimates were larger than 99% of the null estimates, and 13% of the commonly-used estimates were smaller than any of the null estimates (mostly involving surface area; STable 4.1). The latter indicates that in some cases less variance is explained when using one of the established atlases (Desikan-Killiany, Destrieux, Gordon, Glasser, Schaefer) compared to using random parcellations when mapping surface area measures onto behavioural traits.

3.6. Atlas comparisons

We performed cross-atlas LRTs to quantify whether a more fine-grained atlas added explanatory variance in addition to

Table 1 − Linear regression results describing the relationship between morphometricity and logarithmic transformed atlas dimensionality.

Non-brain trait	Brain measurement type	$\mathfrak{a}$	β	β Standard error	p-value	R^2 (%)
Age	SurfArea	12	4.5	1	.0013	66
	GrayVol	13	3.5	.8	.0014	65
	ThickAvg	21	1.2	.17	2.8e-05	84
Alcohol consumption	SurfArea	−3.4	1	.1	2.4e-06	90
	GrayVol	−2.5	.86	.081	8.8e-07	92
	ThickAvg	.14	.56	.054	1.1e-06	92
Body Mass Index	SurfArea	−9.9	3.7	.27	8.5e-08	95
	GrayVol	−8.7	3.4	.26	1.3e-07	94
	ThickAvg	4.3	1.3	.11	2.3e-07	94
Cigarette consumption	SurfArea	−6.8	1.7	.3	.00022	76
	GrayVol	−3.4	1	.12	7.7e-06	88
	ThickAvg	.019	.38	.018	1.3e-09	98
Education	SurfArea	−5.1	1.6	.18	2.9e-06	90
	GrayVol	−2.6	1.1	.09	3.6e-07	93
	ThickAvg	.028	.64	.048	1e-07	95
General cognitive ability	SurfArea	−3.9	2.6	.46	.00019	77
	GrayVol	−1.1	1.6	.28	.00019	77
	ThickAvg	4.5	.6	.085	3.4e-05	83
Sex	SurfArea	13	4.9	1.4	.0054	56
	GrayVol	14	4.3	.8	.00032	74
	ThickAvg	24	2.1	.11	3.9e-09	97

GrayVol = cortical gray-matter volume, SurfArea = cortical surface area, ThickAvg = average cortical thickness. We observed linear-log relationships between morphometricity and atlas dimensionality which we describe through a linear regression analysis of logarithmically transformed atlas dimensionality. $\mathfrak{a}$ is the intercept and β is the regression coefficient. Expected morphometricity estimates may be calculated as follows: E($morphometricity$) = $\beta \times \log(number\ of\ ROIs) + \mathfrak{a}$

the variance already explained by a coarser atlas. For example, first, we modelled morphometricity using Desikan-Killiany (68 ROIs) and Destrieux (148 ROIs) brain measures in one model. Second, we dropped Destrieux from the model and recalculated morphometricity for Desikan-Killiany alone, which allowed quantification of the percentage of variance explained by Destrieux in addition to the variance already explained by Desikan-Killiany. The LRT quantifies whether the explanatory variance added by the higher dimensional atlas (i.e., Destrieux in this example) is larger than zero. To summarise the results, we display an index of the relative improvement made to the model by adding the higher dimensional atlas, which is calculated as the sum of the variance explained by both atlases together, divided by the morphometricity of the lower dimensional atlas alone (Fig. 6, SFig. 5). A ratio of 1 indicates no increase in morphometricity, and a ratio larger than 1 indicates the proportional increase gained by adding the higher dimensional atlas. None of the indices are below 1 because the combined morphometricity from two atlases cannot be smaller than the estimate yielded by one atlas alone.

A total of 576 LRTs (out of 588) were significant, and we found an average ratio of 2.73 for surface area, 2.70 for cortical volume, and 1.67 for cortical thickness (STable 5), indicating that the morphometricity roughly doubled when adding a second atlas to the equation. Vertex-wise measures added considerable explanatory variance in addition to all other atlases, but the greatest model improvements were gained when adding vertex-wise measures to low-dimensional atlases. Across the board, commonly-used atlases accounted for unique trait variance rather than variance shared between atlases suggesting they captured atlas-specific trait variance.

Surprisingly, some low-dimensional pairs of atlases surpassed vertex-wise estimates, for example, average thickness estimates for age were 29% [21−35%] for Desikan-Killiany alone, and 30% [25−36%] for Destrieux alone, but together Desikan-Killiany and Destrieux accounted for 63% [58−68%] which was substantially larger than the vertex-wise estimates of 37% [36−39%]. We re-calculated morphometricity and joint-atlas effects using simple linear regression (LM) (which was not pre-registered) to test whether these LMM estimates were biased, potentially due to the violation of LMM assumptions.

The rationale was that LMMs enforce a normal distribution on the ROI effects, which may be problematic for low-dimensional atlases with individual ROIs dominating and driving brain-wide effects. LM does not impose assumptions on the distribution of effects which we suggest allows to test for a violation of LMM assumptions (though LM will more likely produce inflated estimates, the more predictors are included). LM estimates are displayed in SFig. 7, confirming that LMMs are likely overestimating joint atlas effects by low-dimensional atlases. Overall, LMs validate most of the LMM results, but are likely overestimating effects by high-dimensional atlases.

Only twelve tests were not statistically significant ($p > .05/$ 588), all involving brain measures of surface area and alcohol and cigarette consumption. In these cases, adding the higher dimensional atlas to the equation did not significantly increase morphometricity. The low morphometricity estimates of the traits (<4%), probably resulted in a lower statistical power, which may, in part, explain these results.

3.7. *Sensitivity analysis: LASSO-based prediction*

We trained LASSO models to investigate whether primary association results aligned with results from a machine learning approach. In other words, we sought to investigate if a larger morphometricity also led to a larger prediction accuracy. Indeed, we found that prediction accuracy improved with atlas dimensionality (and morphometricity), with the vertex-wise representation yielding the best prediction accuracy (i.e., R^2 between observed and LASSO-predicted values). Like morphometricity, prediction accuracy was substantial for most traits, but low for education, alcohol, and cigarette consumption (SFig. 6). We observed a gain in prediction accuracy by 1.95-fold (range 1.12-fold−9.46-fold) between the coarsest (34 ROIs, Yeo) and the most fine-grained cortical representations (vertices; prediction accuracies in STable 6). For example, predicted values of cognitive ability by volumetric Yeo measures explained 7% of the variance in observed values, and vertex-wise predictions explained 13% of observed cognitive ability values.

Wilcoxon Signed-Rank tests were used to test for statistical differences in absolute errors between two atlases. Out of 588 comparisons, only 3 reached a significance level of .05/588 (STable 6). Effects sizes were overall very small suggesting that we have no evidence to conclude that one atlas yielded smaller median prediction errors than another atlas. The three significant improvements in prediction were found between the Destrieux and Glasser measurements of ThickAvg in sex ($p = 1.62 \times 10^{-5}$; $r = .049$; $N_{evaluation} = 7701$), Destrieux and Schaefer measurements of ThickAvg in sex ($p = 6.51 \times 10^{-7}$; $r = .057$; $N_{evaluation} = 7701$), and Desikan-Killiany and Yeo measures of GrayVol in alcohol consumption ($p = 1.53 \times 10^{-5}$; $r = .056$; $N_{evaluation} = 5203$).

4. Discussion

In this Registered Report, we provide a quantitative comparison of the information retained by commonly-used cortical atlases of varying dimensionality (i.e., number of considered ROIs) through reporting morphometricity estimates across seven behavioural traits. We calculated whole-brain morphometricity using linear mixed models (LMMs) and compared percentages of variance accounted for by cortical atlases using likelihood ratio tests (LRTs). As hypothesised, we found that using more fine-grained atlases to describe the cortical grey-matter structure resulted in larger morphometricity. This is consistent with our pre-registered hypothesis that lower dimensional cortical representations (i.e., fewer ROIs) tend to mask inter-individual variance, and that this variance can be retained when representing the brain using ~300,000 vertex-wise measures. We found, across all traits and types of measurements, that morphometricity increased linearly as a function of the log-atlas dimension. There was no evidence for a "tipping point" of atlas dimensionality beyond which morphometricity ceased to improve, nor was there evidence for trait-dependent optimal levels of atlas dimensionality.

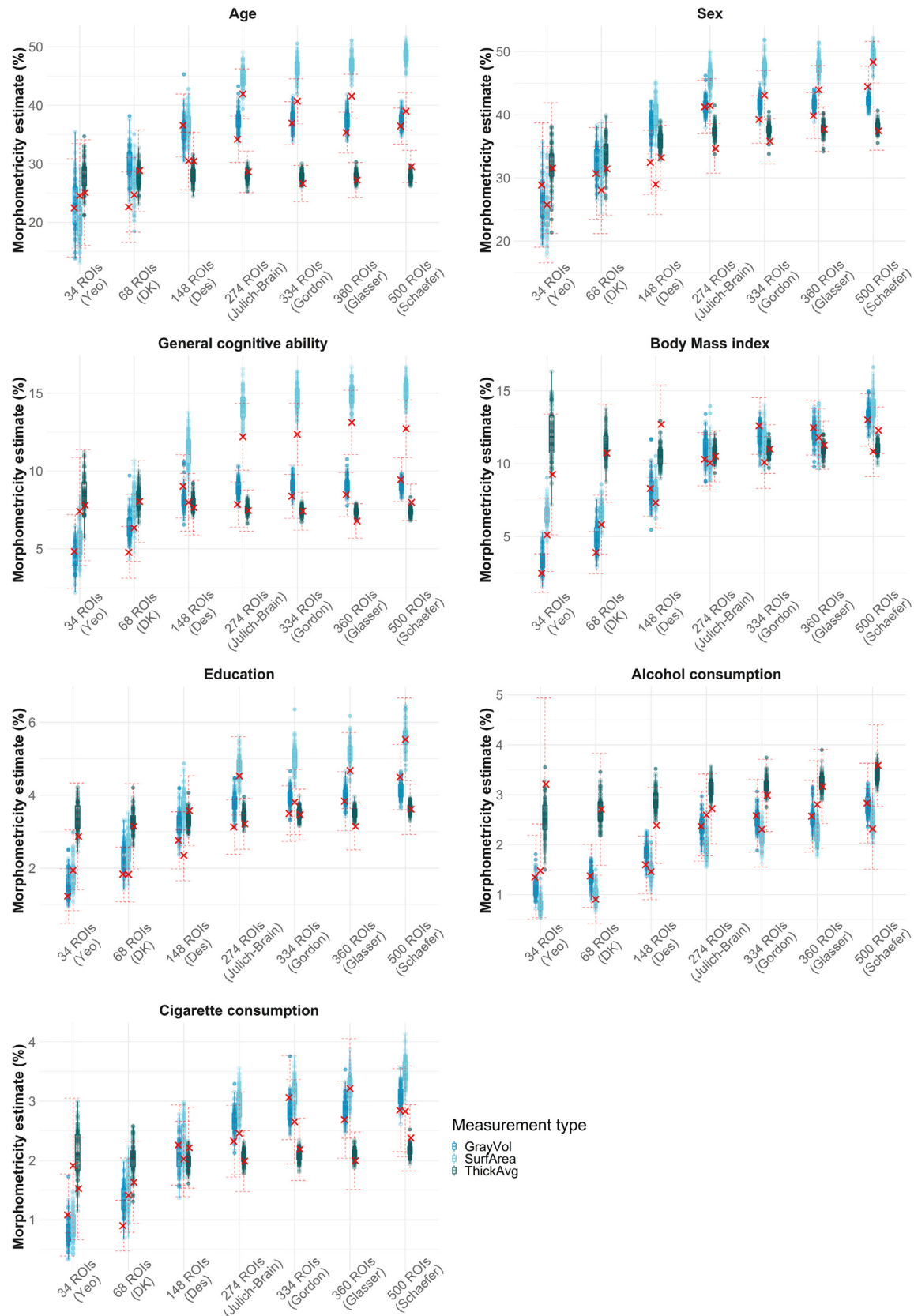Our findings should give studies of cortical structure reason to favour finer-grained cortical representations over

Fig. 5 — Distributions of morphometricity estimates from random parcellations. Each distribution was generated based on morphometricity estimates of random parcellations with matched numbers of ROIs. Red crosses indicate the corresponding estimate yielded by the commonly-used atlas; red lines indicate 95% confidence intervals around the point estimate. Random parcellations of 34 ROIs were compared to Yeo, 68 ROIs were compared to Desikan-Killiany, 148 ROIs were

coarser ones, which should enable better brain-behaviour mapping and prediction accuracy of behaviours from cortical measurements. We suggest that available neuro-imaging samples (such as the UKB) and computational resources are now large enough to accurately estimate associations between vertices and behavioural variables, promising future studies that more systematically account for brain-wide associations than can be obtained from ROI-behaviour associations (Smith & Nichols, 2018).

Should the use of lower-dimensional atlases be necessary in a study mapping interindividual differences in structural cortical measures onto behavioural traits, we would advise researchers, based on the findings presented here, to repeat analyses across multiple commonly-used atlases, or to iterate over multiple atlases with random region boundaries. This may complicate the interpretation of findings but promises more replicable brain-behaviour associations. While statistical power of these multidimensional approaches is limited by multiple testing correction, techniques to overcome this limitation have been derived. For example, one may group vertex-wise measures based on atlas ROIs and fit each set of vertices as a random effect, which permits to perform a single association test per ROI, while still modelling the fine-grained cortical structure (Couvy-Duchesne et al., 2020).

4.1. Greater atlas dimensionality yields substantially greater morphometricity estimates

Differences in morphometricity between the coarsest (Yeo, 34 ROIs) and the most fine-grained atlas (vertices, 300,000 ROIs) were considerable (median increase of 3.41-fold). One previous functional MRI study reported a doubling in prediction accuracy of cognitive ability between estimates represented by the Glasser atlas (360 ROIs) compared with vertex-wise representations (Feilong, Guntupalli, & Haxby, 2021). Though using structural neuroimaging data and an association framework, our results were comparable, showing 1.83-fold (thickness), 2.81-fold (volume) and 2.90-fold (surface area) improvements in morphometricity when using vertex-wise measurements relative to the Glasser atlas. Findings reported here are consistent with previous studies and together this indicates that selecting fine-grained cortical atlases substantially improves variance overlap between brain and behaviour. Future studies are needed to understand whether this applies to functional neuroimaging and diffusion tractography.

4.2. Commonly-used atlases yield the same morphometricity as expected from random parcellations

To assess whether the commonly-used atlases included here were superior to random parcellations, we generated 100 atlases with random ROI boundaries and, for better comparability, the same number of ROIs as in commonly-used atlases. Random parcellations were mapped onto participants' brain images and used to estimate morphometricity.

We observed that the morphometricity from commonly-used atlases was on par with the average morphometricity from random parcellations, which suggests commonly-used atlases do not maximise their potential in capturing specific trait variance. That random atlases yielded similar estimates as commonly-used atlases is in line with a recent study demonstrating that using random parcellations in predicting structure–function correlations resulted in similar power (Revell et al., 2022).

Surface area-based associations involving Desikan-Killiany, Destrieux et al. (2010), Gordon et al. (2014), Glasser et al. (2016) and Schaefer et al. (2017) produced estimates below any of the 100 random parcellations. This suggests for surface area-based associations that most random atlases would outperform the commonly-used ones. We suggest future studies should consider iterating over random parcellations to establish more robust association results. Note that this is referring to Schaefer morphometricity estimates that were corrected for non-normal distributions, and future studies should account for non-normal distributions when using Schaefer in FreeSurfer processing.

4.3. Morphometricity differs between surface area, cortical thickness, and volume

Morphometricity estimated from surface area measures were most affected by atlas dimensionality (largest linear-log regression coefficients; Table 1), which suggests that choosing fine-grained atlases will benefit studies of surface area most, compared with thickness measures, for example. The observation that atlas dimensionality has differential impact on estimates from different measurement types may fit with evidence that surface area and thickness measures have low phenotypic and genetic correlations (Panizzon et al., 2009; Wierenga, Langen, Oranje, & Durston, 2014; Winkler et al., 2010). Here, we did not specifically test for this, and future work is needed to understand whether different brain measurement types represent common or unique information. Morphometricity analyses in Couvy-Duchesne et al. (2020) suggested that surface area and thickness measures capture some common trait variance, but that they are mostly unique.

4.4. Low-dimensional atlases capture unique rather than common trait variance

We used LRTs to examine whether one of a pair of atlases outperformed the other in its overlap with behavioural traits. We found that most comparisons between atlases were significant, and that the variance accounted for was roughly doubled when jointly considering two low-dimensional atlases. This considerable improvement indicates that atlases explain unique, rather than shared trait variance. By selecting one coarse atlas to parcellate participants' brain images, researchers likely restrict their analyses to a specific dimension of variance that only partly overlaps with behavioural traits.

compared to Destrieux, 274 ROIs to Jülich-Brain, 334 to Gordon, 360 to Glasser, and 500 to Schaefer. Estimates for Schaefer were calculated based on Rank-based Inverse Normal Transformed data and are therefore not the same estimates as displayed in Fig. 4. GrayVol = gray matter volume, SurfArea = surface area, ThickAvg = average thickness.
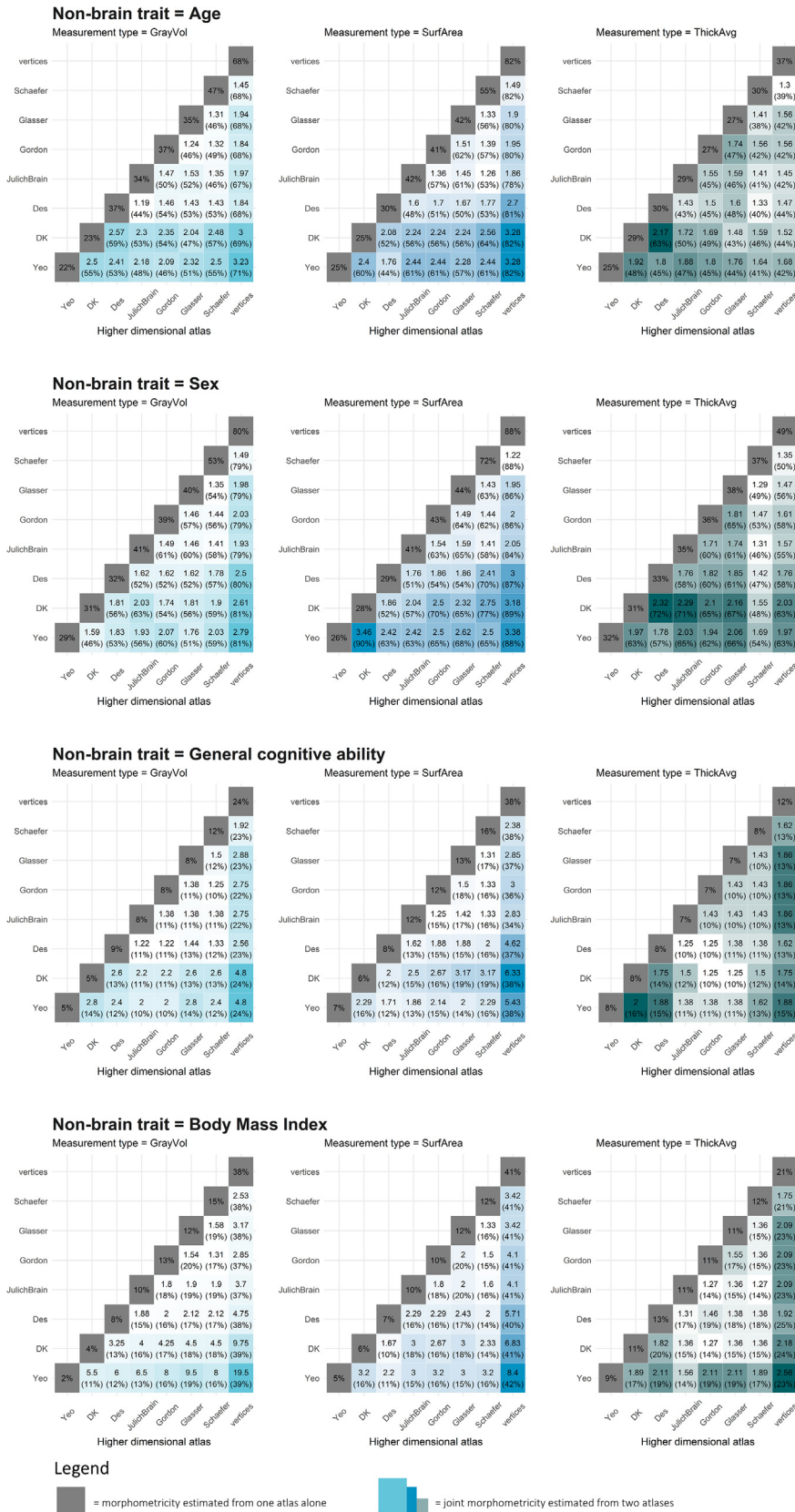
Fig. 6 — Atlas comparisons calculated from likelihood ratio tests for age, sex, cognitive ability, and body mass index. Find the equivalent visualisation for education, cigarette, and alcohol consumption in SFig. 5. Percentages displayed on the diagonal are morphometricity estimates for individual atlases, as indicated in Fig. 4. Indices on the off-diagonal show the relative improvement made to the model by adding the higher dimensional atlas, which we calculated as the sum of the

Alongside small sample sizes and large sampling variation (Genon, Eickhoff, & Kharabian, 2022; Marek et al., 2022), this is likely a contributor to highly heterogenous reports of brain-behaviour correlates and the lack of reproducibility across ROI-based studies. We demonstrate that two atlases can capture non-overlapping trait variance, which implies that results between two ROI-based studies cannot be translated if different atlases were used. Future studies by researchers who still wish to use low-dimensional atlases may yield more replicable results when repeating their analyses across multiple atlases, which may recover variance hidden by one, but not another low-dimensional atlas. This should yield more robust results but may complicate their interpretation. Future studies may parse the likely sources of brain-based signal using atlases derived from different principles (macro-anatomy, resting-state networks etc.).

4.5. LASSO-based predictions confirm morphometricity estimates

The general trend that atlas dimensionality was positively associated with increasing morphometricity was confirmed by LASSO-based predictions. This is in line with previous demonstrations that association and prediction results are related (Couvy-Duchesne et al., 2020). If unbiased and accurate, morphometricity estimated from LMMs represents the total linear association, that, in theory, LASSO-based predictions may reach if the algorithm was trained in an infinite sample allowing accurate estimates of predictor weights. Here, LASSO prediction accuracy is indeed smaller than LMMs in almost all cases. We observed few exceptions where cortical thickness-based predictions using LASSO surpassed LMM-based morphometricity (e.g., vertex-wise thickness-age: LASSO out-of-sample prediction = 59%; LMM association = 37%). This may suggest that the cortical thickness LMM estimates are downward biased, maybe due to violation of a normal distribution in effect sizes for traits like age and sex. It is known, for example, that different areas of the brain age unevenly across the life-span (Raz, Ghisletta, Rodrigue, Kennedy, & Lindenberger, 2010), which may introduce a skewed distribution of effect sizes.

4.6. Parallels between multi-dimensional neuroimaging and genetic analyses

The notion that morphometricity marks the maximum trait variance that surface-based structural brain measures could account for is analogous to narrow-range heritability. Both morphometricity and heritability have the same statistical definition. In parallel with the 'First Law' in behavioural genetics stating that all traits are heritable (Turkheimer, 2000), we found that all traits considered here were morphometric. Note that not all traits were morphometric in Couvy-Duchesne et al. (2020), but they used a considerably smaller

sample (N~9,000). Using statistical genetics techniques to analyse neuroimaging data promises exciting avenues for future research that could generate innovative insights using statistical techniques that have already been derived and thoroughly tested using genetic data (Couvy-Duchesne et al., 2022).

We found exceptions in the analogy with heritability, where the morphometricity estimated by vertex-wise measures was surpassed by two joint low-dimensional atlases (prediction of age and sex by cortical thickness). Post-hoc analyses suggested that model assumptions imposed by LMMs were violated by low-dimensional atlases including individual ROIs with large effects. We suggest that the parallel between heritability and morphometricity is most consistent when considering vertex-wise data, as opposed to coarse cortical atlases, however, results must be interpreted in their specific genetic or neuroimaging context (e.g., genetic markers are stable across the lifespan, while brain structure changes in response to environments and behaviours).

4.7. Limitations

We emphasise that LMMs employed here fit all predictor variables as a single random effect, no matter how many predictor variables are included by an atlas. This makes the variance accounted for comparable between atlases, and it does not automatically increase with the number of predictors, as it would, for example, in a simple regression model. However, reported estimates are likely confounded (for example by disease or environmental influences), but this should be constant across atlases and should not impede a fair comparison between atlases. Some LMMs of joint morphometricity by low-dimensional atlases are likely overestimated due to individual ROIs driving brain-wide effects, but we demonstrate that we can test LMM assumptions using linear regression in this low-dimensional space. Furthermore, some categorical behavioural variables used in this study may be suboptimal. Cigarette and alcohol consumption categories were based on arbitrary cut-offs, and the education phenotype was dominated by the category capturing university education. Low levels of trait morphometricity may reflect these caveats and may require re-evaluation in other samples.

It is unknown whether our results are directly translatable to analyses using different software and pipelines. For reproducibility, we make all our analysis code available online (https://annafurtjes.github.io/Comparing_atlases/). Study results would likely differ across FreeSurfer versions (Gronenschild et al., 2012; we used v6.0.0), and between processing software, especially compared with other approaches than surface-based analyses (e.g., volumetric analyses in Statistical Parametric Mapping, SPM, or FMRIB Software Library, FSL). Future studies are needed to test whether results would replicate in functional MRI data, and

---

variance explained by two atlases together, divided by the morphometricity estimate of the lower dimensional atlas alone (joint$_{morphometricity}$/individual$_{morphometricity}$). **Squares are coloured according to this ratio, i.e., larger ratio, darker colour. The raw sum of variance explained is printed in brackets below the respective index. Non-significant results are marked with *n.s.***

whether more elaborate means of summarising vertex-wise measures across ROIs could provide better cortical representation (e.g., taking the maximum thickness within an ROI).

5. Conclusions

This Registered Report demonstrates the importance of appropriate cortical atlas choices across neuroimaging studies by demonstrating that using commonly-used cortical atlases resulted in considerable loss of information compared to vertex-wise representations of cortical structure. These atlases included Yeo et al. (2011), Desikan-Killiany (Desikan et al., 2006), Destrieux et al. (2010), Jülich-Brain (Amunts et al., 2020), Gordon et al. (2014), Glasser et al. (2016), and Schaefer et al. (2017), which all accounted for magnitudes of variance on par with random parcellations across seven behavioural traits using cortical thickness, volume, and surface area measures. In the interest of more replicable results, our findings should give researchers reason to leverage large-scale samples (like the UKB) to conduct association and prediction analyses using vertex-wise brain data which promises more systematic and localised accounts of the relationships between behaviours and structural measures of the cortex.

We further demonstrate that studies using one coarse atlas only (for example, Yeo, Desikan-Killiany, or Destrieux) tend to capture atlas-specific trait variance, implying that study results often cannot be translated between atlases. Hence, atlas choice is likely a contributor to the lack in reproducibility in the neuroimaging literature. We suggest that studies for which the use of coarse atlases is necessary, should either repeat analyses across multiple commonly-used atlases, or iterate over random atlases to produce more robust results.

5.1. Statement of transparency for secondary analyses

AEF is the lead analyst on this project and has been granted access to the UKB data through UKB application 40933, of which JHC is the principal investigator. AEF previously worked with pre-processed Desikan-Killiany IDPs through another UKB application, but has not seen or worked with the bulk MRI data used in this project, or any of the behavioural traits. She therefore is naïve to any potential associations between variables considered in this study. The UKB application used here is unrelated to the one used in Couvy-Duchesne et al. (2020).

An application to download the required data was submitted to UKB on the 28th of March 2021. It has been approved on the 11th of May, neuroimaging data download occurred successively between November 2021 and January 2022 (our pipeline downloaded individual-by-individual, processed, and deleted data to keep cluster storage on King's College London's high performance computer *Rosalind* at a minimum). AEF had not downloaded or investigated any of the data until in-principle acceptance had been granted in November 2021. The analysis plan pre-registered as a Registered Report (https://osf.io/dkw9t), and analysis code was made available on GitHub (https://annafurtjes.github.io/Comparing_atlases/).

Credit author statement

Anna E. Fürtjes: Conceptualization, Formal analysis, Investigation, Project administration, Software, Visualization, Writing - original draft.

James H. Cole: Conceptualization, Resources, Supervision.

Baptiste Couvy-Duchesne: Conceptualization, Methodology, Supervision, Software.

Stuart J. Ritchie: Conceptualization, Resources, Supervision.

The final review and editing will be done by all authors.

Open practices

The study in this article earned Open Materials e Protected Access and Preregistered badges for transparent practices. Materials and data for the study are available at https://osf.io/dkw9t.

Declaration of competing interest

None.

Data availability

The authors do not have permission to share data.

Acknowledgements

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cortex.2022.11.001.

REFERENCES

Alexander-Bloch, A. F., Shou, H., Liu, S., Satterthwaite, T. D., Glahn, D. C., Shinohara, R. T., … Raznahan, A. (2018). On testing for spatial correspondence between maps of human brain structure and function. *Neuroimage, 178*, 540—551. https://doi.org/10.1016/j.neuroimage.2018.05.070

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. https://doi.org/10.1016/j.neuroimage.2017.10.034

Amunts, K., Mohlberg, H., Bludau, S., & Zilles, K. (2020). Julich-brain: A 3d probabilistic atlas of the human brain's cytoarchitecture. *Science, 369*(6506), 988—992. https://doi.org/10.1126/science.abb4588

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*(9), Article e7200. https://doi.org/10.1371/journal.pone.0007200

Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., … Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry, 23*(5), 1385—1392. https://doi.org/10.1038/mp.2017.62

Couvy-Duchesne, B., Strike, L. T., Zhang, F., Holtz, Y., Zheng, Z., Kemper, K. E., … Visscher, P. M. (2020). A unified framework for association and prediction from vertex-wise grey-matter structure. *Human Brain Mapping, 41*(14), 4062—4076. https://doi.org/10.1002/hbm.25109

Couvy-Duchesne, B., Zhang, F., Kemper, K., Sidorenko, J., Wray, N., Visscher, P., … Yang, J. (2022). Parsimonious model for mass-univariate vertexwise analysis. *The Journal of Medical Investigation: JMI, 9*(5), Article 052404. Retrieved from 10.1117/1.JMI.9.5.052404.

Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M., & Deary, I. J. (2019). Structural brain imaging correlates of general intelligence in Uk Biobank. *Intelligence, 76*, Article 101376. https://doi.org/10.1016/j.intell.2019.101376

Dadi, K., Varoquaux, G., Machlouzarides-Shalit, A., Gorgolewski, K. J., Wassermann, D., Thirion, B., et al. (2020). Fine-grain atlases of functional modes for fmri analysis. *Neuroimage, 221*, Article 117126. https://doi.org/10.1016/j.neuroimage.2020.117126

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience, 11*(3), 201—211. https://doi.org/10.1038/nrn2793

de la Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M., & Deary, I. J. (2021). A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nature Human Behaviour, 5*(1), 49—58. https://doi.org/10.1038/s41562-020-00936-2

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Hyman, B. T. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage, 31*(3), 968—980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage, 53*(1), 1—15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and validity of the Uk Biobank cognitive tests. *Plos One, 15*(4), Article e0231627. https://doi.org/10.1371/journal.pone.0231627

Feilong, M., Guntupalli, J. S., & Haxby, J. V. (2021). The neural basis of intelligence in fine-grained cortical topographies. *eLife, 10*, Article e64058. https://doi.org/10.7554/eLife.64058

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., … Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron, 33*(3), 341—355. https://doi.org/10.1016/S0896-6273(02)00569-X

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

Fry, A., Littlejohns, T. J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., … Allen, N. E. (2017). Comparison of sociodemographic and health-related characteristics of Uk Biobank participants with those of the general population. *American Journal of Epidemiology, 186*(9), 1026—1034. https://doi.org/10.1093/aje/kwx246

Genon, S., Eickhoff, S. B., & Kharabian, S. (2022). Linking interindividual variability in brain structure to behaviour. *Nature Reviews Neuroscience, 23*(5), 307—318. https://doi.org/10.1038/s41583-022-00584-7

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., … Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature, 536*(7615), 171—178. https://doi.org/10.1038/nature18933

Gong, W., Beckmann, C. F., & Smith, S. M. (2021). Phenotype discovery from population brain imaging. *Medical Image Analysis, 71*, Article 102050. https://doi.org/10.1016/j.media.2021.102050

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2014). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex, 26*(1), 288—303. https://doi.org/10.1093/cercor/bhu239

Gronenschild, E. H. B. M., Habets, P., Jacobs, H. I. L., Mengelers, R., Rozendaal, N., van Os, J., et al. (2012). The effects of freesurfer version, workstation type, and macintosh operating system version on anatomical volume and cortical thickness measurements. *Plos One, 7*(6), Article e38234. https://doi.org/10.1371/journal.pone.0038234

Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics, 30*(19), 2811—2812. https://doi.org/10.1093/bioinformatics/btu393

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods, 3*(4), 424. https://doi.org/10.1037/1082-989X.3.4.424

Kharabian Masouleh, S., Eickhoff, S. B., Hoffstaedter, F., Genon, S., & Alzheimer's Disease Neuroimaging Initiative. (2019). Empirical examination of the replicability of associations between brain structure and psychological variables. *eLife, 8*, Article e43464. https://doi.org/10.7554/eLife.43464

Lindroth, H., Nair, V. A., Stanfield, C., Casey, C., Mohanty, R., Wayer, D., … Sanders, R. D. (2019). Examining the identification of age-related atrophy between T1 and T1 + T2-flair cortical thickness measurements. *Scientific Reports, 9*(1), Article 11288. https://doi.org/10.1038/s41598-019-47294-2

Littlejohns, T. J., Hollìday, J., Gibson, L. M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., … Allen, N. E. (2020). The Uk Biobank imaging enhancement of 100,000 participants: Rationale, data collection, management and future directions. *Nature Communications, 11*(1), 2624. https://doi.org/10.1038/s41467-020-15948-9

Mangin, J. F., Rivière, D., & Amunts, K. (2021). *Surface projections of julich-brain cytoarchitectonic maps*. EBRAINS. https://doi.org/10.25493/NZGY-6AS (V2.9) [Data Set].

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., … Dosenbach, N. U. F. (2022).

Reproducible brain-wide association studies require thousands of individuals. *Nature, 603*(7902), 654−660. https://doi.org/10.1038/s41586-022-04492-9

Mowinckel, A. M., & Vidal-Piñeiro, D. (2020). Visualization of brain statistics with R packages ggseg and Ggseg3d. *Advances in Methods and Practices in Psychological Science, 3*(4), 466−483. https://doi.org/10.1177/2515245920928009

Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., … Kremen, W. S. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cerebral Cortex, 19*(11), 2728−2735. https://doi.org/10.1093/cercor/bhp026

Prive, F., Aschard, H., Ziyatdinov, A., & Blum, M. G. B. (2018). Efficient analysis of large-scale genome-wide data with two R packages: Bigstatsr and bigsnpr. *Bioinformatics, 34*(16), 2781−2787. https://doi.org/10.1093/bioinformatics/bty185

Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M., & Lindenberger, U. (2010). Trajectories of brain aging in middle-aged and older adults: Regional and individual differences. *Neuroimage, 51*(2), 501−511. https://doi.org/10.1016/j.neuroimage.2010.03.020

Revell, A. Y., Silva, A. B., Arnold, T. C., Stein, J. M., Das, S. R., Shinohara, R. T., … Davis, K. A. (2022). A framework for brain atlases: Lessons from seizure dynamics. *Neuroimage, 254*, Article 118986. https://doi.org/10.1016/j.neuroimage.2022.118986

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., 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*(8), 2959−2975. https://doi.org/10.1093/cercor/bhy109

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1−36. https://doi.org/10.18637/jss.v048.i02

Sabuncu, M. R., Ge, T., Holmes, A. J., Smoller, J. W., Buckner, R. L., & Fischl, B. (2016). Morphometricity as a measure of the neuroanatomical signature of a trait. *Proceedings of the National Academy of Sciences, 113*(39), E5749−E5756. https://doi.org/10.1073/pnas.1604378113

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., … Yeo, B. T. T. (2017). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral Cortex, 28*(9), 3095−3114. https://doi.org/10.1093/cercor/bhx179

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461−464, 464. Retrieved from 10.1214/aos/1176344136.

Smith, S., Alfaro-Almagro, F., & Miller, K. (2020). *UK Biobank brain imaging documentation*. Retrieved from https://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/brain_mri.pdf.

Smith, S. M., & Nichols, T. E. (2018). Statistical challenges in "big data" human neuroimaging. *Neuron, 97*(2), 263−268. https://doi.org/10.1016/j.neuron.2017.12.018

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., … Collins, R. (2015). Uk Biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine, 12*(3). https://doi.org/10.1371/journal.pmed.1001779. e1001779-e1001779.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267−288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science, 9*(5), 160−164. https://doi.org/10.1111/1467-8721.00084

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *Neuroimage, 87*, 120−126. https://doi.org/10.1016/j.neuroimage.2013.11.010

Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., … Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *Neuroimage, 53*(3), 1135−1146. https://doi.org/10.1016/j.neuroimage.2009.12.028

Yaakub, S. N., Heckemann, R. A., Keller, S. S., McGinnity, C. J., Weber, B., & Hammers, A. (2020). On brain atlas choice and automatic segmentation methods: A comparison of maper & freesurfer using three atlas databases. *Scientific Reports, 10*(1), 2837. https://doi.org/10.1038/s41598-020-57951-6

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). Gcta: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics, 88*(1), 76−82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*(3), 1125−1165. https://doi.org/10.1152/jn.00338.2011

Zhang, F., Chen, W., Zhu, Z., Zhang, Q., Nabais, M. F., Qi, T., … Yang, J. (2019). Osca: A tool for omic-data-based complex trait analysis. *Genome Biology, 20*(1), 107. https://doi.org/10.1186/s13059-019-1718-z