Test–retest reliability of brain morphology estimates

Christopher R. Madan · Elizabeth A. Kensinger

Abstract Metrics of brain morphology are increasingly being used to examine inter-individual differences, making it important to evaluate the reliability of these structural measures. Here we used two open-access datasets to assess the intersession reliability of three cortical measures (thickness, gyrification, and fractal dimensionality) and two subcortical measures (volume and fractal dimensionality). Reliability was generally good, particularly with the gyrification and fractal dimensionality measures. One dataset used a sequence previously optimized for brain morphology analyses and had particularly high reliability. Examining the reliability of morphological measures is critical before the measures can be validly used to investigate inter-individual differences.

Keywords Cortical structure · Subcortical · Reliability · Fractal dimensionality · Cortical thickness · Gyrification · Structural complexity

1 Introduction

A growing number of studies have investigated relationships between brain morphology and inter-individual differences. An important assumption that underlies these studies is that estimates of brain morphology are reliable. While numerous studies have investigated the test–retest reliability for estimates of cortical thickness (e.g., [1–7]) and subcortical volume (e.g., [7–12]), the reliability of other measures of brain morphology has been less

C. R. Madan (✉) · E. A. Kensinger
Department of Psychology, Boston College, McGuinn 300, 140 Commonwealth Ave., Chestnut Hill, MA 02467, USA
e-mail: madanc@bc.edu

established and is an important topic of future research [13]. Here we measured the reliability of several measures of cortical and subcortical structures; in addition to cortical thickness and subcortical volume, we examined the reliability of estimates of cortical gyrification and fractal dimensionality.

Gyrification index is a measure of the ratio between the surface area of the cortex, relative to a simulated enclosing surface that surrounds the cortex (e.g., [14–18]). Generally, gyrification has been suggested to be an important characteristic of the human brain [15–19]. In addition to the well-known differences in cortical thickness associated with age, gyrification also differs with age [20–22]; however, age-related differences in gyrification appear to have a distinct topological distribution than thickness [20, 21]. Gyrification has also been associated with a myriad of other inter-individual measures, as reviewed by Mietchen and Gaser [14].

Structural complexity is measured as fractal dimensionality, which uses fractal geometry principles [23] to measure the complexity of brain structures (see [21]). We recently demonstrated robust age differences in the structural complexity of cortical [21] and subcortical structures [24]. Less work has been done examining the relationship between inter-individual differences and variance in complexity of cortical and subcortical regions; however, these approaches have been found to be useful in a variety of disciplines within neuroscience [25, 26].

Here we examined the test–retest reliability of several measures of brain morphology. While volumetric measures—cortical thickness and subcortical volume—have been evaluated previously, we additionally evaluated the reliability of shape-related measures, specifically gyrification and fractal dimensionality. We evaluated the

FreeSurfer implementation of gyrification, as implemented by Schaer et al. [27]. This approach generates an enclosing surface around each hemisphere and computes the 'local' difference in surface between this surface and the pial surface of the cortex. As such, gyrification is highest over the insula and lowest over medial cortical regions. Fractal dimensionality was evaluated based on the calcFD toolbox [21], which computes fractal dimensionality using intermediate files generated as part of the standard FreeSurfer pipeline. Madan and Kensinger [21] previously compared different algorithms for calculating fractal dimensionality using simulated 3D structures, but here we instead used multiple anatomical volumes acquired from the same participant (i.e., test–retest reliability).

Structural measurements are often used to assess longitudinal changes or inter-individual differences. For instance, advancements in measuring relationships between brain morphology and inter-individual differences have become increasingly relevant as a complementary approach to fMRI, due to aging-related confounds in group comparisons [28]. More recently, age-related differences have been identified in BOLD signal variability [29, 30], which may be related to differences in cerebrovascular reactivity [31, 32]. As brain morphology research advances, it is critical to measure the reliability of these metrics using multiple volume acquisitions. For instance, if the effect of age on a morphological measure is small, poorer reliability may make the effect difficult to detect due to noise in the measure. A number of open-access databases include multiple scans on the same participants, enabling such reliability to be calculated. Appendix 1 summarizes a number of additional open-access datasets—in addition to those we consider here—that also include intersession test–retest reliability data.

Here we examined test–retest reliability from two open-access datasets in which participants were scanned several times over a short interval (i.e., intersession, intrascanner). In the first dataset, 30 participants were scanned 10 times within a 1-month period [33]. In the original work, Chen et al. sought to estimate test–retest reliability of resting-state networks across intra- and inter-individual variability of six rs-fMRI measures (CCBD [Center for Cognition and Brain Disorders] dataset). In the second dataset, 69 participants were scanned twice within a 6-month period [34]. Holmes et al. collected data for a large-scale exploration ($N = 1570$) of the relations among brain function, behavior, and genetics (GSP [Brain Genomics Superstruct Project] dataset). As one demonstration of the uses of this dataset, Holmes et al. [3] examined the relationship between cortical thickness and several measures of cognitive control.

In each of these datasets, we examined the reliability of three cortical measures: cortical thickness, gyrification, and fractal dimensionality—both of the entire cortical ribbon and across regional measures of parcellated cortex (62 regions, based on the DKT atlas; [35]). We additionally evaluated different approaches to calculating fractal dimensionality to establish the reliability of each of these approaches. Finally, reliability of volume and fractal dimensionality of segmented subcortical and ventricular structures also was evaluated. We consider each dataset separately, as would be the typical approach for examining test–retest reliability, and then discuss the conclusions reached using both datasets in the general discussion.

2 Study 1: CCBD

2.1 Procedure

2.1.1 Dataset

MR images were acquired using a GE MR750 3 T scanner at the Centre for Cognition and Brain Disorders (CCBD) at Hangzhou Normal University [33]. Thirty participants aged 20–30 years old were each scanned for 10 sessions, occurring 2–3 days apart over a 1-month period. T_1-weighted data were acquired using a FSPGR sequence (TR: 8.06 s; TE: 3.1 ms; flip angle: 8°; voxel size: $1.0 \times 1.0 \times 1.0$ mm). This dataset is included as part of the Consortium for Reliability and Reproducibility (CoRR; [36]) as HNU1.

2.1.2 Preprocessing of the structural data

Data were analyzed using FreeSurfer 5.3.0 (https://surfer.nmr.mgh.harvard.edu) on a machine running CentOS 6.6. FreeSurfer was used to automatically volumetrically segment and parcellate cortical and subcortical structures from the T_1-weighted images [37–40]. FreeSurfer's standard pipeline was used (i.e., recon-all). No manual edits were made to the surface meshes, but surfaces were visually inspected.

Cortical thickness is calculated as the distance between the white matter surface (white–gray interface) and pial surface (gray–CSF interface) [38]. Thickness estimates have previously been found to be in agreement with manual measurements from MRI images [41, 42], as well as ex vivo tissue measurements [43, 44]. Subcortical volume estimates have also been found to correspond well with manual segmentation protocols, particularly in young adults [45–52].

Gyrification was also calculated using FreeSurfer, as described in Schaer et al. [27]. Cortical regions were delineated based on the Desikan–Killiany–Tourville (DKT) atlas, also part of the standard FreeSurfer analysis pipeline

[35]. Intracranial volume (ICV) was also calculated using FreeSurfer [53].

Fractal dimensionality was quantified using the calcFD toolbox (http://cmadan.github.io/calcFD/), which we previously developed and distribute freely [21, 24]. calcFD is a MATLAB toolbox that calculates the fractal dimensionality of 3D structures and was developed to work with intermediate files from the standard FreeSurfer pipeline. Apart from when otherwise stated, FD was calculated for filled structures (FD_f) using the dilation algorithm. Here we additionally modified calcFD in two ways. First, we improved it to additionally calculate the fractal dimensionality of cortical parcellations for all regions delineated in the DKT atlas (see Appendix 2). An important consideration in decreasing the size of cortical parcellations, however, is that they inherently have decreased fractal dimensionality, i.e., becoming closer to a 'truncated rectangular pyramid.' Second, we adjusted the toolbox to calculate fractal dimensionality using the spherical harmonics (e.g., [54–58]). Additional details about this spherical harmonics approach are outlined in Appendix 3.

2.1.3 Measuring reliability

Reliability was calculated as intraclass correlation coefficient (ICC), which can be used to quantify the relationship between multiple measurements [59–62]. McGraw and Wong [63] provide a comprehensive review of the various ICC formulas and their applicability to different research questions. ICC was calculated as the one-way random effects model for the consistency of single measurements, i.e., ICC(1). As a general guideline, ICC values between .75 and 1.00 are considered 'excellent,' .60–.74 is 'good,' .40–.59 is 'fair,' and below .40 is 'poor' [64]. For the cortical parcellated regions, distributions of mean reliability measures (e.g., lower panel of Fig. 4) were compared using a Mann–Whitney U test, a nonparametric for testing whether two sets of values belong to the same distribution.

In the current study, we focused on regional estimates of brain morphology; a complimentary approach that we did not evaluate here is the reliability in spatial segmentation. This alternative approach evaluates the volumetric overlap between 3D structures within the same space, often quantified as a Dice coefficient (e.g., [5, 10, 48, 50, 65]). This overlap approach is often used when comparing manual and automatic segmentation protocols of the same anatomical volume; however, it can be applied to test–retest reliability by co-registering the individual anatomical volumes from the same participant to each other and comparing the resulting segmented structures' overlap. In contrast, the present goal was to evaluate 'summary statistics' of the structures, such as thickness, volume, and fractal dimensionality.

2.2 Results

2.2.1 Cortical ribbon

We first examined the test–retest reliability of cortical thickness and gyrification, as shown in Fig. 1 and Table 1. Across both measures, estimates clustered closely for all scans from the same individual. This qualitative finding was corroborated by high ICC values, .816 and .945 for thickness and gyrification, respectively.

Fractal dimensionality We computed the reliability of five calculations of fractal dimensionality. First, we used both the dilation and box-counting algorithms, as implemented in the calcFD toolbox, for both the filled volumes and surfaces only. We additionally used a spherical harmonics (SPHARM) approach (surface only). See Appendix 3 for further details about calculating fractal dimensionality using spherical harmonics. Figure 1 shows estimates of fractal dimensionality based on the dilation-filled approach.

As shown in Table 1, we consistently found higher reliability for the dilation algorithm than the box-counting algorithm, though this difference was not statistically significant. We found higher reliability for the spherical harmonics approach; however, this approach can only be used for surfaces of structures (rather than filled volumes).

2.2.2 Cortical parcellations

Mean regional cortical thickness was highest in lateral temporal regions, followed by frontal regions (Fig. 2). This pattern is consistent with prior findings (e.g., [20, 21, 38, 66, 67]). Regional thickness estimates were highly consistent across regions, as shown by the low mean deviation (between scans) for each region in Fig. 2. ICC values for each region are shown in Figs. 3 and 4. Regions with the greatest intersession variability are convergent with prior reliability analyses (see [2] (Fig. 2), [3] (Fig. 1), [4] (Fig. 3), [6] (Fig. 1)). Generally, thickness estimates are less reliable around the temporal pole and would be most affected in the inferior temporal gyrus using the DKT parcellation scheme, and the anterior and medial cingulate. Thickness estimates are often highest in parietal (particularly superior parietal) and occipital cortices. Nonetheless, despite the spatial variability in thickness reliability, mean deviations are often small in magnitude, often around .10 mm (Fig. 2) (see [2] (Fig. 2)).

As expected (as in [15]), gyrification was highest in the insula and lowest over medial cortical regions (Fig. 2). Beyond this, we additionally observed greater gyrification over parietal regions, convergent with prior studies (e.g., [20, 21]). Test–retest reliability of regional gyrification was generally quite high (Figs. 3, 4) and was significantly
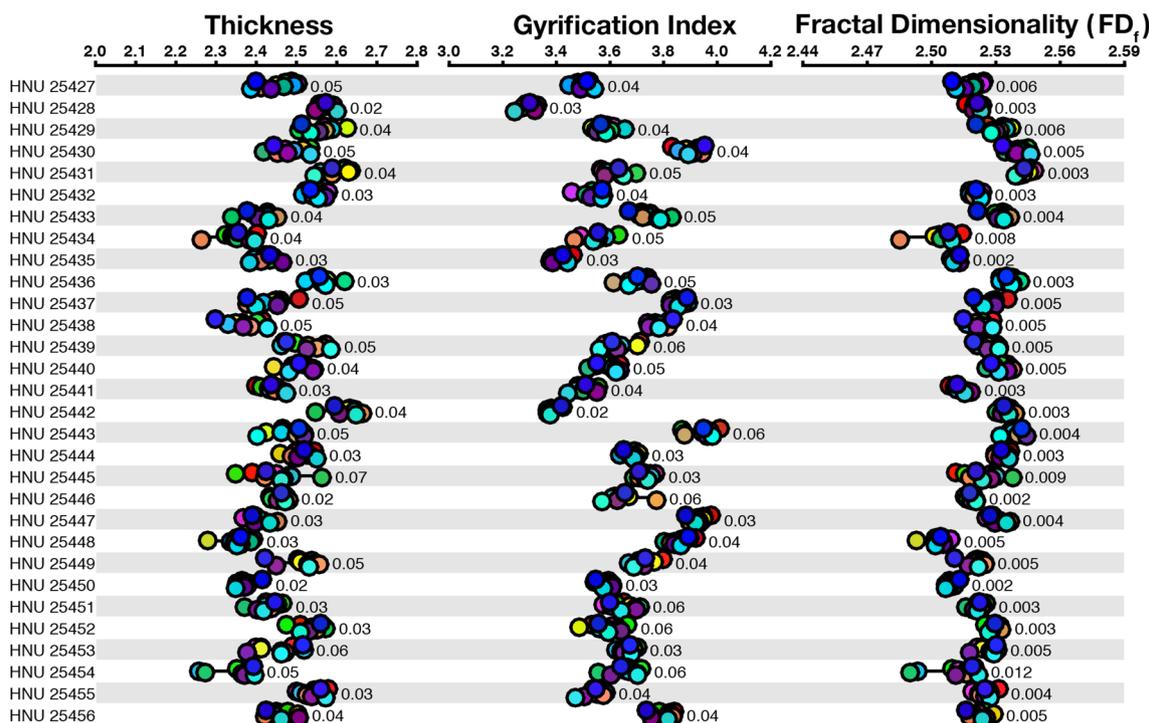
Fig. 1 Dot plot for the structural estimates for each measure for the cortical ribbon, for the CCBD dataset. Participant labels are presented on the left, such that *each row* represents structural metrics for a single participant. Each *dot* within a measure (e.g., 'Thickness') represents a different scan volume. Within each row, markers in the same *color* denote measures taken from the same scan volume. Values beside each set of markers denote the mean deviation between estimates. (Color figure online)

Table 1 Test–retest reliability (ICC) for each measure and dataset, for the cortical ribbon data

Measure	Study 1 CCBD	Study 2 GSP
Thickness (CT)	.816	.890
Gyrification (GI)	.945	.941
Fractal dimensionality		
Dilation filled (FD$_f$)	.842	.936
Dilation surface	.845	.936
Boxcount filled	.799	.879
Boxcount surface	.769	.849
SPHARM surface	.977	.982

SPHARM refers to spherical harmonics. When not otherwise stated, FD$_f$ represents FD as calculated using the dilation-filled approach

higher for gyrification than cortical thickness [$Z = 5.98$, $p < .001$].

Regional fractal dimensionality is shown in Fig. 2. Smaller regions had lower fractal dimensionality, as smaller segmented structures inherently have less structural complexity due to both limitations MRI acquisition precision and biological constraints (also see [24]). Intraclass correlations (ICCs) are shown for each structural measure and brain region in Fig. 3; Fig. 4 shows the 95%

confidence intervals of the inter-class correlations (ICCs) for each measure and region. Across regions, mean ICC was not significantly related to the size of the region for any of the measures [thickness: $r(60) = .206$, $p = .11$; gyrification: $r(60) = .154$, $p = .23$; fractal dimensionality: $r(60) = .251$, $p = .05$]. Test–retest reliability of regional fractal dimensionality was generally high (Figs. 3, 4) and was also significantly higher than for cortical thickness [$Z = 5.46$, $p < .001$]. Reliability did not differ between gyrification and fractal dimensionality [$Z = .31$, $p = .75$].

2.2.3 Subcortical structures

Test–retest reliability was relatively high for most structures and was quite similar for both volume and fractal dimensionality (Fig. 5). Reliability was lowest for the hippocampus; reliability was the highest for the caudate, putamen, and thalamus. Reliability estimates were significantly higher for the ventricles than the subcortical structures.

2.2.4 Summary

The results indicate that gyrification and fractal dimensionality have high test–retest reliability. Indeed, reliability using these measures was higher than for cortical thickness.

Fig. 2 Mean regional morphology measures for each parcellated region plotted on inflated surfaces, for the CCBD dataset
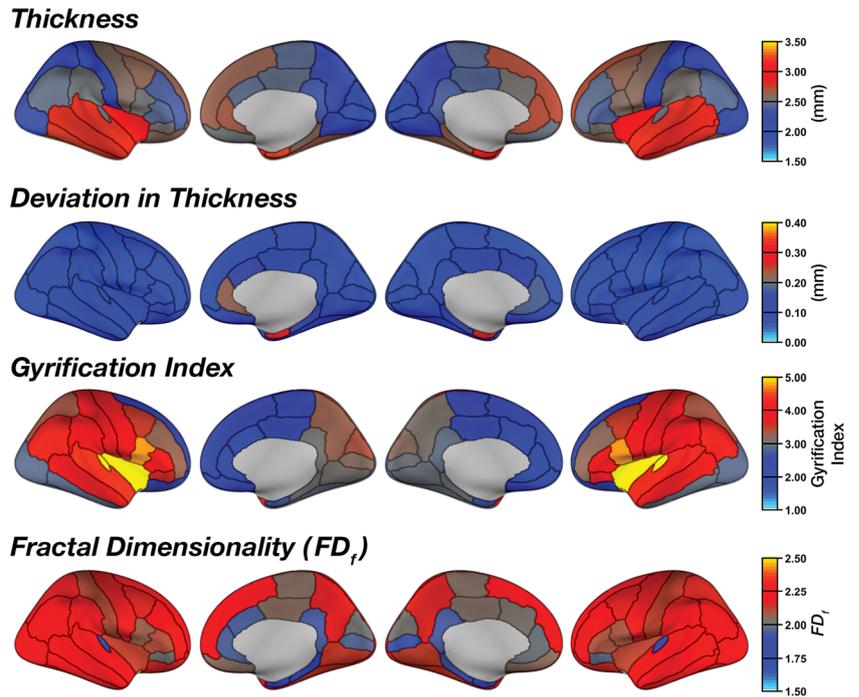
Fig. 3 Test–retest reliability (ICC) for cortical thickness, gyrification, and fractal dimensionality of the cortical parcellations, for the CCBD dataset
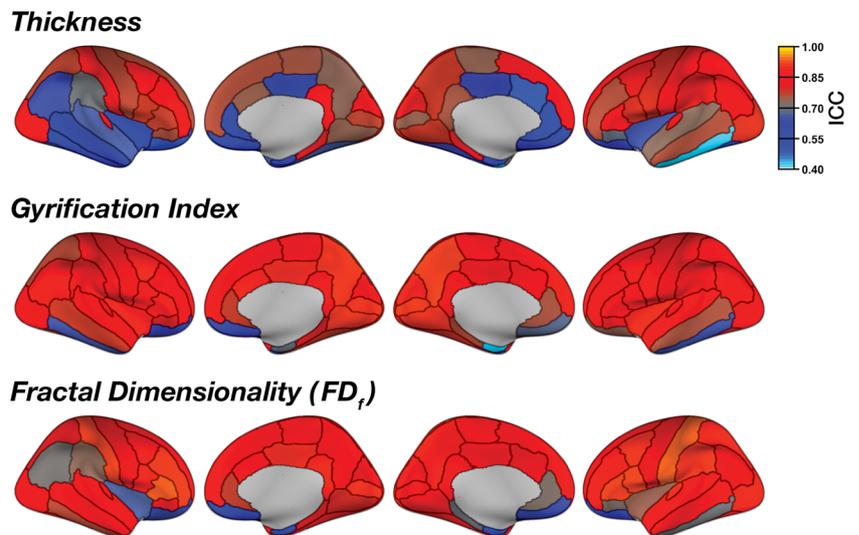
3 Study 2: GSP

To further assess the replicability of these findings, we calculated these same measures in a second dataset. While this dataset had only two MRI sessions, rather than 10, this dataset used an anatomical MRI sequence that was optimized for brain morphology research (based on prior validation work assessing cortical thickness and subcortical volume) [7, 68]. While this prior validation work suggests that reliability for cortical thickness and subcortical volume should be higher for this dataset, it is not clear how these improvements to volumetric measures may influence

shape-related measures of morphology (i.e., gyrification and fractal dimensionality).

3.1 Procedure

3.1.1 Dataset

MR images were acquired on Siemens Trio 3 T scanners at Harvard University and Massachusetts General Hospital, as part of the Brain Genome Superstruct Project (GSP; [34]). This dataset includes 1570 participants from aged 18 to 25 years old. Test–retest reliability data were available for
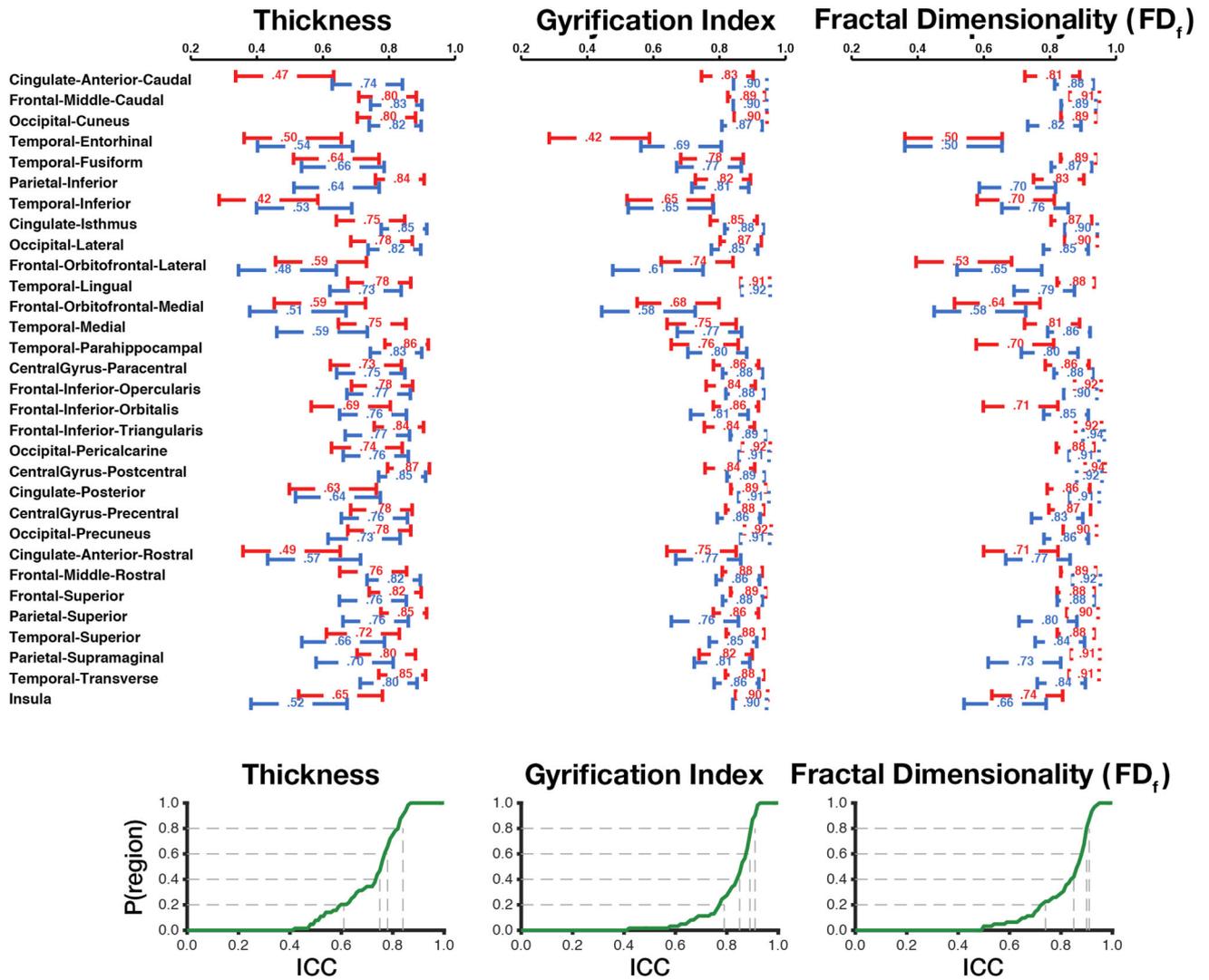
Fig. 4 Test–retest reliability (ICC) for cortical thickness, gyrification, and fractal dimensionality of the cortical parcellations, for the CCBD dataset. *Upper* mean ICC values, with 95% confidence intervals, for each region and measure. Right hemisphere regions are displayed in *red*; left hemisphere regions are displayed in *blue*. Lower: empirical cumulative distribution functions (CDFs) of the mean ICC values. *Gray lines* show the proportion of regions with at least a mean ICC of *x*. (Color figure online)

Fig. 5 Test–retest reliability (ICC; mean and 95% confidence interval) for volume and fractal dimensionality of the subcortical structures, for the CCBD dataset
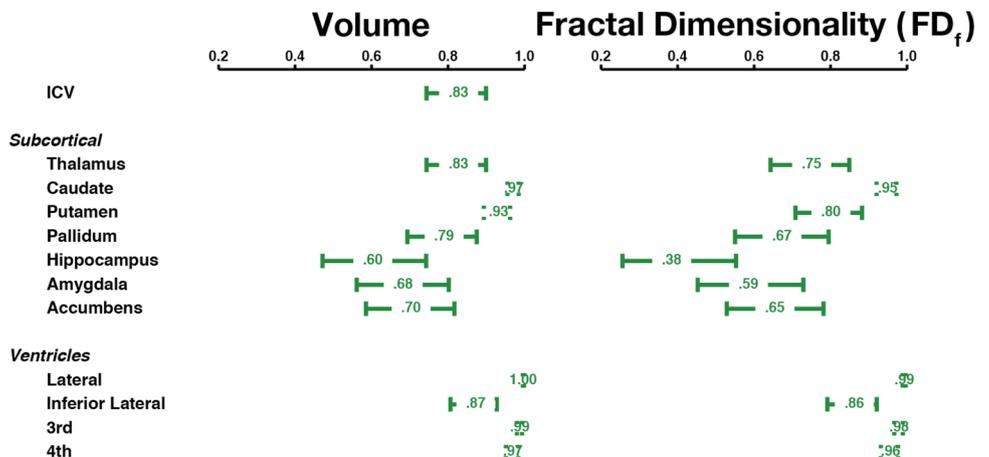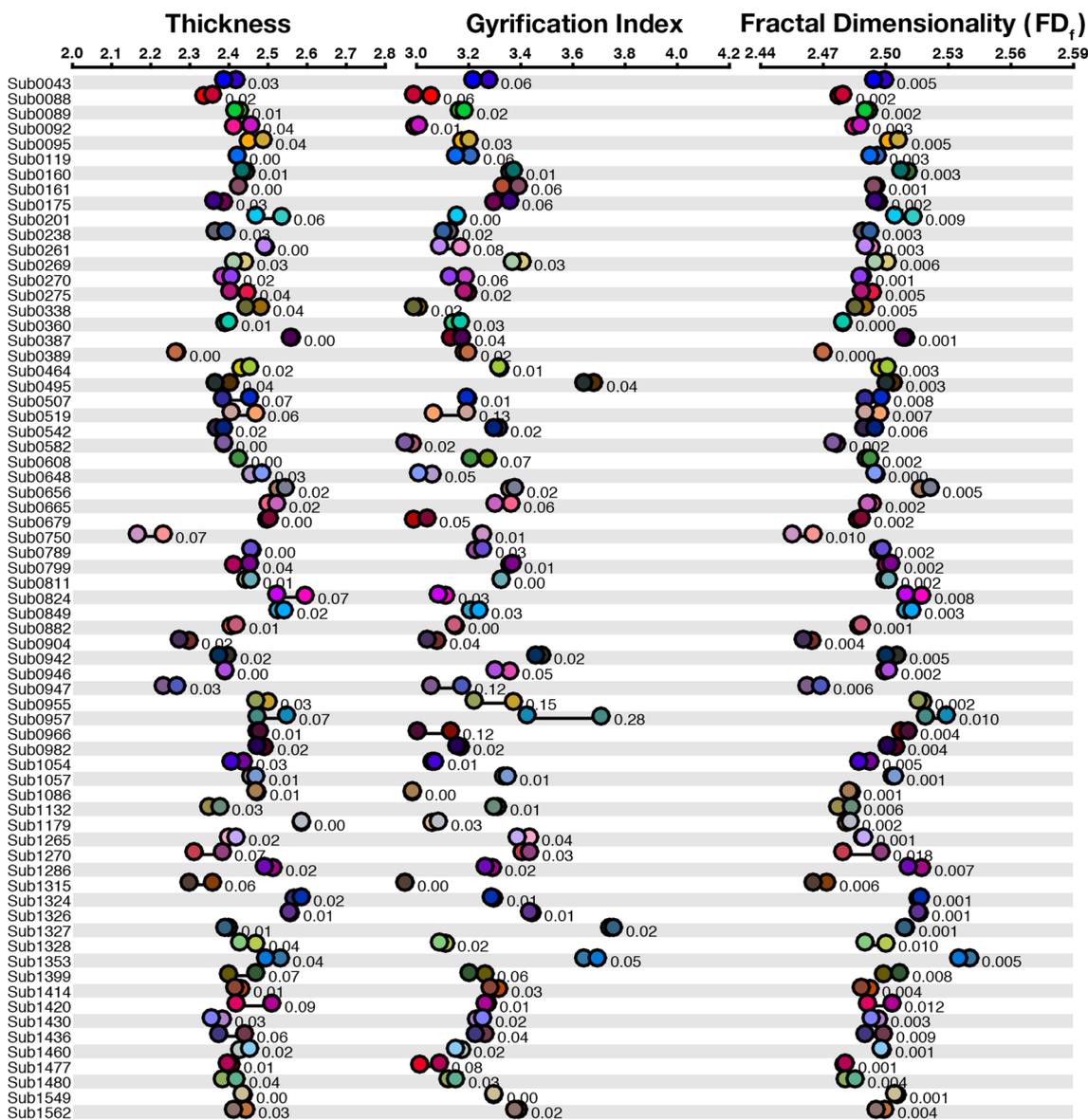
Fig. 6 Dot plot for the structural estimates for each measure for the cortical ribbon, for the GSP dataset. *Each row* represents structural metrics for a single participant, and each dot within a measure (e.g., 'Thickness') represents a scan volume. Within each row, markers in the same *color* denote measures from the same scan volume, across measures. Values beside each set of markers denote the mean deviation between estimates. (Color figure online)

69 participants who were scanned within 6 months of their first session (also see [3]). T_1-weighted data were acquired using a MEMPRAGE sequence optimized for brain morphology (TR: 2.20 s; TE: 1.5, 3.4, 5.2, 7.0 ms; flip angle: 7°; voxel size: 1.2 × 1.2 × 1.2 mm) [7, 68].

3.1.2 Data analysis

The MR images were processed using an identical procedure as in Study 1. ICC was also evaluated using the same approach.

3.2 Results

3.2.1 Cortical ribbon

As shown in Fig. 6, morphology estimates from the two sessions were generally highly concordant, though estimates did markedly differ for some participants (e.g., Sub0955, Sub0957). Nonetheless, test–retest reliability (ICC) was comparable as with the CCBD dataset (see Table 1). In almost all cases, reliability was numerically higher for the GSP dataset than for the CCBD dataset, though this difference was not statistically significant.

Fig. 7 Mean regional
morphology measures for each
parcellated region plotted on
inflated surfaces, for the GSP
dataset
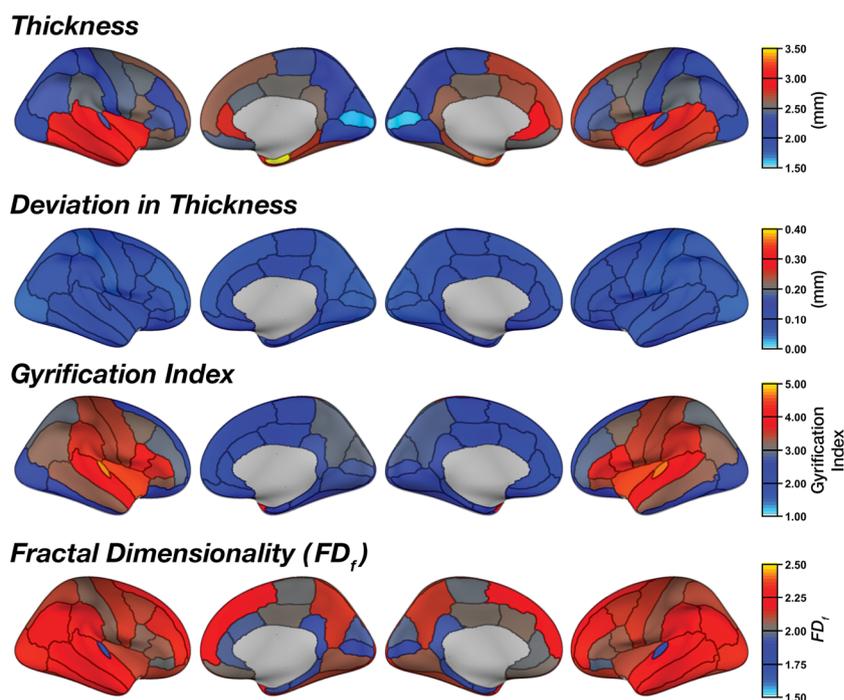
3.2.2 Cortical parcellations

Regional estimates of thickness, gyrification, and fractal
dimensionality were nearly identical between the two
datasets (see Figs. 2, 7). However, it is important to note
that test–retest reliability of regional estimates was very
high across all regions and measures (Fig. 8a) and was
indeed numerically higher than in the CCBD dataset. It is
likely the increased reliability in this dataset, relative to the
CCBD dataset, is related to the prior work optimizing the
anatomical sequence optimized for brain morphology
analyses [7, 68]. In this GSP dataset, the reliability differed
between all three measures (Fig. 8b): Regional thickness
had greater reliability than regional gyrification [$Z = 2.27$,
$p = .023$]. Regional fractal dimensionality had greater
reliability than both thickness [$Z = 7.21$, $p < .001$] and
gyrification [$Z = 4.91$, $p < .001$].

3.2.3 Subcortical structures

As shown in Fig. 8c, test–retest reliability was near per-
fect for both volume and fractal dimensionality of the
subcortical structures. The regions that had relatively
lower reliability (pallidum, amygdala, accumbens) were
also relatively lower in Study 1, demonstrating the
replicability of lower test–retest reliability in these
regions—at least when segmented using FreeSurfer's
automated algorithms. Reliability was particularly high
for the hippocampus and was significantly higher than in
the CCBD dataset (Study 1).

4 Discussion

Here we evaluated the test–retest reliability of several brain
morphology measures using open-access datasets. Prior
work had examined the reliability of volumetric mea-
sures—cortical thickness and subcortical volume; however,
the present study is the first to assess reliability of *shape*-
related measures, gyrification and fractal dimensionality.

Both datasets showed relatively high reliability for all
morphology measures and additionally revealed that reli-
ability was particularly good for the gyrification and fractal
dimensionality measures. Additionally, we provide empir-
ical evidence that the dilation approach for calculating
fractal dimensionality was superior in reliability to the
'standard' box-counting method. These findings held
across two datasets, but reliability was particularly good in
the GSP dataset, where the anatomical sequence had been
previously optimized for use in brain morphology studies.

Although reliability was good in these datasets, there is
still the question of how reliability may be increased in
future studies. A number of factors have been found to
influence estimates of brain morphology. Broadly, these
factors can be divided into three categories: MR acquisi-
tion, biological, and analysis related. For MR acquisition,
there are not yet enough datasets available to systemati-
cally examine how reliability is affected by the particular
acquisition protocols, although the current data suggest that
sequences previously optimized for brain morphology
analyses (i.e., those used in GSP dataset) will have better
reliability. Another acquisition-related factor is head

Fig. 8 Test–retest reliability (ICC) for regional parcellations and subcortical structures, for the GSP dataset. **a** ICCs for cortical thickness, gyrification, and fractal dimensionality of the cortical parcellations. **b** Empirical cumulative distribution functions (CDFs). *Gray lines* show the proportion of regions with at least a mean ICC of *x*. (C) ICCs (mean and 95% confidence interval) for volume and fractal dimensionality of the subcortical structures
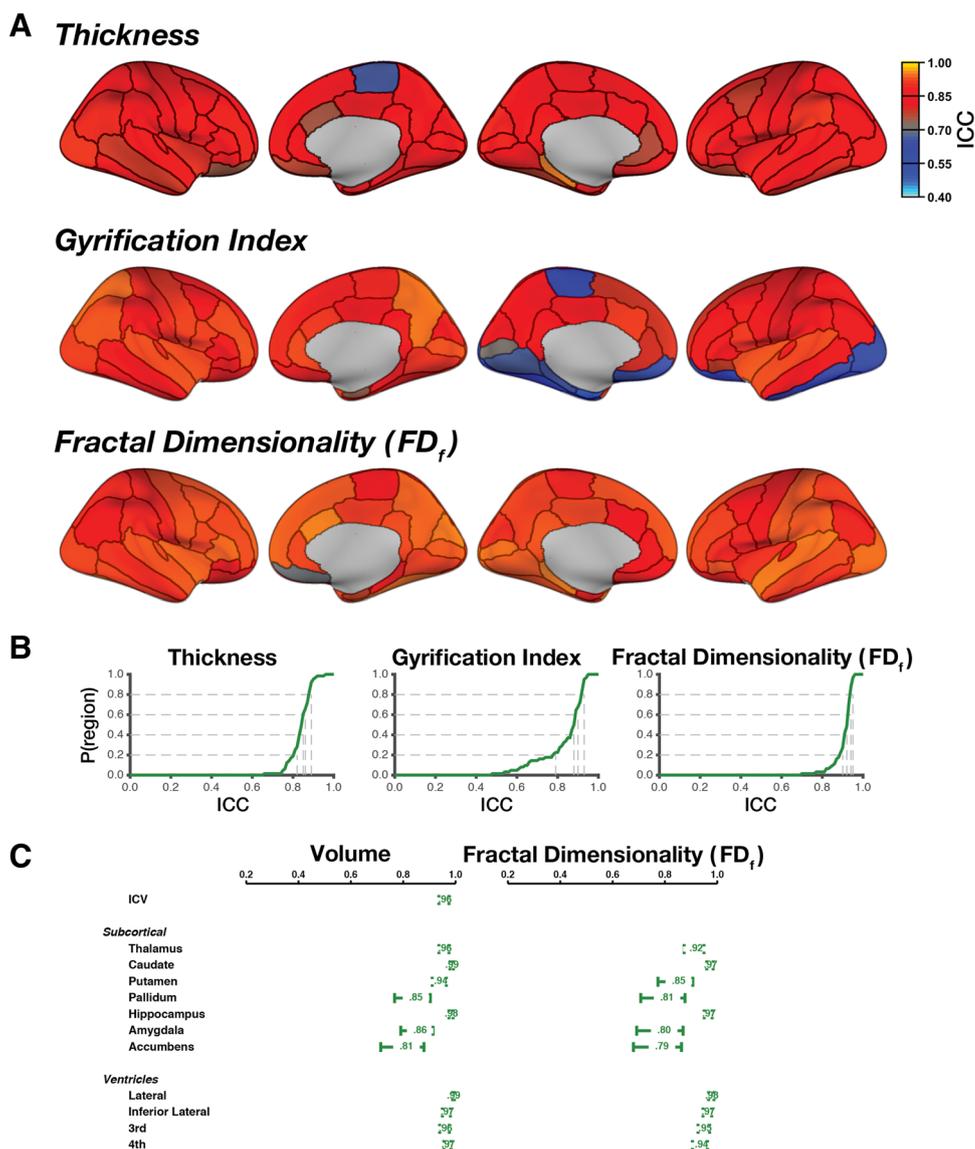
movement; movement has been shown to lead to decreased estimates of cortical thickness [69–72], though it is unclear how movement would affect measures of gyrification and fractal dimensionality. This issue may become less critical in future studies, as recent advances in structural imaging have been able to attenuate movement-related artifacts (e.g., [73–76]). Morphological measures can also be influenced by biological confounds, such as hydration [77–80] or circadian rhythms [81, 82]. Additionally, it is important to control for variations in analysis software and operating system, which can also affect brain morphology estimates [65, 83, 84].

While the surface reconstructions were visually inspected, the surfaces were not manually edited, for two reasons. First and foremost, the quality of the automatic reconstructions was judged to be acceptable and did not require manual intervention. While manual editing is more

necessary with older adult and patient populations, all of the individuals included in the present work were young adults. Additionally, manual editing introduces a subjective component and is often not conducted in studies of reconstruction reliability [2, 5, 6, 46], though some reliability studies have included minimal manual editing [4, 7]. Given that no manual editing was conducted, the reliability estimates presented here may serve as a lower bound, where manual editing would be expected to increase reliability [4, 6]; however, there is evidence that editing may not sufficiently influence regional estimates [85, 86].

Fractal dimensionality was used here as a measure of the complexity in the shape of a structure. Results indicate that this measure was generally more reliable than volumetric morphological measures, likely because fractal dimensionality is influenced by both shape and volumetric characteristics that often covary [21, 24, 87–89]. By

pooling from both of these characteristics, fractal dimensionality appears to be more reliable and should be considered in future research investigating the relationship between brain morphology and inter-individual differences.

In sum, here we evaluated the reliability of several brain morphology estimates using two open-access datasets. Reliability was generally high, providing support for using gyrification and fractal dimensionality measures to evaluate inter-individual or between-sample differences in morphology.

Compliance with ethical standards

Appendix 1

See Table 2.

Table 2 List of open-access datasets that include intersession test–retest structural MRIs

References	N	Notes
Analyzed TRT datasets		
Chen et al. [33]	30	Ten sessions acquired over 1-month period (2–3 days between sessions); part of CoRR (HNU1; see below)
Holmes et al. [34]	69	Two sessions within 6-month period; part of larger dataset with $N = 1570$; also see Holmes et al. [3]
Additional TRT datasets		
Boekel et al. [99]	34	Two sessions within same day; subsample of $N = 15$ had a third session in same day and a 2-week follow-up
Marcus et al. [100]	20	Three–four volumes within session, for two sessions within a 90-day period; part of larger cross-sectional dataset of aging and dementia with $N = 416$
Morey et al. [11]	23	Two sessions within single day, follow-up in 7–9 days with another two sessions within single day
Gorgolewski et al. [101]	22	Two sessions 1 week apart, three rs-fMRI scans per session (includes high-res prefrontal cortex scan); acquired on a 7 T scanner
Landman et al. [102]	21	Two sessions within single day; multiple sequences
Gorgolewski et al. [103]	10	Two sessions acquired 2–3 days apart
Highly sampled individual participant datasets		
Maclaren et al. [104]	3	Two volumes within single session, for each of 20 sessions over 1-month period
Poldrack et al. [105]	1	One hundred and four sessions; scanned intermittently over 18 months (10 usable T_1 volumes); also see Laumann et al. [106]
Choe et al. [107]	1	One hundred and fifty-eight sessions; scanned weekly for 3.5 years
Froeling et al. [108]	1	Eighteen sessions, comprising 8000 dMRI volumes (5 sessions included T_1 volumes)
Datasets part of CoRR		
Zuo et al. [36]	–	Consortium for Reliability and Reproducibility (CoRR), aggregates many TRT datasets
Orban et al. [109]	80	Two volumes within single session, for each of two sessions within 3-month period; part of CoRR (UM1)
Lin et al. [110]	57	Two sessions within 6-week period; part of CoRR (BNU1)
Huang et al. [111]	61	Two sessions within 6-month period; part of CoRR (BNU2)

Note, nearly all of these datasets also include test–retest rs-fMRI data, some additionally collected task-based fMRI data

Appendix 2

Generating DKT volumes for the calcFD toolbox

calcFD toolbox (build 28 [and above]; http://cmadan.github.io/calcFD/) can calculate the fractal dimensionality for parcellated cortical regions based on the DKT atlas. This is done based on the `aparc.DKTaltas40+aseg.mgz` volume, which must first be generated using FreeSurfer, but is not part of the standard analysis pipeline. The FreeSurfer command to accomplish this is:

```
mri_aparc2aseg − s[SUBJECTID]
   − annotaparc.DKTatlas40
```

where [SUBJECTID] corresponds to the individual subject folder.

After these volumes have been generated by FreeSurfer, the calcFD toolbox only needs the options to be set to DKT (set `options.aparc` to 'DKT').

Appendix 3

Measuring fractal dimensionality via spherical harmonics

In addition to the fractal dimensionality measures considered in Madan and Kensinger [21], we additionally considered an approach based on spherical harmonics (often abbreviated as 'SPHARM'). Yotter et al. [58] demonstrated that fractal dimensionality can be calculated using spherical harmonics and compared this to the standard box-counting approach. Importantly, they found that the spherical harmonics approach was more robust to rotations of the structure than the box-counting method. We additionally implemented this approach when evaluating the test–retest reliability of fractal dimensionality estimates.

Briefly, spherical harmonics can be used to reconstruct complex 3D surfaces based on space–frequency deformations to a sphere, based on similar principles as used to reconstruct complex wave functions using Fourier transforms. Some of the spherical harmonics basis functions are shown in Fig. 9. Chung [54, 90] provides a comprehensive introduction to these principles.

We used weighted spherical harmonics, a generalized form of traditional spherical harmonics, which substantially reduces ringing artifacts related to the Gibbs phenomenon [54, 56]. Previous studies have used spherical harmonics to study the shape of cortical and subcortical structures (e.g., [55–57, 87, 89, 91–98]), but spherical harmonics have not been connected with fractal dimensionality approaches until recently [58].

The fractal dimensionality approach we took, using spherical harmonics, was conducted based on weighted spherical harmonics equations provided by Chung et al. [54–56] and the fractal dimensionality equations from Yotter et al. [58]. The spherical parameterization provided by FreeSurfer (`?h.sphere, ?h.pial`) is used as the input surfaces for this processes. Cortical surfaces were reconstructed for each hemisphere with a maximum degree of $l = \{11, 16, 20, 29\}$ (a subset of those suggested by [58]) and a bandwidth of $\sigma = .001$ (as recommended by [54]). The reconstruction of one hemisphere, across a variety of degrees, is shown in Fig. 10. The calculations involved in reconstructing cortical surfaces using weighted spherical harmonics are discussed in detail in Chung [54].
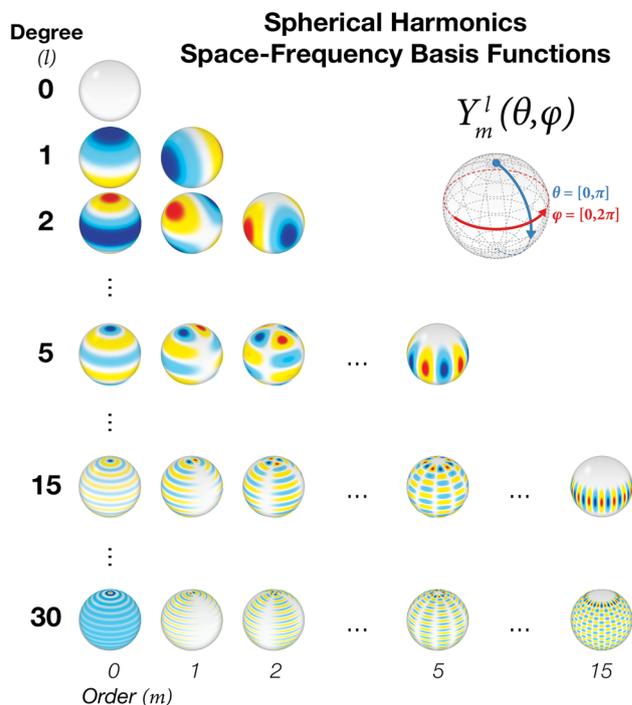
Fig. 9 Spherical harmonics space–frequency basis functions, organized by degree (l) and order (m)
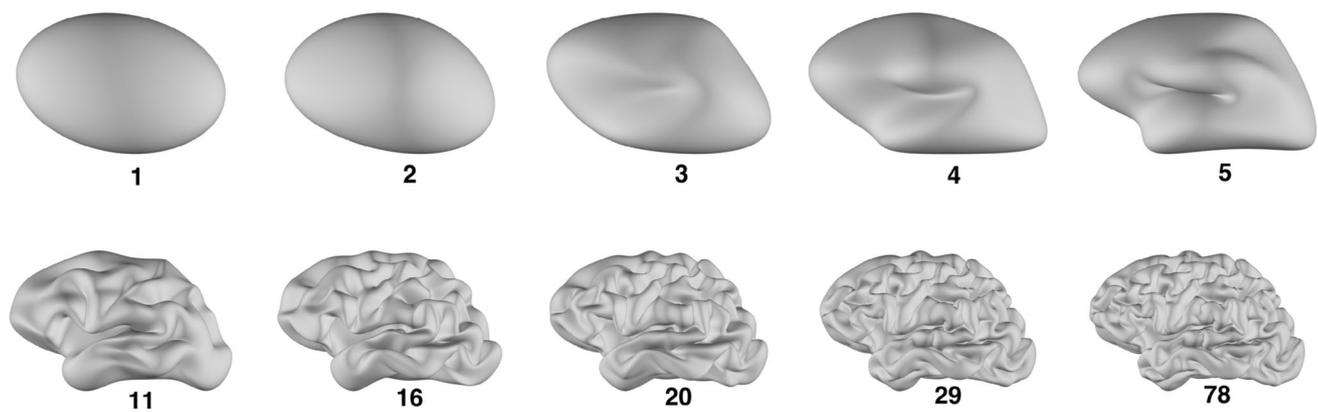
Fig. 10 Reconstructed cortical surface for the left hemisphere for one participant, with the maximum number of degrees as denoted under each surface

References

1. Dickerson BC, Fenstermacher E, Salat DH, Wolk DA, Maguire RP, Desikan R et al (2008) Detection of cortical thickness correlates of cognitive performance: reliability across MRI scan sessions, scanners, and field strengths. NeuroImage 39:10–18. doi:10.1016/j.neuroimage.2007.08.042

2. Han X, Jovicich J, Salat D, van der Kouwe A, Quinn B, Czanner S et al (2006) Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. NeuroImage 32:180–194. doi:10.1016/j.neuroimage.2006.02.051

3. Holmes AJ, Hollinshead MO, Roffman JL, Smoller JW, Buckner RL (2016) Individual differences in cognitive control circuit anatomy link sensation seeking, impulsivity, and substance use. J Neurosci 36:4038–4049. doi:10.1523/jneurosci.3206-15.2016

4. Iscan Z, Jin TB, Kendrick A, Szeglin B, Lu H, Trivedi M et al (2015) Test–retest reliability of FreeSurfer measurements within and between sites: effects of visual approval process. Hum Brain Mapp 36:3472–3485. doi:10.1002/hbm.22856

5. Jovicich J, Marizzoni M, Sala-Llonch R, Bosch B, Bartrés-Faz D, Arnold J et al (2013) Brain morphometry reproducibility in multi-center 3T MRI studies: a comparison of cross-sectional and longitudinal segmentations. NeuroImage 83:472–484. doi:10.1016/j.neuroimage.2013.05.007

6. Liem F, Mérillat S, Bezzola L, Hirsiger S, Philipp M, Madhyastha T, Jäncke L (2015) Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. NeuroImage 108:95–109. doi:10.1016/j.neuroimage.2014.12.035

7. Wonderlick JS, Ziegler DA, Hosseini-Varnamkhasti P, Locascio J, Bakkour J, van der Kouwe A et al (2009) Reliability of MRI-derived cortical and subcortical morphometric measures: effects of pulse sequence, voxel geometry, and parallel imaging. NeuroImage 44:1324–1333. doi:10.1016/j.neuroimage.2008.10.037

8. Bartzokis G, Mintz J, Marx P, Osborn D, Gutkind D, Chiang F et al (1993) Reliability of in vivo volume measures of hippocampus and other brain structures using MRI. Magn Reson Imaging 11:993–1006. doi:10.1016/0730-725x(93)90218-3

9. Goodro M, Sameti M, Patenaude B, Fein G (2012) Age effect on subcortical structures in healthy adults. Psychiatry Res Neuroimaging 203:38–45. doi:10.1016/j.pscychresns.2011.09.014

10. Jovicich J, Czanner S, Han X, Salat D, van der Kouwe A, Quinn B et al (2009) MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths. NeuroImage 46:177–192. doi:10.1016/j.neuroimage.2009.02.010

11. Morey RA, Selgrade ES, Wagner HR, Huettel SA, Wang L, McCarthy G (2010) Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Hum Brain Mapp 31:1751–1762. doi:10.1002/hbm.20973

12. Nugent AC, Luckenbaugh DA, Wood SE, Bogers W, Zarate CA, Drevets WC (2013) Automated subcortical segmentation using FIRST: test–retest reliability, interscanner reliability, and comparison to manual segmentation. Hum Brain Mapp 34:2313–2329. doi:10.1002/hbm.22068

13. Pestilli F (2015) Test–retest measurements and digital validation for in vivo neuroscience. Sci Data 2:140057. doi:10.1038/sdata.2014.57

14. Mietchen D, Gaser C (2009) Computational morphometry for detecting changes in brain structure due to development, aging, learning, disease and evolution. Front Neuroinf 3:25. doi:10.3389/neuro.11.025.2009

15. Toro R, Perron M, Pike B, Richer L, Veillette S, Pausova Z, Paus T (2008) Brain size and folding of the human cerebral cortex. Cereb Cortex 18:2352–2357. doi:10.1093/cercor/bhm261

16. Armstrong E, Schleicher A, Omran H, Curtis M, Zilles K (1995) The ontogeny of human gyrification. Cereb Cortex 5:56–63. doi:10.1093/cercor/5.1.56

17. Zilles K, Armstrong E, Schleicher A, Kretschmann H-J (1988) The human pattern of gyrification in the cerebral cortex. Anat Embryol 179:173–179. doi:10.1007/BF00304699

18. Zilles K, Armstrong E, Moser KH, Schleicher A, Stephan H (1989) Gyrification in the cerebral cortex of primates. Brain Behav Evol 34:143–150. doi:10.1159/000116500

19. Toro R (2012) On the possible shapes of the brain. Evol Biol 39:600–612. doi:10.1007/s11692-012-9201-8

20. Hogstrom LJ, Westlye LT, Walhovd KB, Fjell AM (2013) The structure of the cerebral cortex across adult Life: age-related patterns of surface area, thickness, and gyrification. Cereb Cortex 23:2521–2530. doi:10.1093/cercor/bhs231

21. Madan CR, Kensinger EA (2016) Cortical complexity as a measure of age-related brain atrophy. NeuroImage 134:617–629. doi:10.1016/j.neuroimage.2016.04.029

22. Magnotta VA, Andreasen NC, Schultz SK, Harris G, Cizadlo T, Heckel D et al (1999) Quantitative in vivo measurement of gyrification in the human brain: changes associated with aging. Cereb Cortex 9:151–160. doi:10.1093/cercor/9.2.151

23. Mandelbrot BB (1967) How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 156:636–638. doi:10.1126/science.156.3775.636

24. Madan CR, Kensinger EA (2017) Age-related differences in the structural complexity of subcortical and ventricular structures. Neurobiol Aging 50:87–95. doi:10.1016/j.neurobiolaging.2016.10.023

25. Di Ieva A, Esteban FJ, Grizzi F, Klonowski W, Martin-Landrove M (2015) Fractals in the neurosciences, part II: clinical applications and future perspectives. Neuroscientist 21:30–43. doi:10.1177/1073858413513928

26. Di Ieva A, Grizzi F, Jelinek H, Pellionisz AJ, Losa GA (2014) Fractals in the neurosciences, part I: general principles and basic neurosciences. Neuroscientist 20:403–417. doi:10.1177/1073858413513927

27. Schaer M, Cuadra MB, Schmansky N, Fischl B, Thiran J-P, Eliez S (2012) How to measure cortical folding from MR images: a step-by-step tutorial to compute local gyrification index. J Vis Exp 59:e3417. doi:10.3791/3417

28. Samanez-Larkin GR, D'Esposito M (2008) Group comparisons: imaging the aging brain. Soc Cognit Affect Neurosci 3:290–297. doi:10.1093/scan/nsn029

29. Garrett DD, Samanez-Larkin GR, MacDonald SWS, Lindenberger U, McIntosh AR, Grady CL (2013) Moment-to-moment brain signal variability: a next frontier in human brain mapping? Neurosci Biobehav Rev 37:610–624. doi:10.1016/j.neubiorev.2013.02.015

30. McIntosh AR, Vakorin V, Kovacevic N, Wang H, Diaconescu A, Protzner AB (2014) Spatiotemporal dependency of age-related changes in brain signal variability. Cereb Cortex 24:1806–1817. doi:10.1093/cercor/bht030

31. Thomas BP, Liu P, Park DC, van Osch MJ, Lu H (2014) Cerebrovascular reactivity in the brain white matter: magnitude, temporal characteristics, and age effects. J Cereb Blood Flow Metab 34:242–247. doi:10.1038/jcbfm.2013.194

32. Tsvetanov KA, Henson RNA, Tyler LK, Davis SW, Shafto MA, Taylor JR et al (2015) The effect of ageing on fMRI: correction for the confounding effects of vascular reactivity evaluated by joint fMRI and MEG in 335 adults. Hum Brain Mapp 36:2248–2269. doi:10.1002/hbm.22768

33. Chen B, Xu T, Zhou C, Wang L, Yang N, Wang Z et al (2015) Individual variability and test–retest reliability revealed by ten repeated resting-state brain scans over one month. PLoS ONE 10:e0144963. doi:10.1371/journal.pone.0144963

34. Holmes AJ, Hollinshead MO, O'Keefe TM, Petrov VI, Fariello GR, Wald LL et al (2015) Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. Sci Data 2:150031. doi:10.1038/sdata.2015.31

35. Klein A, Tourville J (2012) 101 labeled brain images and a consistent human cortical labeling protocol. Front Neurosci 6:171. doi:10.3389/fnins.2012.00171

36. Zuo X-N, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J et al (2014) An open science resource for establishing reliability and reproducibility in functional connectomics. Sci Data 1:140049. doi:10.1038/sdata.2014.49

37. Fischl B (2012) FreeSurfer. NeuroImage 62:774–781. doi:10.1016/j.neuroimage.2012.01.021

38. Fischl B, Dale AM (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc Natl Acad Sci USA 97:11050–11055. doi:10.1073/pnas.200033797

39. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C et al (2002) Whole brain segmentation: automated labelling of neuroanatomical structures in the human brain. Neuron 33:341–355. doi:10.1016/s0896-6273(02)00569-x

40. Fischl B, Salat DH, van der Kouwe AJW, Makris N, Ségonne F, Quinn BT, Dale AM (2004) Sequence-independent segmentation of magnetic resonance images. NeuroImage 23:S69–S84. doi:10.1016/j.neuroimage.2004.07.016

41. Kuperberg GR, Broome MR, McGuire PK, David AS, Eddy M, Ozawa F et al (2003) Regionally localized thinning of the cerebral cortex in Schizophrenia. Arch Gen Psychiatry 60:878. doi:10.1001/archpsyc.60.9.878

42. Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RSR et al (2004) Thinning of the cerebral cortex in aging. Cereb Cortex 14:721–730. doi:10.1093/cercor/bhh032

43. Cardinale F, Chinnici G, Bramerio M, Mai R, Sartori I, Cossu M et al (2014) Validation of FreeSurfer-estimated brain cortical thickness: comparison with histologic measurements. Neuroinformatics 12:535–542. doi:10.1007/s12021-014-9229-2

44. Rosas HD, Liu AK, Hersch S, Glessner M, Ferrante RJ, Salat DH et al (2002) Regional and progressive thinning of the cortical ribbon in Huntington's disease. Neurology 58:695–701. doi:10.1212/wnl.58.5.695

45. Grimm O, Pohlack S, Cacciaglia R, Winkelmann T, Plichta MM, Demirakca T, Flor H (2015) Amygdalar and hippocampal volume: a comparison between manual segmentation, Freesurfer and VBM. J Neurosci Methods 253:254–261. doi:10.1016/j.jneumeth.2015.05.024

46. Keller SS, Gerdes JS, Mohammadi S, Kellinghaus C, Kugel H, Deppe K et al (2012) Volume estimation of the thalamus using FreeSurfer and stereology: consistency between methods. Neuroinformatics 10:341–350. doi:10.1007/s12021-012-9147-0

47. Lehmann M, Douiri A, Kim LG, Modat M, Chan D, Ourselin S et al (2010) Atrophy patterns in Alzheimer's disease and semantic dementia: a comparison of FreeSurfer and manual volumetric measurements. NeuroImage 49:2264–2274. doi:10.1016/j.neuroimage.2009.10.056

48. Morey RA, Petty CM, Xu Y, Pannu Hayes J, Wagner HR, Lewis DV et al (2009) A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. NeuroImage 45:855–866. doi:10.1016/j.neuroimage.2008.12.033

49. Mulder ER, de Jong RA, Knol DL, van Schijndel RA, Cover KS, Visser PJ et al (2014) Hippocampal volume change measurement: quantitative assessment of the reproducibility of expert manual outlining and the automated methods FreeSurfer and FIRST. NeuroImage 92:169–181. doi:10.1016/j.neuroimage.2014.01.058

50. Pardoe HR, Pell GS, Abbott DF, Jackson GD (2009) Hippocampal volume assessment in temporal lobe epilepsy: how good is automated segmentation? Epilepsia 50:2586–2592. doi:10.1111/j.1528-1167.2009.02243.x

51. Tae WS, Kim SS, Lee KU, Nam E-C, Kim KW (2008) Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. Neuroradiology 50:569–581. doi:10.1007/s00234-008-0383-9

52. Wenger E, Mårtensson J, Noack H, Bodammer NC, Kühn S, Schaefer S et al (2014) Comparing manual and automatic segmentation of hippocampal volumes: reliability and validity issues in younger and older brains. Hum Brain Mapp 35:4236–4248. doi:10.1002/hbm.22473

53. Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, Snyder AZ (2004) A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. NeuroImage 23:724–738. doi:10.1016/j.neuroimage.2004.06.018

54. Chung MK (2014) Statistical and computational methods in brain image analysis. CRC Press, New York

55. Chung MK, Dalton KM, Davidson RJ (2008) Tensor-based cortical surface morphometry via weighted spherical harmonic

representation. IEEE Trans Med Imaging 27:1143–1151. doi:10.1109/tmi.2008.918338

56. Chung MK, Dalton KM, Shen L, Evans AC, Davidson RJ (2007) Weighted Fourier series representation and its application to quantifying the amount of gray matter. IEEE Trans Med Imaging 26:566–581. doi:10.1109/tmi.2007.892519

57. Shen L, Firpi HA, Saykin AJ, West JD (2009) Parametric surface modeling and registration for comparison of manual and automated segmentation of the hippocampus. Hippocampus 19:588–595. doi:10.1002/hipo.20613

58. Yotter RA, Nenadic I, Ziegler G, Thompson PM, Gaser C (2011) Local cortical surface complexity maps from spherical harmonic reconstructions. NeuroImage 56:961–973. doi:10.1016/j.neuroimage.2011.02.007

59. Asendorpf J, Wallbott HG (1979) Maße der Beobachterübereinstimmung: ein systematischer Vergleich. Zeitschrift für Sozialpsychologie 10:243–252

60. Bartko JJ (1966) The intraclass correlation coefficient as a measure of reliability. Psychol Rep 19:3–11. doi:10.2466/pr0.1966.19.1.3

61. Rajaratnam N (1960) Reliability formulas for independent decision data when reliability data are matched. Psychometrika 25:261–271. doi:10.1007/bf02289730

62. Shrout PE, Fleiss JL (1979) Intraclass correlations: uses in assessing rater reliability. Psychol Bull 86:420–428. doi:10.1037/0033-2909.86.2.420

63. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–46. doi:10.1037/1082-989x.1.1.30

64. Cicchetti DV (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 6:284–290. doi:10.1037/1040-3590.6.4.284

65. Glatard T, Lewis LB, Ferreira da Silva R, Adalat R, Beck N, Lepage C et al (2015) Reproducibility of neuroimaging analyses across operating systems. Front Neuroinf. doi:10.3389/fninf.2015.00012

66. Fjell AM, Westlye LT, Amlien I, Espeseth T, Reinvang I, Raz N et al (2009) High consistency of regional cortical thinning in aging across multiple samples. Cereb Cortex 19:2001–2012. doi:10.1093/cercor/bhn232

67. Hutton C, Draganski B, Ashburner J, Weiskopf N (2009) A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. NeuroImage 48:371–380. doi:10.1016/j.neuroimage.2009.06.043

68. van der Kouwe AJW, Benner T, Salat DH, Fischl B (2008) Brain morphometry with multiecho MPRAGE. NeuroImage 40:559–569. doi:10.1016/j.neuroimage.2007.12.025

69. Alexander-Bloch A, Clasen L, Stockman M, Ronan L, Lalonde F, Giedd J, Raznahan A (2016) Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. Hum Brain Mapp 36:2385–2397. doi:10.1002/hbm.23180

70. Pardoe HR, Kucharsky Hiess R, Kuzniecky R (2016) Motion and morphometry in clinical and nonclinical populations. NeuroImage 135:177–185. doi:10.1016/j.neuroimage.2016.05.005

71. Reuter M, Tisdall MD, Qureshi A, Buckner RL, van der Kouwe AJW, Fischl B (2015) Head motion during MRI acquisition reduces gray matter volume and thickness estimates. NeuroImage 107:107–115. doi:10.1016/j.neuroimage.2014.12.006

72. Savalia NK, Agres PF, Chan MY, Feczko EJ, Kennedy KM, Wig GS (2017) Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. Hum Brain Mapp 38:472–492. doi:10.1002/hbm.23397

73. Federau C, Gallichan D (2016) Motion-correction enabled ultra-high resolution in vivo 7T-MRI of the brain. PLoS ONE 11:e0154974. doi:10.1371/journal.pone.0154974

74. Maclaren J, Herbst M, Speck O, Zaitsev M (2013) Prospective motion correction in brain imaging: a review. Magn Reson Med 69:621–636. doi:10.1002/mrm.24314

75. Stucht D, Danishad KA, Schulze P, Godenschweger F, Zaitsev M, Speck O (2015) Highest resolution in vivo human brain MRI using prospective motion correction. PLoS ONE 10:e0133921. doi:10.1371/journal.pone.0133921

76. Tisdall MD, Reuter M, Qureshi A, Buckner RL, Fischl B, van der Kouwe AJW (2016) Prospective motion correction with volumetric navigators (vNavs) reduces the bias and variance in brain morphometry induced by subject motion. NeuroImage 127:11–22. doi:10.1016/j.neuroimage.2015.11.054

77. Duning T, Kloska S, Steinstrater O, Kugel H, Heindel W, Knecht S (2005) Dehydration confounds the assessment of brain atrophy. Neurology 64:548–550. doi:10.1212/01.wnl.0000150542.16969.cc

78. Kempton MJ, Ettinger U, Schmechtig A, Winter EM, Smith L, McMorris T et al (2009) Effects of acute dehydration on brain morphology in healthy humans. Hum Brain Mapp 30:291–298. doi:10.1002/hbm.20500

79. Nakamura K, Brown RA, Araujo D, Narayanan S, Arnold DL (2014) Correlation between brain volume change and T2 relaxation time induced by dehydration and rehydration: implications for monitoring atrophy in clinical studies. NeuroImage Clin 6:166–170. doi:10.1016/j.nicl.2014.08.014

80. Streitbürger D-P, Möller HE, Tittgemeyer M, Hund-Georgiadis M, Schroeter ML, Mueller K (2012) Investigating structural brain changes of dehydration using voxel-based morphometry. PLoS ONE 7:e44195. doi:10.1371/journal.pone.0044195

81. Nakamura K, Brown RA, Narayanan S, Collins DL, Arnold DL (2015) Diurnal fluctuations in brain volume: statistical analyses of MRI from large populations. NeuroImage 118:126–132. doi:10.1016/j.neuroimage.2015.05.077

82. Trefler A, Sadeghi N, Thomas AG, Pierpaoli C, Baker CI, Thomas C (2016) Impact of time-of-day on brain morphometric measures derived from T1-weighted magnetic resonance imaging. NeuroImage 133:41–52. doi:10.1016/j.neuroimage.2016.02.034

83. Chepkoech J-L, Walhovd KB, Grydeland H, Fjell AM (2016) Effects of change in FreeSurfer version on classification accuracy of patients with Alzheimer's disease and mild cognitive impairment. Hum Brain Mapp 37:1831–1841. doi:10.1002/hbm.23139

84. Gronenschild EHBM, Habets P, Jacobs HIL, Mengelers R, Rozendaal N, van Os J, Marcelis M (2012) The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS ONE 7:e38234. doi:10.1371/journal.pone.0038234

85. McCarthy CS, Ramprashad A, Thompson C, Botti J-A, Coman IL, Kates WR (2015) A comparison of FreeSurfer-generated data with and without manual intervention. Front Neurosci 9:379. doi:10.3389/fnins.2015.00379

86. Ronan L, Alexander-Bloch AF, Wagstyl K, Farooqi S, Brayne C, Tyler LK et al (2016) Obesity associated with increased brain age from midlife. Neurobiol Aging 47:63–70. doi:10.1016/j.neurobiolaging.2016.07.010

87. Gerig G, Styner M, Jones D, Weinberger D, Lieberman J (2001b) Shape analysis of brain ventricles using SPHARM. In: Proceedings of the IEEE workshop on mathematical methods in biomedical image analysis (MMBIA 2001), pp 171–178. doi:10.1109/mmbia.2001.991731

88. King RD, George AT, Jeon T, Hynan LS, Youn TS, Kennedy DN, Dickerson B (2009) Characterization of atrophic changes in the cerebral cortex using fractal dimensional analysis. Brain Imaging Behav 3:154–166. doi:10.1007/s11682-008-9057-9

89. Nitzken MJ, Casanova MF, Gimelfarb G, Inanc T, Zurada JM, El-Baz A (2014) Shape analysis of the human brain: a brief

survey. IEEE J Biomed Health Inf 18:1337–1354. doi:10.1109/jbhi.2014.2298139

90. Chung MK (2013) Computational neuroanatomy: the methods. World Scientific Publishing, Hackensack

91. Chung MK, Nacewicz BM, Wang S, Dalton KM, Pollak S, Davidson RJ (2008) Amygdala surface modeling with weighted spherical harmonics. Lect Notes Comput Sci 5128:177–184. doi:10.1007/978-3-540-79982-5_20

92. Chung MK, Worsley KJ, Nacewicz BM, Dalton KM, Davidson RJ (2010) General multivariate linear modeling of surface shapes using SurfStat. NeuroImage 53:491–505. doi:10.1016/j.neuroimage.2010.06.032

93. Dombroski B, Nitzken M, Elnakib A, Khalifa F, Switala A, El-Baz A, Casanova M (2014) Cortical surface complexity in a population-based normative sample. Transl Neurosci. doi:10.2478/s13380-014-0202-1

94. Gerig G, Styner M, Shenton ME, Lieberman JA (2001) Shape versus size: improved understanding of the morphology of brain structures. Lect Notes Comput Sci 2208:24–32. doi:10.1007/3-540-45468-3_4

95. Gong Z, Lu J, Chen J, Wang Y, Yuan Y, Zhang T et al (2011) Ventricle shape analysis for centenarians, elderly subjects, MCI and AD patients. Lect Notes Comput Sci 7012:84–92. doi:10.1007/978-3-642-24446-9_11

96. Shen L, Saykin AJ, Kim S, Firpi HA, West JD, Risacher SL et al (2010) Comparison of manual and automated determination of hippocampal volumes in MCI and early AD. Brain Imaging Behav 4:86–95. doi:10.1007/s11682-010-9088-x

97. Styner M, Oguz I, Xu S, Brechbühler C, Pantazis D, Levitt JJ et al (2006) Framework for the statistical shape analysis of brain structures using SPHARM-PDM. Insight Journal 1071:242–250

98. Yu P, Yeo BTT, Grant PE, Fischl B, Golland P (2007) Cortical folding development study based on over-complete spherical wavelets. In: International Conference on Computer Vision 2007 proceedings of the workshop on mathematical methods in biomedical image analysis (MMBIA). doi:10.1109/iccv.2007.4409137

99. Boekel W, Keuken MC, Forstmann BU (2017) A test–retest reliability analysis of diffusion measures of white matter tracts relevant for cognitive control. Psychophysiology 54:24–33. doi:10.1111/psyp.12769

100. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL (2007) Open Access Series of Imaging Studies (OASIS): cross-sectional MRI Data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 19:1498–1507. doi:10.1162/jocn.2007.19.9.1498

101. Gorgolewski KJ, Mendes N, Wilfing D, Wladimirow E, Gauthier CJ et al (2015) A high resolution 7-Tesla resting-state fMRI dataset with cognitive and physiological measures. Sci Data 2:140054. doi:10.1038/sdata.2014.54

102. Landman BA, Huang AJ, Gifford A, Vikram DS, Lim IAL, Farrell JAD et al (2011) Multi-parametric neuroimaging reproducibility: a 3-T resource study. NeuroImage 54:2854–2866. doi:10.1016/j.neuroimage.2010.11.047

103. Gorgolewski KJ, Storkey AJ, Bastin ME, Whittle I, Pernet C (2013) Single subject fMRI test–retest reliability metrics and confounding factors. NeuroImage 69:231–243. doi:10.1016/j.neuroimage.2012.10.085

104. Maclaren J, Han Z, Vos SB, Fischbein N, Bammer R (2014) Reliability of brain volume measurements: a test–retest dataset. Sci Data 1:140037. doi:10.1038/sdata.2014.37

105. Poldrack RA, Laumann TO, Koyejo O, Gregory B, Hover A, Chen M-Y et al (2015) Long-term neural and physiological phenotyping of a single human. Nat Commun 6:8885. doi:10.1038/ncomms9885

106. Laumann TO, Gordon EM, Adeyemo B, Snyder AZ, Joo SJ, Chen M-Y et al (2015) Functional system and areal organization of a highly sampled individual human brain. Neuron 87:657–670. doi:10.1016/j.neuron.2015.06.037

107. Choe AS, Jones CK, Joel SE, Muschelli J, Belegu V, Caffo BS et al (2015) Reproducibility and temporal structure in weekly resting-state fMRI over a period of 3.5 years. PLoS ONE 10:e0140134. doi:10.1371/journal.pone.0140134

108. Froeling M, Tax CMW, Vos SB, Luijten PR, Leemans A (in press) "MASSIVE" brain dataset: multiple acquisitions for standardization of structural imaging validation and evaluation. Magn Reson Med. doi:10.1002/mrm.26259

109. Orban P, Madjar C, Savard M, Dansereau C, Tam A, Das S et al (2015) Test–retest resting-state fMRI in healthy elderly persons with a family history of Alzheimer's disease. Sci Data 2:150043. doi:10.1038/sdata.2015.43

110. Lin Q, Dai Z, Xia M, Han Z, Huang R, Gong G et al (2015) A connectivity-based test–retest dataset of multi-modal magnetic resonance imaging in young healthy adults. Sci Data 2:150056. doi:10.1038/sdata.2015.56

111. Huang L, Huang T, Zhen Z, Liu J (2016) A test–retest dataset for assessing long-term reliability of brain morphology and resting-state brain activity. Sci Data 3:160016. doi:10.1038/sdata.2016.16

Christopher R. Madan is currently a Post-Doctoral Research Fellow at Boston College. He received his Ph.D. in Psychology from the University of Alberta (Canada), and previously was a visiting scientist at the University Medical Center Hamburg-Eppendorf (Germany). Chris primarily studies human memory, using a combination of cognitive psychology, neuroimaging, and computational modeling. Recently, Chris has extended his research interests to include inter-individual differences in brain morphology and methods development for characterizing structural properties of the brain.

Elizabeth A. Kensinger is a Professor of Psychology at Boston College. She received her Ph.D. in Neuroscience from the Massachusetts Institute of Technology prior to completing a postdoctoral fellowship at Harvard University and Massachusetts General Hospital. Her laboratory combines psychological, physiological, and neuroimaging measures to investigate the links between emotion and cognition and to examine how those links change as adults age.