# The within-subject stability of cortical thickness, surface area, and brain volumes across one year

Meng-Yun Wang[1,2], Max Korbmacher[2,3], Rune Eikeland[1,2], Stener Nerland[4,5], Didac Vidal Pineiro[6], Karsten Specht[1,2,7]

1. Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway

2. Mohn Medical Imaging and Visualization Centre (MMIV), Haukeland University Hospital, Bergen, Norway

3. Department of Health and Functioning, Western Norway University of Applied Sciences, Bergen, Norway

4. Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway

5. Norwegian Center for Mental Disorders Research (NORMENT), Institute of Clinical Medicine, University of Oslo, Oslo, Norway

6. Center for Lifespan changes in brain and cognition, Department of Psychology, University of Oslo, Oslo, Norway

7. Department of Education, UiT The Arctic University of Norway, Tromsø, Norway

*Correspondence: Karsten Specht, karsten.specht@uib.no or Meng-Yun Wang mengyun.wang@uib.no

**Abstract:**

With the feature of noninvasively monitoring the human brain, magnetic resonance imaging (MRI) has become a ubiquitous means to understand how the brain works. Specifically, T1-weighted (T1w) imaging is widely used to examine the brain structure where the cortical thickness, surface area, and brain volumes have been investigated. These T1w-derived phenotypes undergo radical changes during childhood and adolescence, while remaining relatively stable during adulthood. However, stability over a short time (e.g. one year) during adulthood is still unknown. Additionally, how environmental factors such as time-of-day and different daylight lengths could impact the structural brain is also elusive. The main purpose of this study, therefore, was to assess the stability of T1w-derived phenotypes, i.e., cortical thickness, surface area, and brain volumes including subcortical volumes, and to explore the time-of-day and daylight length effects. Accordingly, three subjects in their late 20s, and early 30s and 40s were scanned repeatedly on the same scanner over one year from which a deep brain imaging dataset was constructed with 38, 40, and 25 sessions for subjects 1, 2, and 3, respectively. The T1w-derived phenotypes demonstrated percentage changes within 5% and CVs (coefficients of variance) within 2% for the majority of brain regions. However, several brain regions did show larger variations with percentage changes around 10% and CVs around 5%, such as the temporal pole, the frontal pole, and the entorhinal cortex. More importantly, there were no significant effects of time-of-day and daylight length. Moreover, cortical thickness change was strongly and positively correlated with that of volume while being negatively correlated with that of surface area, illustrating their distinct roles in brain anatomy. Additionally, it was found that apparent head motion causes cortical thickness and volume to be underestimated and surface area to be overestimated. These results indicate that T1w-derived phenotypes are reasonably stable regardless of time-of-day or daylight length, but that head motion should be taken into consideration.

**Significance:**

Assessing the measurement precision and within-subject stability of T1w-derived phenotypes is crucial for accurately estimating brain changes induced by treatments or interventions. Furthermore, understanding within-subject variation enhances our ability to predict behavior and associations with brain phenotypes, which rely heavily on between-subject variation.

**Keywords:** dense sampling; cortical thickness; surface area; volume; T1w; time-of-day; daylight length

## Introduction

Our understanding of how the human brain works has been substantially deepened over the last few decades (Finn, Poldrack, & Shine, 2023; Lerch et al., 2017; Poldrack & Farah, 2015). Specifically, magnetic resonance imaging (MRI), with the capability of obtaining *in vivo* brain images, has revolutionized our understanding of the brain structure (Lerch et al., 2017), function (Finn et al., 2023), and their associations with behavior (Genon, Eickhoff, & Kharabian, 2022; Wu, Li, Eickhoff, Scheinost, & Genon, 2023), genetics (Chen et al., 2012; Elliott et al., 2018; Grasby et al., 2020), and environment (Holz et al., 2023; Thompson et al., 2020). Without any doubt, MRI has become a ubiquitous means of investigating the human brain both in basic brain research and clinical settings.

MRI can provide various brain imaging modalities by exploiting different MRI sequences (Lerch et al., 2017), among which T1-weighted (T1w) brain imaging has been widely used to explore cortical and subcortical brain characteristics. For example, the phenotypes derived from the T1w images have been utilized to establish normal brain development trajectories (Bethlehem et al., 2022; Lemaitre et al., 2012; Sowell et al., 2003; Sowell et al., 2004; Vijayakumar et al., 2016) and uncover brain abnormalities (Frisoni, Fox, Jack, Scheltens, & Thompson, 2010). The T1w-derived phenotypes can be constructed with the surface-based algorithm, which has been predominantly used in structural brain research including several big data initiatives, such as the UK Biobank (Alfaro-Almagro et al., 2018), the Human Connectome Project (Glasser et al., 2013), and the Adolescent Brain Cognitive Development Study (Hagler et al., 2019). Therefore, the surface-based analysis will also be utilized in this study. Simply put, the surface-based algorithm will construct the pial and white matter surfaces in the subject's native space, with which the phenotypes including the cortical thickness, area, and volumes can be calculated (Dale, Fischl, & Sereno, 1999; Fischl & Dale, 2000; Fischl, Sereno, & Dale, 1999).

These three abovementioned T1w-derived phenotypes have been ubiquitously used to investigate longitudinal brain development (Bethlehem et al., 2022), especially in childhood (Gilmore, Knickmeyer, & Gao, 2018; Sowell et al., 2004), adolescence (Sowell et al., 2004; Vijayakumar et al., 2016), and the elderly (Lemaitre et al., 2012; Sele, Liem, Merillat, & Jancke, 2021; Storsve et al., 2014) where the brain changes rapidly. Almost no studies, however, have focused on brain change during adulthood, which is reasonable since almost all the brain features will reach the plateau at this time and not change dramatically (Bethlehem et al., 2022). For example, the human cortical thickness peaks at 1.7 years (Bethlehem et al., 2022) which varies between 1 and 4.5 mm with an overall average of approximately 2.5 mm (Fischl & Dale, 2000). Furthermore, the total surface area peaks at 10.97 years while cortical and subcortical gray matter volume peaks at 5.9 years and 14.4 years, respectively (Bethlehem et al., 2022). After the peaks, the three T1-derived phenotypes remain relatively stable during adulthood and gradually decline in late life. Although it is considerably stable, the degree of stability across a short time (e.g. across one year) at adulthood is still largely unknown. Additionally, the influence of external environmental factors on the structural brain is still elusive.

Although time-of-day effects have been documented in functional brain organization studies (Orban, Kong, Li, Chee, & Yeo, 2020), the impact on brain structure is still unclear. Some studies have reported associations between time-of-day and cortical measures (Karch et al., 2019; Nakamura et al., 2015; Trefler et al., 2016). Specifically, the cortical thickness decreased from morning to afternoon (Trefler et al., 2016). In the same vein, the brain volume is larger during the morning than during the afternoon (Karch et al., 2019; Nakamura et al., 2015; Trefler et al., 2016). However, two of the studies employed a cross-sectional

design where between-subject variation could interfere with the reported time-of-day effect (Nakamura et al., 2015; Trefler et al., 2016). For example, in one study, each subject only had two data points for morning and afternoon sessions (Trefler et al., 2016), and the other just pooled the data from other datasets and separated them based on the collection time (Nakamura et al., 2015). Although one is within-subject design (Karch et al., 2019), they did not use the longitudinal pipeline as recommended (Reuter, Schmansky, Rosas, & Fischl, 2012), which could hinder the robustness and reliability of the results. More critically, none of the studies showed the data distribution of the morning and afternoon sessions. Therefore, in this study, within-subject design together with the longitudinal processing pipeline will be adopted meanwhile the data distribution will be provided.

Besides the time-of-day effect, the seasonal effect on functional brain organization has recently attracted some attention (M. Y. Wang, Korbmacher, Eikeland, & Specht, 2023b; Zhang, Shokri-Kojori, & Volkow, 2023). Specifically, daylight length comes out as a potential contributor (Di, Woelfer, Kühn, Zhang, & Biswal, 2022) to the functional brain organization (M. Y. Wang et al., 2023b). For example, the salience network, one of the resting-state brain networks, is larger when daylight length is short compared to when it is long (M. Y. Wang et al., 2023b). To the best of our knowledge, there is still no study investigating the daylight length effect on brain structure. Located in the Arctic Circle, the daylight length in Norway varies substantially throughout the year. Specifically, the daylight length of the place in Bergen can be as short as around 2 hours daylight and as long as around 16 hours daylight, which makes it suitable for exploring the daylight length effect.

The main aim of this study was to quantify the stability of cortical thickness, surface area, and brain volumes across one year, and to investigate the effects of time-of-day and daylight length. For this purpose, and to mitigate the influence of different scanners, field strengths (Han et al., 2006; Iscan et al., 2015), different processing platforms, software versions (Gronenschild et al., 2012), and inter-subject interference, three subjects were repeatedly scanned in this study across one year in the same scanner at the same location (M. Y. Wang, Korbmacher, Eikeland, & Specht, 2022, 2023a) where the data were processed with the latest released version of the software (FreeSurfer 7.2) on the same computer (M. Y. Wang et al., 2023a). In addition, the image quality metrics were included as covariates since it is shown that image quality can affect measurements (Ducharme et al., 2016; Reuter et al., 2015).

## Methods

### Participants

Three participants (**Table 1**) were scanned as part of a precision brain mapping project named the Bergen Breakfast Scanning Club (BBSC) project (Korbmacher et al., 2023; M. Y. Wang, Korbmacher, Eikeland, Craven, & Specht, 2024; M. Y. Wang et al., 2022, 2023a). The BBSC project aims to illustrate the individual precision brain networks and to examine the stability of MRI measurements over one year. For this purpose, the participants were scanned twice a week between February 2021 and February 2022 with two breaks (Jun. to Oct. 2021, and Jan. 2022) where functional and structural MRI data (M. Y. Wang et al., 2022, 2023a) were collected. In total, there are 38, 40, and 25 T1w MRI sessions for subjects 1, 2, and 3, respectively. The exact dates of these sessions can be found in **Table S1.**

All three participants speak at least two languages (their native languages, and English). Notably, the first participant has also been acquiring an additional language since January 2021 and remained COVID-19 free during the period of data collection. The second participant contracted COVID-19 around December 2021, while the third participant had COVID-19 in approximately August 2021.

**Table1**. Basic demographic information of subjects

| | Sub1 | Sub2 | Sub3 |
|---|---|---|---|
| **Gender** | Male | Male | Male |
| **Onset Age** | 31 | 27 | 40 |
| **Laterality** | Right | Right | Right |
| **Regular caffeine consumption?** | No | Yes | Yes |
| **Regular nicotine consumption?** | No | No | No |

## Data collection

Data collection was embedded in the functional protocol of the BBSC project (M. Y. Wang et al., 2022, 2023a), which lasted around 25 minutes in total. The procedure of the data collection in the functional protocol is seven mins T1w MRI, 5 mins MR spectroscopy (MRS) data, and 12 mins rs-fMRI, where the data were collected with a 3T MR scanner (GE Discovery MR750) with a 32-channel head coil at the Haukeland University Hospital in Bergen, Norway. The other modalities in the BBSC project have been described previously (M. Y. Wang et al., 2023a).

Seven-minute structural T1w images were acquired using a 3D Fast Spoiled Gradient-Recalled Echo (FSPGR) sequence with the following parameters: 188 contiguous slices acquired, with repetition time (TR) = 6.88 ms, echo time (TE) = 2.95 ms, FA (flip angle) = 12°, slice thickness = 1 mm, in-plane resolution = 1 mm × 1 mm, and field of view (FOV) = 256 mm, with an isotropic voxel size of 1 $mm^3$.

## Data processing

Cortical thickness, area, and volume were computed with FreeSurfer 7.2 (freesurfer-darwin-macOS-7.2.0-20210713-aa8f76b) (Dale et al., 1999; Fischl et al., 1999). First, each session was processed with the cross-sectional *recon-all* pipeline. Then, the results went through the longitudinal pipeline (Reuter et al., 2012), where an unbiased within-subject template space and image are created by using a robust, inverse consistent registration (Reuter, Rosas, & Fischl, 2010). For more details, please refer to the data processing part in the **Supplementary Material**.

It has become common practice to report brain imaging results in a common brain atlas (Desikan et al., 2006; Yeo et al., 2011) to aggregate or contrast findings across different studies, among which the Desikan- Killiany (DK) atlas (Desikan et al., 2006) has become the common atlas used in the T1w brain imaging studies (Alfaro-Almagro et al., 2018; Bethlehem et al., 2022; Ducharme et al., 2016; Grasby et al., 2020; Hagler et al., 2019; Iscan et al., 2015; Lemaitre et al., 2012; Sele et al., 2021; Storsve et al., 2014; Vijayakumar et al., 2016). The DK atlas was developed based on curvature-based information such as the sulcal/gyral representations and contains 34 cortical brain regions in each hemisphere (Desikan et al., 2006). In this study, therefore, cortical thickness, surface area, and volume were extracted from 68 brain regions. Additionally, 8 subcortical regions (thalamus, pallidum, amygdala, hippocampus, putamen, accumbens area, caudate nucleus, and lateral ventricles) were used for the subcortical brain volume analysis.

Moreover, the T1w images were quality assessed with MRIQC 23.1.0 (Esteban et al., 2017), where the following image quality metrics (IQMs) were used (**Fig. 1AB**): contrast-to-noise

ratio (CNR); signal-to-noise ratio (SNR); white-matter to maximum intensity ratio (WM2MAX); coefficient of joint variation (CJV); entropy focus criterion (EFC); intensity non-uniformity (INU); foreground-to-background energy ratio (FBER); full-width half maximum (FWHM); residual partial volume effect feature (rPVE).

CNR measures the difference in mean intensities between gray and white matter, normalized by the standard deviation of non-brain areas, with higher values indicating better contrast between gray and white matter (Magnotta, Friedman, & First, 2006). SNR quantifies the mean intensity within gray matter relative to the standard deviation outside the brain, where higher values denote higher quality (Magnotta et al., 2006). WM2MAX estimates the median white matter intensity against the 95th percentile of the entire image intensity, pinpointing hyperintensities such as those in carotid vessels and fat with ideal values falling between 0.6 and 0.8 (Esteban et al., 2017).

CJV was proposed as a measure for INU correction algorithms, higher CJV indicates head motion or INU artifacts, whereas lower values imply better quality (Ganzetti, Wenderoth, & Mantini, 2016). EFC uses Shannon entropy to measure image blurring due to head motion with lower values reflecting a superior image clarity (Atkinson, Hill, Stoyle, Summers, & Keevil, 1997). INU summary statistics (max, min and median) of the INU field (bias field) as estimated with the N4ITK algorithm (Tustison et al., 2010). Values closer to 1.0 are better, values further from zero indicate greater field inhomogeneity.

FBER is defined as the mean energy of image values within the head relative to that outside the head with higher values indicating better images. The FWHM of the spatial distribution represents the image intensity values in voxel units where lower values are better with higher values indicating a blurrier image (Forman et al., 1995). The rPVE is a tissue-wise sum of partial volumes that fall in the range [5-95%] of the total volume of a pixel, computed on the partial volume maps, in which a lower rPVE is better (Esteban et al., 2017).

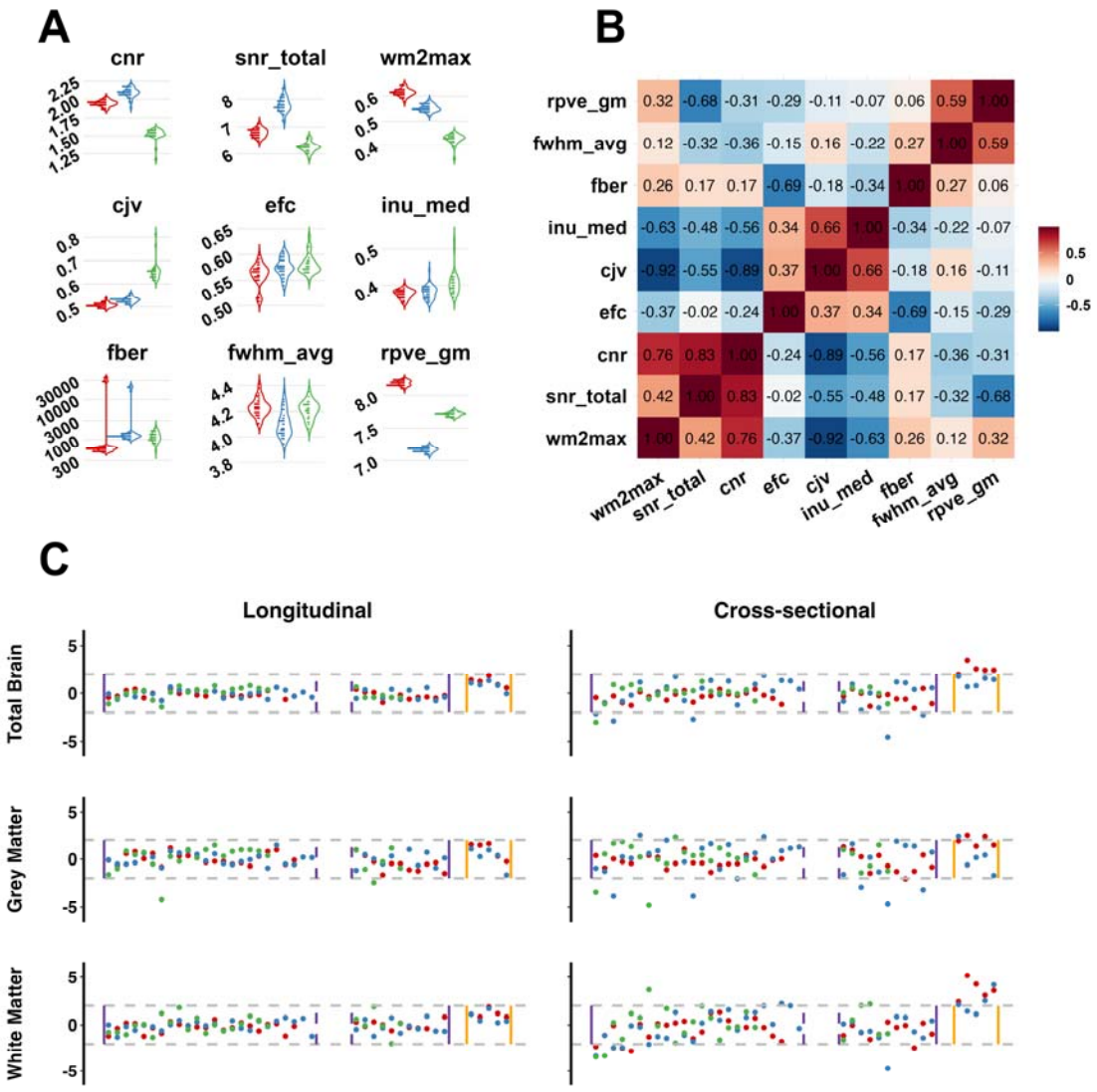



**Fig. 1. Distributions and correlations of IQMs and percentage changes of different brain tissues from different pipelines.** **(A)** Violin plots of the IQM distributions for each subject. **(B)** Correlation matrix between the IQMs, where the cnr, snr_total, and wm2max form one cluster representing the image intensity, the cjv, efc, and inu_med constructs another cluster representing head motion or related artifacts, and the fber, fwhm_avg, and rpve_gm form the last cluster representing the technique qualities. **(C)** Percentage changes along the data collections generated from the longitudinal and cross-sectional pipelines. The x-axis represents the time sequence of data collection while the y-axis depicts the percentage change. The gray dashed lines denote 2% percentage change. For subplots (A) and (C), the red color represents subject 1, the blue depicts subject 2, and the green represents subject 3. Contrast-to-noise ratio (CNR); signal-to-noise ratio (SNR); white-matter to maximum intensity ratio (wm2max); coefficient of joint variation (CJV); entropy-focus criterion (EFC); intensity non-uniformity (INU); foreground-to-background energy ratio (FBER); full-width half-maximum (FWHM); residual partial volume effect (rPVE).

## Data Analysis

### *Coefficient of variance:*

The coefficient of variation (CV) is utilized to assess the consistency of T1w-derived brain phenotypes across different testing sessions:

$$CV(\%) = \left(\frac{std}{mean}\right) * 100$$

where *std* denotes the standard deviation of the measurements taken during different sessions for the same subject, while *mean* refers to their average.

CVs facilitate comparison of variability across varying units or scales. It has also been instrumental in assessing the stability of measurements over some time (Borga et al., 2020; Carbonell et al., 2022; M. Y. Wang et al., 2024; Y. Wang et al., 2021), where it is especially suitable for the within-subject design (M. Y. Wang et al., 2024). Lower CVs indicate less variability relative to the mean, signifying higher precision or consistency. Conversely, higher CVs indicate more variability relative to the mean, suggesting lower precision or consistency (M. Y. Wang et al., 2024).

***Linear mixed-effects models***

To evaluate the impact of time-of-day and daylight length on various brain metrics such as cortical thickness, surface area, and brain volumes, linear mixed-effects models were employed. These models designated cortical measurements as dependent variables, influenced by the independent variables of time-of-day and daylight length. To account for discrepancies among different brain regions, we used percentage changes of each T1-weighted (T1w) derived phenotype, calculated as the deviation from the mean divided by the mean itself.

Time-of-day was classified based on the data collection time, labeling sessions before noon as 'morning' and those after as 'afternoon.' The distribution of sessions was as follows: in sub1, there were 22 morning and 16 afternoon sessions; in sub2, 23 morning and 16 afternoon sessions; and in sub3, 15 morning and 10 afternoon sessions. Daylight length data, ranging from 120 to 948 minutes, was obtained from the weather station at Florida maintained by the University of Bergen (https://veret.gfi.uib.no/). The time-of-day indices and daylight length have been described in **Table S1**.

To control the influence of the IQMs on the stability of T1w-derived phenotypes, the IQMs were used as covariates in the linear mixed-effects models. In addition, to alleviate the different scales of IQMs, the IQMs were normalized by subtracting the mean and being divided by the standard deviation.

Then, the following model was used for the cortical thickness, surface area, and brain volumes within each subject:

$$\begin{aligned} &percentagechange \\ &= \beta_0 + \beta_1\left(time_{effect}\right) + \beta_2(region) + \beta_3\left(time_{effect} \times region\right) \\ &+ \beta_{4-12}(IQMs) + u(hemi) + \epsilon \end{aligned}$$

Where, $\beta_0$ is the intercept; $\beta_1$ to $\beta_{12}$ are the fixed effect coefficients for each predictor; *time_effect* and *region* are fixed effects, with an interaction term included to examine the combined influence of day length and regional differences on cortical thickness; *time_effect* is time-of-day or daylight length; *IQMs* are included as covariates to control for various imaging and physiological parameters; $u$ (hemi) representing the random intercepts for each level of the hemisphere, and $\epsilon$ is the residual error.

To control for multiple comparisons, the modified Bonferroni correction method (Holm, 1979) was applied. Statistical analyses and visualizations were done with R 4.3.2 (R Core Team, 2022), where the regression was modeled with lme4 (v1.1-35.1) (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (v3.1-3) (Kuznetsova, Brockhoff, & Christensen,

2017), and visualization was performed with ggplot2 (v3.5.0) (Wickham, 2016), ggseg (v1.6.6) (Mowinckel & Vidal-Piñeiro, 2020), and patchwork (v1.2.0).

## Results

### Longitudinal vs cross-sectional pipelines

The CVs of longitudinal and cross-sectional processing pipelines from FreeSurfer are described in **Table 2**. Meanwhile, the percentage changes against the average of different brain tissues for each subject are illustrated in **Fig. 1C**.

As shown in **Table 2 and Fig. 1C**, the CVs and percentage changes generated by the longitudinal processing pipeline are smaller than those of the cross-sectional pipeline, which indicates the longitudinal pipeline generates more reliable results. Therefore, the following results are derived from the longitudinal pipeline.

**Table 2. The longitudinal pipeline generates smaller CVs**

| | Sub1 | | Sub2 | | Sub3 | |
|---|---|---|---|---|---|---|
| | Longitudinal | Cross-sectional | Longitudinal | Cross-sectional | Longitudinal | Cross-sectional |
| **Total Brain** | 0.57 | 1.13 | 0.50 | 1.47 | 0.58 | 0.85 |
| **Gray Matter** | 0.81 | 1.06 | 0.71 | 1.86 | 1.31 | 1.60 |
| **Left - GM** | 0.89 | 1.17 | 0.80 | 1.97 | 1.40 | 1.61 |
| **Right - GM** | 0.91 | 1.13 | 0.74 | 1.84 | 1.35 | 1.69 |
| **White Matter** | 0.66 | 1.79 | 0.65 | 1.74 | 0.90 | 1.70 |
| **Left - WM** | 0.69 | 1.86 | 0.63 | 1.76 | 1.20 | 1.72 |
| **Right - WM** | 0.76 | 1.77 | 0.74 | 1.76 | 0.87 | 1.81 |

All CVs are reported as percentages, where smaller is better.
Total Brain: '*BrainSegVolNotVent*' generated from FreeSurfer.

### Cortical Thickness

#### *Mean thickness for each brain region across sessions*

The mean cortical thicknesses of the three subjects were 2.59 mm, 2.70 mm, and 2.58 mm, respectively with a shared standard deviation of 0.02 mm (**Table S2)**. In addition, subject 1 showed a range of 1.59 to 4.04 mm, subject 2 with a range of 1.83 to 3.75 mm, and subject 3 with a range of 1.50 to 4.03 mm (**Table S2)**. Moreover, analysis of the cortical thickness distribution across various brain regions revealed a consistent ranking of regions within subjects, identifying the insula, entorhinal cortex, and temporal pole as the thickest regions for all subjects (**Fig. 2AB** and **Table S2**). Conversely, regions such as the lingual gyrus, cuneus, and pericalcarine cortex were consistently ranked as the thinnest across the subjects, as also shown in **Fig. 2AB** and **Table S2**. Furthermore, standard deviations ranged from 0.02 to 0.15 mm, where the thickest brain regions had the largest standard deviations (**Table S2**).

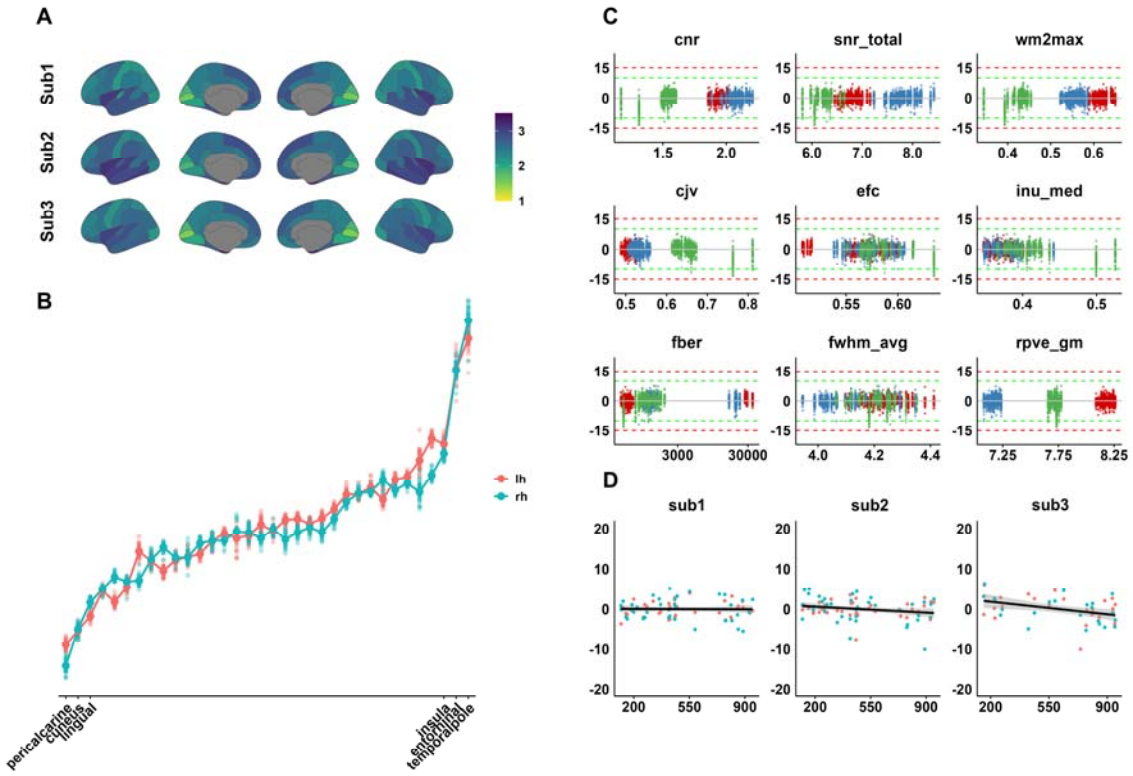



**Fig. 2. Mean cortical thickness, percentage change distributions along with IQMs, and percentage changes along the daylight length. (A)** Average cortical thickness across all sessions of different brain regions for each subject where the unit is a millimeter. **(B)** Ranked brain regions are illustrated from the thinnest to the thickest (example data from sub1). All three subjects have the same the most three thinnest and thickest brain regions although the brain regions ranking in the middle vary. **(C)** The distribution of the percentage changes along with different IQMs, where the red, blue, and green dots represent subjects 1, 2, and 3, respectively. **(D)** Percentage change of thickness along different daylight lengths in the pericalcarine region for sub1, sub2, and sub3, respectively. No significant results for sub1 and sub2 whereas sub3 showed significant results.

### *The stability and percentage changes for each brain region*

The distribution of percentage changes along the IQMs and percentage change is illustrated in **Fig. 2C**, which indicated that sub3 had two sessions with considerable apparent head motion evidenced by the larger CJV, EFC, INU, and smaller CNR values. These two sessions could lead to the larger percentage changes and CVs in sub3 as illustrated in **Fig. 3**. To confirm the assumption, the two sessions with excessive head motion indicated by the CJV (sessions 1 and 7) were excluded, after which the CVs (**Table 3**) and percentage changes (Sub3_qc in **Fig. 3**) significantly improved (left hemi: $t_{33}$ = -7.6, $p = 4.8 \times 10^{-9}$, *Cohen's d* = 1.3; right hemi: $t_{33}$ = -7.13, $p = 1.8 \times 10^{-8}$, *Cohen's d* = 1.22).

After quality control, the CVs of within-subject thickness ranged from 0.67 to 4.40, from 0.74 to 2.64 for sub1, 0.67 to 3.24 for sub2, and 0.84 to 4.40 for sub3 (**Table 3**). Furthermore, the percentage changes in almost all brain regions across the three subjects were well-constrained within the 5% as illustrated in **Fig. 3.** Several brain regions, however, showed larger CVs and percentage changes across all three subjects (**Fig. 3** and **Table 3**), such as the temporal pole, frontal pole, pericalcarine, and entorhinal cortex.

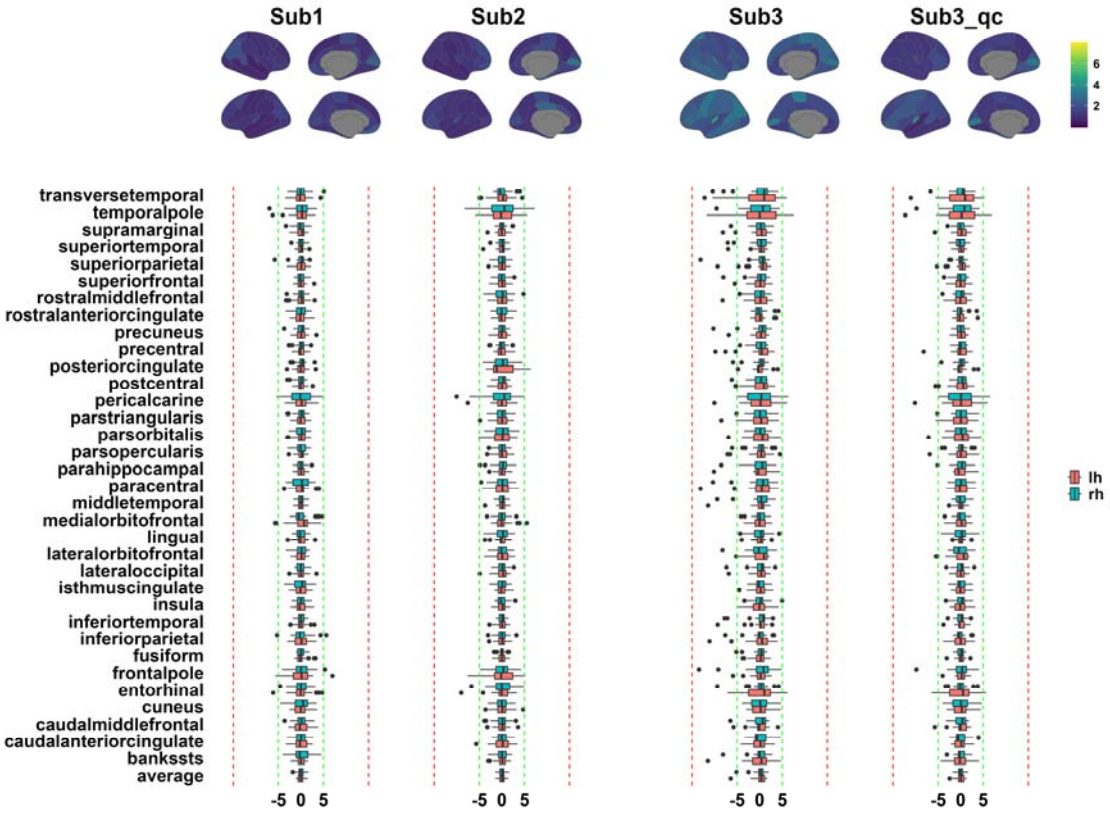



**Fig. 3. The CVs and percentage changes in the cortical thickness.** The upper panel depicts the coefficient of the variant (CV) for each subject. The lower boxplots show percentage change against the average for each brain region and each subject. Sub3 represents the results from the original data while sub3_qc shows the results from the data after excluding two sessions with excessive head motion. The green lines are in the range of ±5 % while the red lines depict the range of ±15%. All values in the boxplot are given as percentages.

**Table 3. The CVs of the cortical thickness based on the DK atlas**

| | Sub1 | | Sub2 | | Sub3 | | | |
|---|---|---|---|---|---|---|---|---|
| | Left | Right | Left | Right | Left | Left_qc | Right | Right_qc |
| **transversetemporal** | 1.59 | 1.45 | 1.90 | 1.40 | 4.61 | 3.94 | 3.43 | 2.05 |
| **temporalpole** | 1.85 | 2.25 | 2.72 | 3.24 | 4.71 | 4.40 | 3.26 | 3.28 |
| **supramarginal** | 1.00 | 1.20 | 1.07 | 0.92 | 2.45 | 1.81 | 1.91 | 1.32 |
| **superiortemporal** | 0.76 | 0.74 | 0.98 | 0.85 | 1.77 | 0.92 | 2.31 | 1.30 |
| **superiorparietal** | 1.21 | 1.33 | 1.11 | 1.18 | 2.64 | 1.72 | 2.97 | 1.04 |
| **superiorfrontal** | 0.98 | 0.92 | 1.11 | 1.14 | 1.78 | 1.42 | 2.19 | 1.43 |
| **rostralmiddlefrontal** | 1.38 | 0.98 | 1.46 | 1.74 | 2.35 | 1.58 | 1.74 | 1.46 |
| **rostralanteriorcingulate** | 1.36 | 1.38 | 1.21 | 1.51 | 1.35 | 1.22 | 1.47 | 1.23 |
| **precuneus** | 1.33 | 1.09 | 1.12 | 1.00 | 1.90 | 1.28 | 2.66 | 1.15 |
| **precentral** | 0.92 | 1.06 | 1.11 | 1.05 | 3.20 | 2.22 | 1.61 | 1.16 |
| **posteriorcingulate** | 1.16 | 1.13 | 2.67 | 1.88 | 1.62 | 1.27 | 1.78 | 1.33 |
| **postcentral** | 0.91 | 0.97 | 1.20 | 1.13 | 2.52 | 2.26 | 1.90 | 1.25 |
| **pericalcarine** | 1.76 | 2.64 | 1.86 | 3.16 | 3.43 | 3.41 | 3.16 | 3.25 |
| **parstriangularis** | 1.07 | 1.18 | 1.66 | 1.41 | 2.80 | 2.22 | 2.02 | 1.91 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **parsorbitalis** | 1.24 | 1.35 | 2.07 | 1.97 | 2.62 | 2.60 | 1.74 | 1.81 |
| **parsopercularis** | 0.98 | 1.20 | 1.32 | 1.16 | 2.64 | 2.26 | 2.18 | 1.81 |
| **parahippocampal** | 0.95 | 0.99 | 1.18 | 1.76 | 2.63 | 1.58 | 2.55 | 1.77 |
| **paracentral** | 1.67 | 1.85 | 1.89 | 1.71 | 3.43 | 1.70 | 2.97 | 1.54 |
| **middletemporal** | 0.83 | 0.86 | 0.97 | 0.91 | 2.65 | 1.17 | 2.67 | 1.29 |
| **medialorbitofrontal** | 2.39 | 1.77 | 1.70 | 1.31 | 1.51 | 1.48 | 1.63 | 1.44 |
| **lingual** | 1.16 | 1.14 | 1.12 | 1.62 | 1.39 | 1.06 | 2.09 | 2.14 |
| **lateralorbitofrontal** | 1.23 | 1.36 | 1.67 | 1.33 | 2.24 | 2.14 | 2.63 | 1.91 |
| **lateraloccipital** | 1.25 | 1.05 | 1.36 | 1.04 | 1.98 | 1.28 | 2.06 | 1.34 |
| **isthmuscingulate** | 1.48 | 1.75 | 1.04 | 1.16 | 1.65 | 1.25 | 1.30 | 1.24 |
| **insula** | 1.16 | 1.15 | 1.09 | 1.20 | 2.04 | 1.64 | 1.73 | 1.38 |
| **inferiortemporal** | 1.10 | 0.88 | 1.09 | 0.72 | 2.32 | 1.22 | 2.49 | 1.10 |
| **inferiorparietal** | 1.56 | 1.95 | 1.05 | 1.16 | 3.17 | 1.71 | 2.17 | 1.23 |
| **fusiform** | 1.08 | 0.80 | 0.85 | 0.67 | 1.53 | 1.12 | 1.59 | 0.84 |
| **frontalpole** | 2.28 | 2.11 | 2.94 | 2.14 | 1.83 | 1.30 | 4.01 | 2.87 |
| **entorhinal** | 2.03 | 1.55 | 2.19 | 2.43 | 3.59 | 3.16 | 2.49 | 1.60 |
| **cuneus** | 1.51 | 2.02 | 1.62 | 1.47 | 1.91 | 1.88 | 2.27 | 2.27 |
| **caudalmiddlefrontal** | 1.69 | 1.23 | 1.36 | 1.45 | 2.04 | 1.88 | 1.92 | 1.41 |
| **caudalanteriorcingulate** | 1.62 | 1.44 | 1.86 | 1.34 | 1.77 | 1.44 | 2.06 | 1.76 |
| **bankssts** | 0.93 | 1.88 | 1.17 | 1.17 | 3.14 | 2.07 | 2.12 | 1.23 |
| **average** | 0.72 | 0.66 | 0.68 | 0.68 | 1.67 | 0.95 | 1.45 | 0.83 |

Percentage CVs after excluding the two sessions with excessive head motion.

### *Time-of-Day*

The linear mixed-effects model did not generate any statistically significant results regarding the time-of-day effect within each subject in our dataset. The percentage changes for each brain region stratified by morning and afternoon sessions were illustrated in **Fig. S1**.

### *Daylight length*

The linear mixed-effects model did not generate any statistically significant results regarding the daylight length effect within sub1 and sub2, but sub3 showed brain thickness decline (**Fig. 2D**) in the pericalcarine ($R^2$ = 0.15, $\beta$ = -1.46, $t$ = -4.09, $p = 3.55 \times 10^{-3}$) along with the increase of daylight length.

## Surface Area

### *Mean brain area across sessions for each brain region*

The total surface areas of one brain hemisphere are around 850, 912, and 908 $cm^2$ for sub1, sub2, and sub3, respectively. Additionally, the range of the surface area of one brain region in each brain hemisphere is from 219.82 to 6815.53 $mm^2$ for sub1, 298.30 to 8046.82 $mm^2$ for sub2, and 248.83 to 8594.52 $mm^2$ for sub3 (**Table S3**). The superior parietal, rostral middle frontal, and superior frontal cortices emerged as the consistently largest surface areas in all subjects (**Fig. 4AB**). In contrast, the entorhinal, transverse temporal, and frontal pole regions were invariably the smallest (**Fig. 4AB** and **Table S3**).

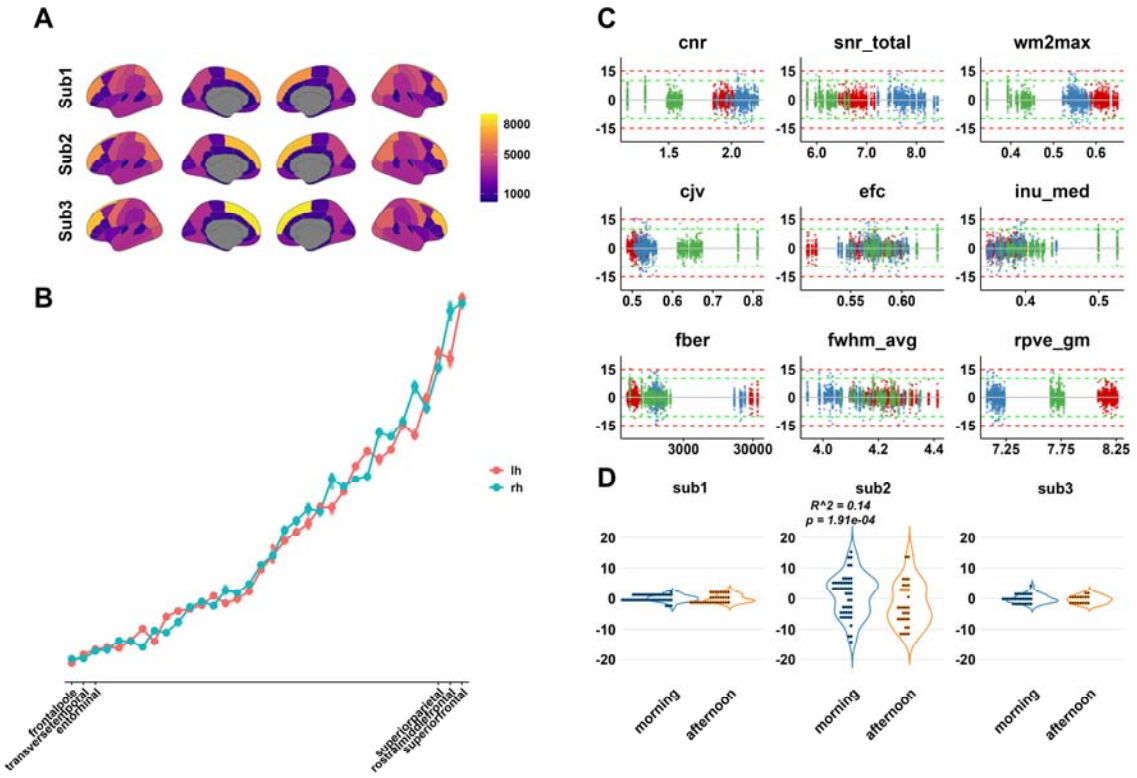



**Fig. 4. Mean surface area, percentage change distribution along with IQMs, and longitudinal and time-of-day change. (A)** Average surface area of different brain regions for each subject and the unit is millimeters per square. **(B)** Surface areas from the smallest to the largest (data from sub1). The three smallest and largest surface areas were the same in all three subjects, but the surface areas ranking in the middle varied. **(C)** The distributions of surface area percentage changes along with the IQMs, where the red, blue, and green dots represent subjects 1, 2, and 3, respectively. **(D)** The distribution of morning and afternoon sessions of the posterior cingulate cortex for each subject.

### *The stability and percentage change for each brain region*

The distributions of percentage changes along with the IQMs are shown in **Fig. 4C**, where the distributions illustrated more skewed positive changes along with larger CJV and smaller CNR values in sub3. This could contribute to the large percentage change in surface area measurement in sub3 as illustrated in **Fig. 5**. Accordingly, if the outliers (sessions 1 and 7) are excluded, the CV (**Table 4**) and percentage changes (sub3_qc in **Fig. 5**) have been improved (left: $t_{33}$ = -6.03, $p = 4.4 \times 10^{-7}$, *Cohen's d* = 1.03; right: $t_{33}$ = -2.27, $p = 7.9 \times 10^{-5}$, *Cohen's d* = 0.73).

The CVs of the within-subject area range from 0.41 to 7.25, where 0.41 to 4.52 for Sub1, 0.49 to 7.25 for Sub2, and 0.53 to 5.34 for Sub3 (**Table 4**). Specifically, sub2 manifested a generally larger percentage change in the cingulate cortex (**Fig. 5**) and greater CVs (**Table 4**) than other brain regions along the data collections. Aside from the cingulate cortex in sub2 and those abovementioned three cortices with greater percentage changes, almost all brain regions showed percentage changes within 5% across all three subjects (**Fig. 5**). Notably, several surface areas showed larger CVs and percentage changes across all three subjects, including the temporal pole, frontal pole, and entorhinal cortex (**Fig. 5** and **Table 4**).

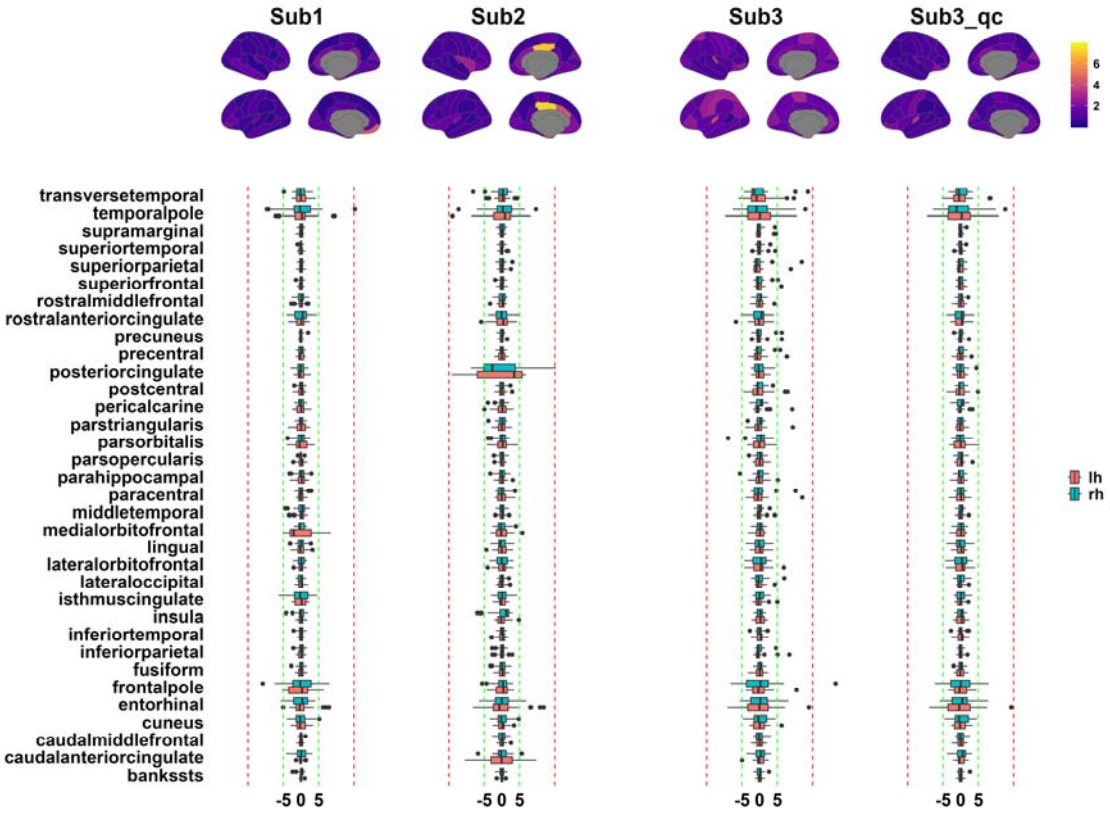



**Fig. 5 CVs and percentage changes in surface area.** The upper panel depicts the coefficient of the variant (CV) for each subject. The lower boxplot shows percentage changes for each brain region and each subject. Sub3 depicts the results generated from the original data while Sub_qc shows the results generated from the data after excluding sessions 1 and 7. The green lines are in the range of ±5 % while the red lines depict the range of ±15%. The unit of all values in the boxplot is percentage.

**Table 4. The CVs of the surface area based on the DK atlas**

| | Sub1 | | Sub2 | | Sub3 | | | |
|---|---|---|---|---|---|---|---|---|
| | Left | Right | Left | Right | Left | Left_qc | Right | Right_qc |
| **transversetemporal** | 1.56 | 1.77 | 1.99 | 2.26 | 4.04 | 3.42 | 4.04 | 1.91 |
| **temporalpole** | 3.50 | 4.83 | 4.58 | 4.01 | 4.72 | 4.81 | 5.04 | 5.13 |
| **supramarginal** | 0.61 | 0.65 | 0.63 | 0.82 | 1.40 | 0.53 | 1.16 | 0.74 |
| **superiortemporal** | 0.59 | 0.40 | 0.73 | 0.64 | 1.38 | 0.90 | 0.99 | 0.73 |
| **superiorparietal** | 0.72 | 0.59 | 0.82 | 0.92 | 2.79 | 0.91 | 2.76 | 0.91 |
| **superiorfrontal** | 0.56 | 0.55 | 0.76 | 0.76 | 1.54 | 0.88 | 1.59 | 0.91 |
| **rostralmiddlefrontal** | 1.10 | 1.24 | 1.01 | 1.21 | 1.36 | 1.05 | 1.03 | 0.92 |
| **rostralanteriorcingulate** | 1.46 | 2.24 | 2.15 | 2.27 | 2.38 | 1.83 | 2.06 | 1.80 |
| **precuneus** | 0.46 | 0.52 | 0.68 | 0.67 | 1.58 | 0.83 | 1.82 | 0.77 |
| **precentral** | 0.81 | 0.76 | 1.06 | 0.80 | 2.60 | 1.24 | 1.75 | 0.89 |
| **posteriorcingulate** | 1.18 | 1.24 | 7.25 | 6.94 | 1.42 | 1.29 | 1.78 | 1.54 |
| **postcentral** | 0.97 | 0.77 | 1.11 | 1.04 | 2.93 | 1.95 | 1.37 | 1.01 |
| **pericalcarine** | 1.38 | 0.94 | 1.70 | 1.11 | 2.31 | 1.13 | 1.39 | 1.43 |
| **parstriangularis** | 1.57 | 0.79 | 1.39 | 1.15 | 2.37 | 1.15 | 1.10 | 0.86 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **parsorbitalis** | 2.04 | 1.51 | 1.86 | 1.43 | 2.24 | 2.05 | 2.53 | 1.49 |
| **parsopercularis** | 0.91 | 0.39 | 0.77 | 0.98 | 1.33 | 1.24 | 1.25 | 1.05 |
| **parahippocampal** | 1.25 | 1.47 | 1.14 | 1.29 | 1.70 | 1.17 | 1.61 | 1.03 |
| **paracentral** | 0.78 | 1.03 | 1.52 | 1.31 | 2.98 | 1.56 | 2.65 | 1.23 |
| **middletemporal** | 1.22 | 1.66 | 0.86 | 0.78 | 1.30 | 0.95 | 0.97 | 0.80 |
| **medialorbitofrontal** | 4.52 | 0.99 | 2.09 | 1.61 | 1.44 | 1.44 | 1.23 | 1.27 |
| **lingual** | 1.39 | 1.30 | 1.50 | 1.65 | 1.53 | 1.33 | 1.99 | 2.01 |
| **lateralorbitofrontal** | 0.97 | 1.13 | 1.57 | 1.76 | 2.38 | 1.96 | 2.26 | 2.27 |
| **lateraloccipital** | 0.86 | 0.80 | 1.06 | 0.86 | 1.26 | 0.92 | 1.89 | 1.23 |
| **isthmuscingulate** | 1.60 | 2.46 | 1.23 | 1.53 | 1.56 | 1.05 | 1.41 | 1.41 |
| **insula** | 0.76 | 1.14 | 1.56 | 2.84 | 1.46 | 1.44 | 1.14 | 1.16 |
| **inferiortemporal** | 0.53 | 0.85 | 1.11 | 0.77 | 1.23 | 1.25 | 1.13 | 1.04 |
| **inferiorparietal** | 0.79 | 0.90 | 0.91 | 0.82 | 2.20 | 0.74 | 1.21 | 0.66 |
| **fusiform** | 0.95 | 0.89 | 1.08 | 1.34 | 1.37 | 1.23 | 0.91 | 0.87 |
| **frontalpole** | 3.20 | 3.80 | 1.98 | 1.95 | 2.99 | 2.15 | 5.77 | 3.80 |
| **entorhinal** | 2.84 | 2.59 | 4.33 | 2.84 | 5.26 | 5.34 | 3.29 | 3.26 |
| **cuneus** | 1.64 | 1.92 | 1.40 | 1.94 | 2.30 | 1.99 | 2.31 | 2.37 |
| **caudalmiddlefrontal** | 0.58 | 0.55 | 0.77 | 1.29 | 1.12 | 1.11 | 1.29 | 1.08 |
| **caudalanteriorcingulate** | 0.64 | 1.93 | 4.38 | 2.28 | 1.54 | 1.18 | 2.10 | 2.17 |
| **bankssts** | 0.41 | 0.88 | 0.49 | 0.77 | 0.65 | 0.65 | 0.95 | 0.85 |
| **Total** | 0.52 | 0.55 | 0.67 | 0.74 | 1.22 | 0.67 | 0.86 | 0.58 |

Unit: percentage; ?h_qc: CVs after excluding the two sessions with excessive head motion

### *Time-of-Day*

The linear mixed-effects model did not generate any statistically significant results regarding the time-of-day effect within sub3. However, in sub2, the posterior cingulate cortex (**Fig. 4D**) showed a higher percentage change during morning than afternoon sessions ($R^2$ = 0.14, $t$ = 4.72, $p = 1.91 \times 10^{-4}$). Moreover, the percentage change in morning sessions is lower than that of the afternoon in the temporal pole region in sub1 ($R^2$ = 0.09, $t$ = - 4.18, $p = 2.27 \times 10^{-3}$) whereas the reversed pattern is in sub2 ($R^2$ = 0.14, $t$ = 4.02, $p = 4.55 \times 10^{-3}$). However, see the influence of outliers for further analysis (**Fig. 9**). The percentage changes for each brain region of morning and afternoon sessions were illustrated in **Fig. S2**.

### *Daylight length*

The linear mixed-effects model did not generate any statistically significant results regarding the daylight length effect for any of the participants.

## Cortical Brain Volume

### *Total volumes of different brain apartments*

The average and standard deviation of the total gray matter volumes, total white matter volumes, and the CSF volumes including the left and right hemispheres are described in

**Table 5**.

**Table 5. The values (Mean ± SD) of volumes of the different brain apartments**

| | Sub1 | Sub2 | Sub3 |
|---|---|---|---|
| **Gray Matter** | 477.87 ± 3.88 | 548.77 ± 3.89 | 507.50 ± 6.67 |
| **Left - GM** | 236.60 ± 2.10 | 275.81 ± 2.21 | 254.96 ± 3.57 |
| **Right - GM** | 241.27 ± 2.20 | 272.96 ± 2.03 | 252.54 ± 3.40 |
| **White Matter** | 453.74 ± 2.99 | 515.36 ± 3.32 | 445.01 ± 4.01 |
| **Left - WM** | 224.54 ± 1.55 | 258.50 ± 1.62 | 223.28 ± 2.68 |
| **Right - WM** | 229.20 ± 1.75 | 256.86 ± 1.90 | 221.73 ± 1.92 |
| **CSF** | 0.81 ± 0.04 | 1.22 ± 0.04 | 1.05 ± 0.04 |
| **IntraCranialVol (ICV)** | 1525.148 | 1658.298 | 1555.418 |

The unit of all values is cubic centimeters.

### *Time change of the total volumes of different brain apartments*

There are no significant results regarding the time-of-day or daylight length change in the total volumes of different brain apartments. The distributions of the morning and afternoon sessions are shown in **Fig. S3**.

### *Mean cortical brain volume across sessions*

For three subjects, brain volume of different regions varied from 731.53 to 21682.50 $mm^3$ for sub1, 981.82 to 26338.03 $mm^3$ for sub2, and 903.65 to 26170.17 $mm^3$ for sub3 (**Table S4**). The intracranial volume (ICV) remained consistent for each subject in all sessions (sub1: 1,525,148 $mm^3$; sub2: 1,658,298 $mm^3$; sub3: 1,555,418 $mm^3$) after running the longitudinal pipeline. Ranking the volumes revealed that the superior frontal and rostral middle frontal cortices were the largest in all subjects (**Fig. 6B**). Conversely, the frontal pole and transverse temporal regions were consistently the smallest across all subjects (**Fig. 6B** and **Table S6**).

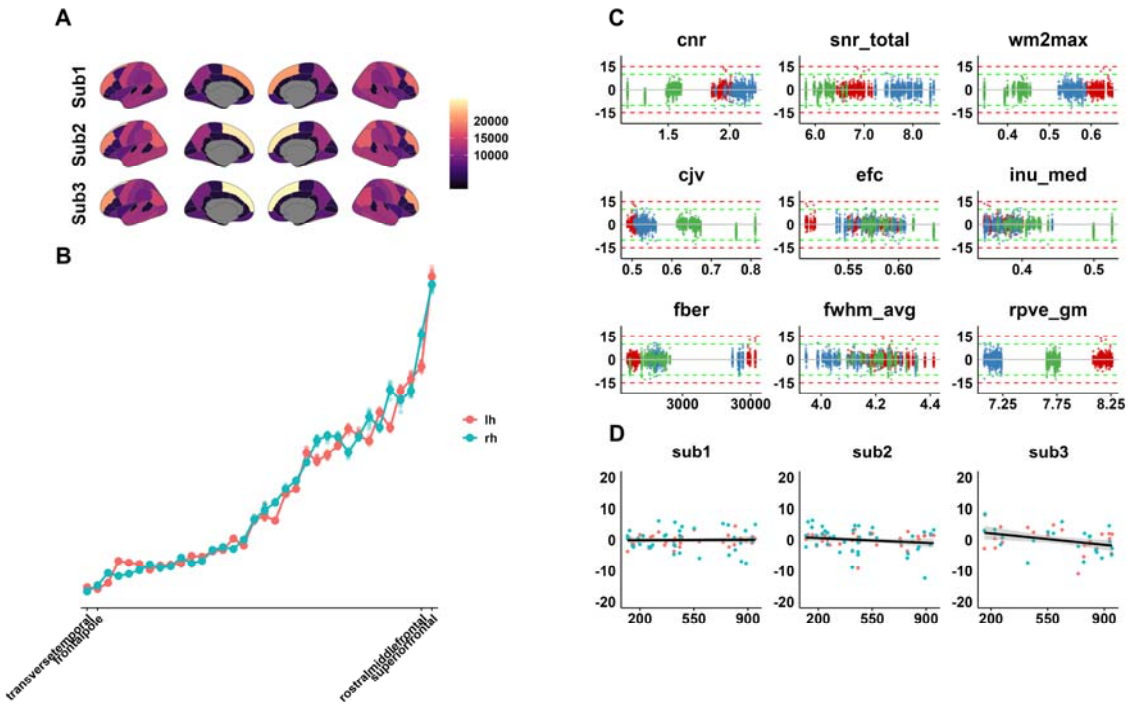



**Fig. 6. Mean cortical brain volume, and percentage change distributions along with IQMs, and the data collections.** **(A)** Average brain volume size of different brain regions for each subject where the unit is cubic millimeters. **(B)** Brain regions from the smallest to the largest (data from sub1). All three subjects have the same the most two smallest and largest brain volumes although the brain regions ranking in the middle vary. **(C)** The distribution of the percentage changes along the IQMs, where the red, blue, and green dots represent subjects 1, 2, and 3, respectively. **(D)** Percentage changes of pericalcarine brain volumes plotted against daylight lengths for each subject.

### *The stability and percentage changes for each brain region*

The distribution of the percentage changes against the IQMs is illustrated in **Fig. 6C**, where it manifested skewed negative changes with larger CJV and smaller CNR induced by the head motion in sub3. These two sessions could contribute to the large percentage change in the cortical brain volume measurement. Accordingly, if we exclude the outliers (sessions 1 and 7), the CV (**Table 6**) and percentage change (sub3_qc in **Fig. 7**) have been significantly improved (left: $t_{33}$ = -5.06, $p = 7.6 \times 10^{-6}$, *Cohen's d* = 0.87; right: $t_{33}$ = -6.21, $p = 2.6 \times 10^{-7}$, *Cohen's d* = 1.06).

The CVs of the within-subject brain volume range from 0.64 to 4.73 for sub1, 0.72 to 4.87 for sub2, and 0.86 to 4.99 for sub3 (**Table 6**). Furthermore, several brain regions showed larger CVs (**Table 6**) and percentage changes (**Fig. 7**), such as the temporal pole, pericalcarine, and entorhinal cortex. Moreover, aside from the abovementioned regions, the percentage changes are well confined within 5% in all subjects (**Fig. 7**).

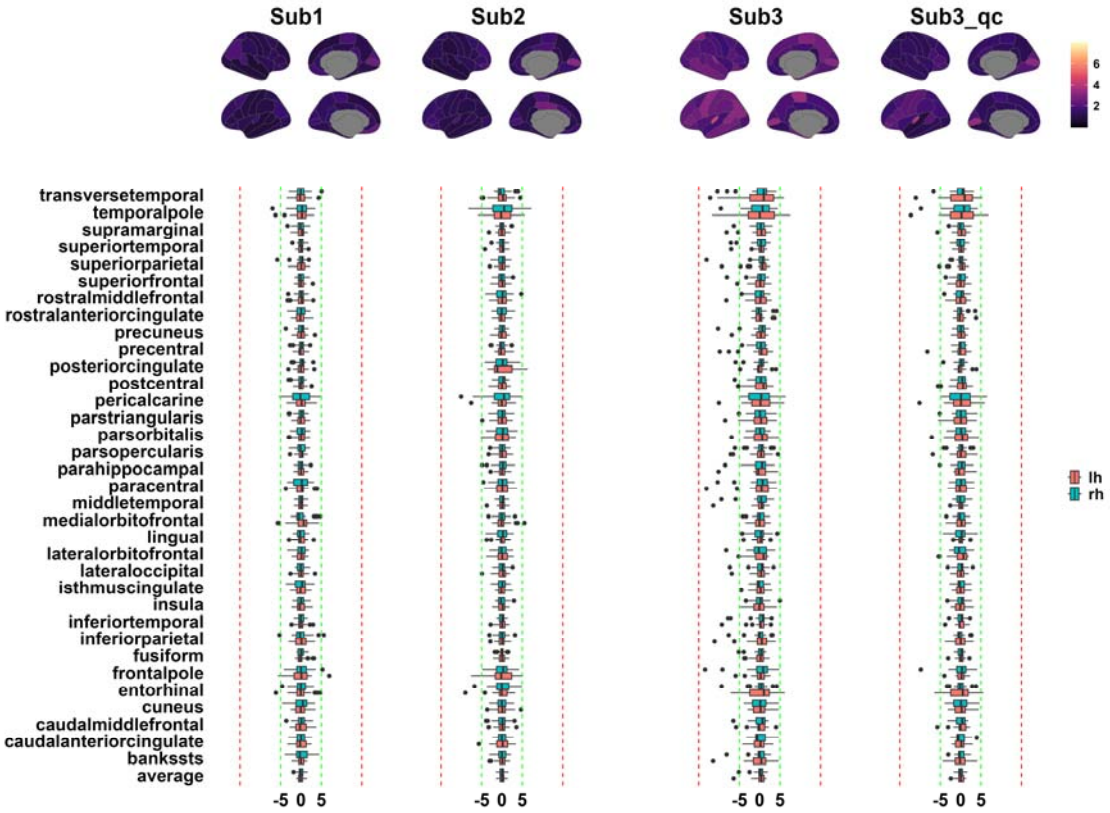



**Fig. 7. CVs and percentage changes in cortical brain volume.** The upper panel depicts the coefficient of the variant (CV) for each subject. The lower boxplot shows the percentage change for each brain region and subject. Sub3 depicts the results generated from the original data while sub3_qc shows results after quality control. The green lines are in the range of ±5 % while the red lines depict the range of ±15%. All values in the boxplot are reported as percentages.

**Table 6. The CVs of the cortical brain volume based on the DK atlas**

| | Sub1 | | Sub2 | | Sub3 | | | |
|---|---|---|---|---|---|---|---|---|
| | Left | Right | Left | Right | Left | Left_qc | Right | Right_qc |
| **transversetemporal** | 1.51 | 1.63 | 1.43 | 1.29 | 3.57 | 2.98 | 2.26 | 1.85 |
| **temporalpole** | 1.96 | 1.99 | 2.63 | 2.96 | 4.96 | 4.99 | 3.31 | 2.96 |
| **supramarginal** | 1.04 | 1.30 | 0.98 | 0.96 | 1.86 | 1.67 | 1.42 | 1.06 |
| **superiortemporal** | 0.95 | 1.01 | 1.15 | 0.83 | 1.16 | 0.98 | 1.65 | 1.06 |
| **superiorparietal** | 1.36 | 1.49 | 1.55 | 1.64 | 1.89 | 1.62 | 1.77 | 0.86 |
| **superiorfrontal** | 1.05 | 1.10 | 1.20 | 1.08 | 1.56 | 1.59 | 1.48 | 1.31 |
| **rostralmiddlefrontal** | 1.27 | 1.31 | 1.34 | 1.46 | 1.67 | 1.58 | 1.71 | 1.70 |
| **rostralanteriorcingulate** | 1.26 | 1.31 | 1.75 | 1.07 | 1.34 | 1.39 | 1.03 | 1.03 |
| **precuneus** | 1.34 | 1.14 | 1.28 | 1.07 | 1.66 | 1.34 | 1.99 | 1.09 |
| **precentral** | 1.45 | 1.32 | 1.66 | 1.37 | 2.27 | 2.09 | 1.02 | 1.00 |
| **posteriorcingulate** | 1.05 | 1.20 | 4.83 | 4.87 | 1.97 | 1.92 | 1.35 | 1.00 |
| **postcentral** | 1.58 | 1.45 | 1.44 | 1.98 | 1.87 | 1.87 | 1.41 | 1.08 |
| **pericalcarine** | 1.83 | 3.54 | 2.20 | 3.94 | 3.99 | 4.05 | 3.95 | 4.12 |
| **parstriangularis** | 1.03 | 1.43 | 1.29 | 1.09 | 1.97 | 1.91 | 1.87 | 1.87 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **parsorbitalis** | 1.47 | 0.98 | 2.13 | 1.81 | 1.87 | 1.95 | 2.08 | 1.78 |
| **parsopercularis** | 1.21 | 1.28 | 1.30 | 1.11 | 2.32 | 1.94 | 2.31 | 1.82 |
| **parahippocampal** | 1.38 | 1.51 | 1.69 | 1.53 | 2.23 | 1.81 | 2.63 | 1.92 |
| **paracentral** | 1.77 | 1.70 | 2.00 | 2.14 | 2.55 | 1.70 | 2.03 | 1.39 |
| **middletemporal** | 1.01 | 1.90 | 1.03 | 0.78 | 1.90 | 1.17 | 2.30 | 1.44 |
| **medialorbitofrontal** | 4.73 | 2.63 | 2.71 | 1.62 | 1.90 | 1.85 | 1.57 | 1.47 |
| **lingual** | 1.08 | 1.19 | 1.27 | 1.54 | 1.22 | 1.02 | 1.61 | 1.63 |
| **lateralorbitofrontal** | 1.05 | 1.24 | 1.31 | 1.07 | 1.73 | 1.63 | 3.04 | 1.74 |
| **lateraloccipital** | 1.38 | 0.93 | 1.18 | 1.23 | 2.02 | 1.44 | 2.06 | 1.53 |
| **isthmuscingulate** | 1.12 | 1.40 | 1.13 | 1.62 | 1.44 | 1.12 | 1.39 | 1.45 |
| **insula** | 1.19 | 1.33 | 1.17 | 1.64 | 2.34 | 2.00 | 1.51 | 1.23 |
| **inferiortemporal** | 1.42 | 1.13 | 1.09 | 0.89 | 2.33 | 1.37 | 2.77 | 1.19 |
| **inferiorparietal** | 1.15 | 1.51 | 1.11 | 1.13 | 2.20 | 1.50 | 1.62 | 1.07 |
| **fusiform** | 1.31 | 0.64 | 0.89 | 0.72 | 1.41 | 0.95 | 1.57 | 0.90 |
| **frontalpole** | 1.87 | 1.73 | 1.67 | 1.32 | 1.56 | 1.44 | 2.45 | 2.20 |
| **entorhinal** | 2.88 | 2.22 | 2.42 | 3.74 | 3.35 | 3.36 | 3.79 | 3.09 |
| **cuneus** | 1.25 | 1.90 | 1.55 | 3.26 | 1.45 | 1.47 | 1.71 | 1.55 |
| **caudalmiddlefrontal** | 1.95 | 1.66 | 1.98 | 2.16 | 2.00 | 2.02 | 1.54 | 1.37 |
| **caudalanteriorcingulate** | 1.49 | 1.23 | 3.76 | 1.85 | 1.80 | 1.69 | 1.54 | 1.40 |
| **bankssts** | 0.77 | 1.84 | 1.03 | 0.98 | 2.40 | 1.55 | 2.26 | 1.28 |

CVs in percentages after excluding the two sessions with excessive head motion

### *Time-of-Day*

The linear mixed-effects model did not generate any statistically significant results regarding the time-of-day effect within each subject in our dataset. The distributions of the morning and afternoon sessions are shown in **Fig. S4.**

### *Daylight length*

The linear mixed-effects model did not generate any statistically significant results regarding the time effect within sub1 and sub2, but sub3 showed brain thickness decline (**Fig. 2D**) in the pericalcarine ($R^2$ = 0.20, $\beta$ = -1.96, $t$ = -5.70, $p = 1.12 \times 10^{-6}$).

## Subcortical volumes

### *Mean subcortical volumes across sessions*

Subcortical structures encompassed the thalamus, caudate nucleus, putamen, pallidum, hippocampus, amygdala, and accumbens area. Subcortical brain volumes ranged from 721.54 to 9063.16 $mm^3$ for sub1, 607.53 to 10865.75 $mm^3$ for sub2, and 497.60 to 7692.36 $mm^3$ for sub3 (**Table 7**). The putamen and and thalamus consistently ranked as the largest subcortical structures, see **Fig. S5** and **Table 7**. Additionally, the pallidum, amygdala, and accumbens area constantly ranked as the smallest subcortical structures across all three subjects (**Fig. S5** and **Table 7**). **Table S5** depicts the average value and standard deviation of the subcortical structures along with the cerebellum, ventricles, brain stem, corpus callosum (CC), and CSF.

**Table 7. Mean and standard deviations of subcortical brain volumes**

| | | Sub1 | Sub2 | Sub3 |
|---|---|---|---|---|
| **Thalamus** | L | 9063.16±94.43 | 10865.75±76.00 | 7692.36±101.40 |

| | | | | |
|---|---|---|---|---|
| | R | 8628.22±120.11 | 10467.27±129.49 | 7426.36±99.28 |
| **Putamen** | L | 6174.55±43.24 | 5999.58±72.23 | 5111.26±66.52 |
| | R | 6458.49±75.02 | 6127.76±74.73 | 5247.80±80.06 |
| **Hippocampus** | L | 4296.58±39.89 | 4758.70±45.44 | 3679.56±56.31 |
| | R | 4398.27±31.41 | 4701.36±54.06 | 4006.50±73.39 |
| **Caudate** | L | 3789.69±42.54 | 4526.46±62.57 | 3750.78±37.03 |
| | R | 4227.56±48.17 | 4568.34±40.35 | 3773.77±33.81 |
| **Pallidum** | L | 2162.79±24.57 | 2129.04±28.34 | 1811.80±24.17 |
| | R | 2463.63±31.62 | 2106.66±27.91 | 1887.07±33.07 |
| **Amygdala** | L | 1618.33±26.01 | 1639.32±39.76 | 1554.44±27.71 |
| | R | 1883.15±27.96 | 1946.16±27.77 | 1834.68±36.75 |
| **Accumbens area** | L | 721.54±44.54 | 607.53±37.92 | 497.60±37.60 |
| | R | 872.33±20.61 | 884.08±23.81 | 725.29±24.23 |
| **Total** | L | 27826.64±126.86 | 30526.37±154.48 | 24097.79±155.04 |
| | R | 28931.65±208.29 | 30801.63±196.29 | 24901.48±177.96 |

The unit of all values is the cubic millimeters.

### *The stability and percentage changes for each subcortical brain region*

The CVs of the within-subject subcortex ranged from 0.64 to 6.17 for Sub1, 0.54 to 6.24 for Sub2, and 0.62 to 7.56 for Sub3 (**Table 8**). As depicted in the upper panel of **Fig. 8**, the subcortex including the thalamus, caudate, putamen, pallidum, hippocampus, amygdala, showed very small CVs (< 2.5%) except the accumbens area (**Table 8**), especially the left side. In addition, **Table S5** depicts the CVs of the subcortices along with the cerebellum, ventricles, brain stem, corpus callosum (CC), and CSF, where it is shown that the corpus callosum showed much higher stability as a whole than its parts such as the anterior, middle, and posterior (**Fig. S6**).

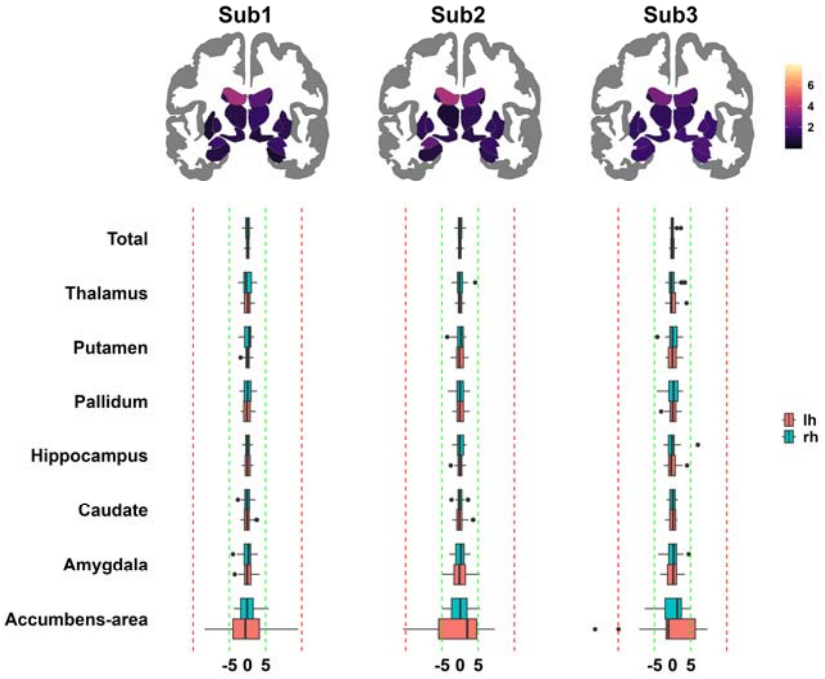


**Fig. 8. CVs and percentage change in brain volume.** The upper panel depicts the coefficient of the variant (CV) for each subject. The lower boxplots shows the percentage change for each brain region and each subject. The green lines are in the range of ±5 % while the red lines depict the range of ±15%. All values in the boxplots are reported as percentages.

**Table 8. The CVs of the non-cortical brain volume based on the DK atlas**

| | Sub1 | | Sub2 | | Sub3 | |
|---|---|---|---|---|---|---|
| | L | R | L | R | L | R |
| **Thalamus** | 1.04 | 1.39 | 0.70 | 1.24 | 1.32 | 1.34 |
| **Caudate** | 1.12 | 1.14 | 1.38 | 0.88 | 0.99 | 0.90 |
| **Putamen** | 0.70 | 1.16 | 1.20 | 1.22 | 1.30 | 1.53 |
| **Pallidum** | 1.14 | 1.28 | 1.33 | 1.32 | 1.33 | 1.75 |
| **Hippocampus** | 0.93 | 0.71 | 0.95 | 1.15 | 1.53 | 1.83 |
| **Amygdala** | 1.61 | 1.48 | 2.43 | 1.43 | 1.78 | 2.00 |
| **Accumbens-area** | 6.17 | 2.36 | 6.24 | 2.69 | 7.56 | 3.34 |
| **Total** | 0.46 | 0.72 | 0.51 | 0.64 | 0.64 | 0.71 |

### *Time-of-day / Length-of-day*

There were no significant effects of time-of-day or daylight length in the percentage change of subcortex volumes. Additionally, the percentage changes were not correlated with any IQMs. The percentage changes for each brain region along the data collection were illustrated in **Fig. S5**.

### The influence of the outliers

As depicted in **Fig. 1A**, sub3 manifested excessive head motion in two sessions (sessions 1 and 7) indicated by the CJV values. Accordingly, after excluding these two sessions, the CVs (**Tables 3, 4, 6**) and the percentage changes (Sub3_qc in **Figs. 3, 5, 7**) have been substantially improved for all the T1w-derived phenotypes, which have been covered by the preceding results.

Moreover, it is shown that sub1 and sub2 manifested larger and smaller percentage changes in the afternoon sessions compared to the morning ones, respectively. If the time-of-day effect exert its power, it should show the similar effect showing the similar pattern not the reversed pattern across most of subjects. Furthermore, the significant results were heavily driven by the last one or two data points, and after excluding these one or two data points, the significant results vanished (**Fig. 9 BD**).

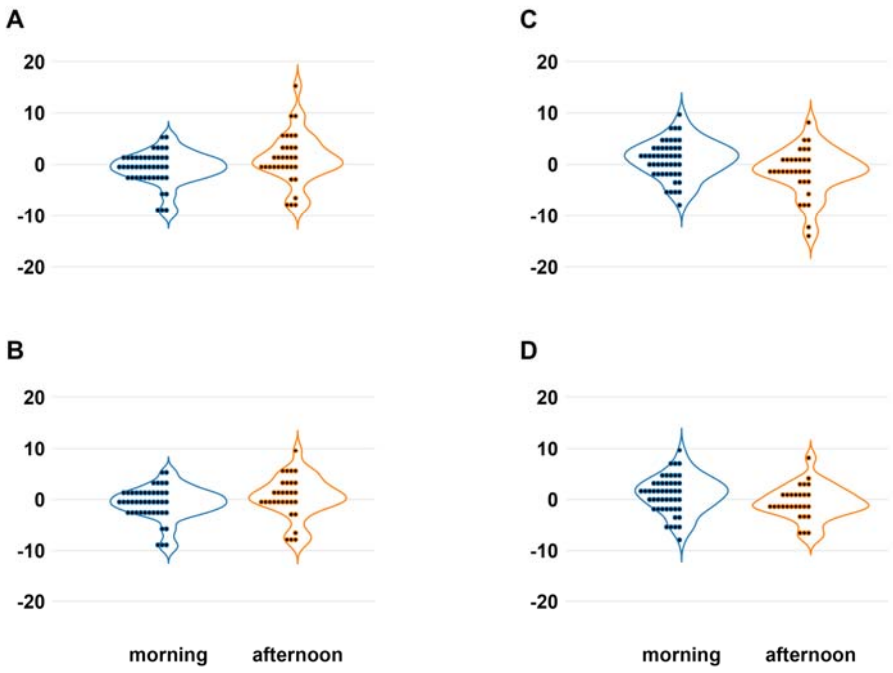



**Fig. 9. The influence of outliers on final results. (A)** Percentage changes of surface areas in the temporal pole in sub1 where percentage changes are larger in the afternoon than in the morning. **(B)** Percentage changes showed no significant differences in sub1 after excluding just one afternoon session. **(C)** Percentage changes of the surface areas in the temporal pole in sub2 where the percentage changes are smaller in the afternoon than the morning. **(D)** Percentage changes showed no significant differences in sub2 after excluding just two afternoon sessions.

### Correlation within and between different phenotypes

In order to explore which brain regions covaried, correlation matrices were computed within and between each phenotype.

Within phenotypes, the average correlation coefficients of thickness (**Fig. 10A**), area (**Fig. 10B**), and volume (**Fig. 10C**) were 0.25, 0.23, and 0.27, respectively. Moreover, between phenotypes, the average correlation coefficients (without the diagonal lines) of thickness-volume (**Fig. 10D**), thickness-area (**Fig. 10E**), and area-volume (**Fig. 10F**) were 0.26, -0.09, and $-3.4 \times 10^{-4}$, respectively. More importantly, the correlation coefficients between brain regions (the diagonal lines in **Fig. 10DEF**) are positive for thickness-volume, negative for thickness-area, and almost none for area-volume (**Table 9**). The distribution of correlation coefficients of each phenotype association is depicted in **Fig. S6**.

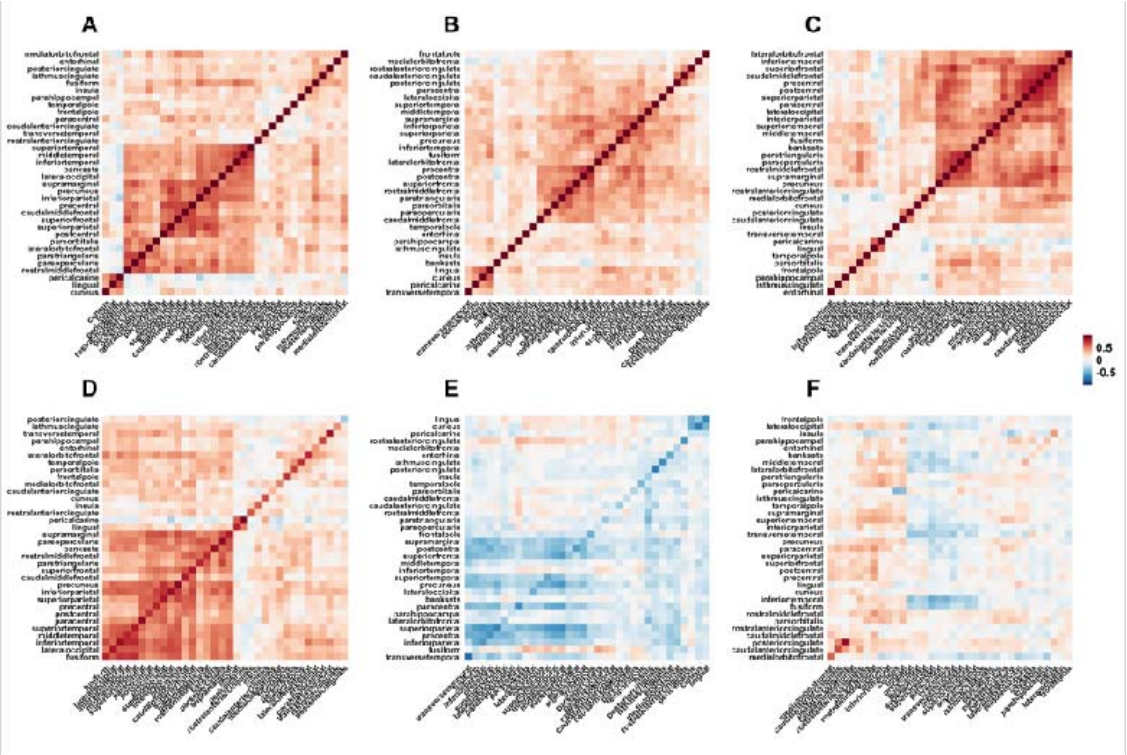



**Fig. 10. Correlation matrices within and between different phenotypes.** Correlation matrices for **(A)** Thickness, **(B)** Surface area, **(C)** Brain volume, **(D)** Thickness and volume, **(E)** Thickness and area, and **(F)** Area and volume.

**Table 9. The correlation coefficients across different phenotypes**

| | Thickness-Volume | Thickness-Area | Area-Volume |
|---|---|---|---|
| **pericalcarine** | 0.96 | -0.22 | -0.02 |
| **precuneus** | 0.91 | -0.48 | -0.18 |
| **parsopercularis** | 0.90 | -0.20 | 0.11 |
| **paracentral** | 0.89 | -0.62 | -0.27 |
| **inferiortemporal** | 0.87 | -0.02 | 0.14 |
| **inferiorparietal** | 0.87 | -0.57 | -0.28 |
| **bankssts** | 0.87 | -0.10 | 0.20 |
| **transversetemporal** | 0.84 | -0.77 | -0.38 |
| **superiorparietal** | 0.83 | -0.66 | -0.23 |
| **lateraloccipital** | 0.82 | -0.49 | -0.06 |
| **superiorfrontal** | 0.82 | -0.41 | 0.11 |
| **superiortemporal** | 0.80 | -0.51 | -0.04 |
| **supramarginal** | 0.80 | -0.43 | 0.04 |
| **caudalmiddlefrontal** | 0.79 | -0.24 | 0.18 |
| **lingual** | 0.78 | -0.65 | -0.13 |
| **fusiform** | 0.78 | -0.23 | 0.26 |
| **parstriangularis** | 0.77 | -0.37 | 0.09 |

| | | | |
|---|---|---|---|
| **precentral** | 0.77 | -0.59 | -0.01 |
| **middletemporal** | 0.76 | -0.22 | 0.35 |
| **lateralorbitofrontal** | 0.76 | -0.36 | 0.01 |
| **rostralmiddlefrontal** | 0.76 | -0.07 | 0.45 |
| **temporalpole** | 0.74 | -0.45 | -0.06 |
| **frontalpole** | 0.73 | -050 | -0.13 |
| **postcentral** | 0.72 | -0.55 | 0.11 |
| **parsorbitalis** | 0.69 | -0.31 | 0.32 |
| **parahippocampal** | 0.67 | -0.22 | 0.28 |
| **cuneus** | 0.65 | -0.70 | -0.04 |
| **rostralanteriorcingulate** | 0.62 | -0.59 | 0.07 |
| **insula** | 0.60 | -0.35 | 0.44 |
| **medialorbitofrontal** | 0.50 | -0.31 | 0.59 |
| **isthmuscingulate** | 0.46 | -0.63 | 0.18 |
| **entorhinal** | 0.43 | -0.51 | 0.22 |
| **caudalanteriorcingulate** | 0.33 | -0.36 | 0.65 |
| **posteriorcingulate** | -0.35 | -0.70 | 0.88 |

## Discussion

In this study, the within-subject stability of T1w-derived phenotypes encompassing cortical thickness, surface area, and brain volume were examined. We showed that the longitudinal pipeline of FreeSurfer generated more stable results compared to the cross-sectional pipeline. Furthermore, the stability of T1w-derived phenotypes was quite high with CVs and percentage change within 5%. Moreover, apparent head motion lead to underestimation of cortical thickness and volume and an overestimation of surface area. Furthermore, percentage changes in the cortical thickness strongly correlated with cortical volume while they correlated negatively with surface area. In addition, outliers in data distribution even with one or two data points could sway the final results, highlighting the importance of reporting the distributions of measurements.

The longitudinal pipeline generates more reliable within-subject results, where it showed much smaller CVs (**Table 1**) and smaller percentage changes (**Fig. 1C**) compared with the cross-sectional pipeline. The longitudinal pipeline was proposed in 2012 by the FreeSurfer team, and since then it has been recommended to analyze longitudinal datasets (Reuter et al., 2012). Specifically, the longitudinal pipeline creates and utilizes an unbiased within-subject template space and images to generate a robust, inverse consistent registration, which will significantly boost the stability and statistical power (Reuter et al., 2012). Indeed, our results are well in line with this argument. Accordingly, we also advocate the longitudinal pipeline should be used when analyzing longitudinal datasets.

### The characteristics of the T1w-derived phenotypes

Cortical thickness is one of the phenotypes that is of great interest in human brain research regarding its role in defining normal cortical maturation (Bethlehem et al., 2022) and its abnormality of various neurological and mental disorders (Bethlehem et al., 2022; Frisoni et

al., 2010; Lemaitre et al., 2012; Thompson et al., 2020). In our dataset, cortical thickness ranged from 1.50 to 4.03 mm which falls well within the known bounds of 1 and 4.5 mm (Fischl & Dale, 2000) indicating the validity of our dataset. Furthermore, it is shown that the cortical thickness increases along the posterior-anterior axis (**Fig. 2A**), which is in concordance with the pattern documented in previous studies (Fischl & Dale, 2000; Hutton, De Vita, Ashburner, Deichmann, & Turner, 2008). Specifically, more detailed information has emerged based on the DK atlas, where we found that the insula, entorhinal cortex, and temporal pole are the thickest while the lingual gyrus, cuneus, and pericalcarine cortex are the thinnest (**Fig. 2B**). What's more, it is shown that the thickest brain regions possess the largest standard deviations which agrees well with previous studies (Fischl & Dale, 2000; Hutton et al., 2008). Previous studies have also shown that the intersubjective standard deviations can be around 0.5mm (Fischl & Dale, 2000; Hutton et al., 2008). Complement that, we found the within-subject standard deviations are much smaller, where most of them lie well below 0.05 mm with the thickest brain regions possessing the largest standard deviation of around 0.10 mm. These results indicate that the cortical thickness measurement is very stable across a short time (about a year) in adults.

The surface area and brain volume are the other two phenotypes that can be constructed from the T1w images which can be used as a proxy index of the brain size (Genon et al., 2022). It is shown that there are big variations between subjects regarding brain sizes where sub2 has the largest brain size followed by sub3 and sub1. This result highlights the importance of counterbalancing the brain size when comparing volume or surface area in a region between subjects. For example, it is common practice that ICV should be divided when comparing brain volumes between subjects instead of using absolute values (Bethlehem et al., 2022; Malone et al., 2015). Even though the brain size varies, based on the DK atlas, the biggest and smallest surface areas and brain volumes are the same within the three subjects. For example, regarding surface area, the superior parietal, rostral middle frontal, and superior frontal cortices are the largest while the entorhinal, transverse temporal, and frontal pole regions are the smallest. Moreover, the superior frontal and rostral middle frontal cortices are the largest for the cortical brain volume whereas the frontal pole and transverse temporal regions are the smallest. Additionally, for subcortical brain volume, it is shown that the thalamus and putamen are the largest and the pallidum, amygdala, and accumbens area are the smallest. The DK atlas was developed based on curvature-based information such as the sulcal representations and the anatomic curvature (Desikan et al., 2006), therefore, the different size of the surface area and brain volume in different brain regions represent their own territories constrained by the curvature boundaries.

**T1w-derived phenotypes are stable across one year**

It is found that the T1w-derived phenotypes including the cortical thickness, surface area, and brain volume in the vast majority of the brain regions are stable, where CVs are well constrained at 2% and the percentage change values are well within 5% over one year. CV and percentage change values have been used to evaluate the stability given that they are unitless and suitable for comparing variables with different sizes (Borga et al., 2020; Carbonell et al., 2022; M. Y. Wang et al., 2024; Y. Wang et al., 2021). CV is the division between standard deviation and average, where the smaller va. In the same vein, percentage change is the division between discrepancy from the average and the average, where smaller values denote high stability and less variation. Generally, a measurement is considered reliable or stable when CVs are within 5 % or percentage changes are within 10% (Borga et al., 2020; Carbonell et al., 2022; M. Y. Wang et al., 2024; Y. Wang et al., 2021). For example, using CVs, it is found that the stability of brain metabolites such as N-acetyl-aspartate was reasonably high with CVs around 4% (M. Y. Wang et al., 2024). Therefore, the 2% of CVs in

this study indicate that the T1w-derived phenotypes encompassing cortical thickness, surface area, and brain volume are fairly stable. In addition, the percentage change of 5% is sensible since it is estimated that during normal aging the phenotypes such as the cortical brain volume decreases at a rate of approximately between 0.5 to 1 % annually in adulthood (Lemaitre et al., 2012; Sele et al., 2021; Storsve et al., 2014).

Although most of the brain regions are quite stable, several cortical brain regions did manifest larger fluctuations across all three T1w-derived phenotypes, such as the temporal pole, frontal pole, pericalcarine, and entorhinal cortex. In addition, subcortical brain region such as the accumbens area showed very high variations. One speculation could be that these brain regions do vary that much, however, the underlying biological reason for it need to be further investigated. Besides, only sub2 manifested a larger variation in the cingulate cortex especially the PCC compared to the other two participants indicating the individual differences.

**Cortical thickness changes together with cortical volume but not with surface area**

It is shown that the overall association within each phenotype is quite similar with an average strength of around 0.25. On the contrary, the relationships between each phenotype are quite divergent. Specifically, the percentage changes in cortical thickness and volume manifested a strong positive correlation while both of them illustrated a weak correlation with the percentage change of surface area. More importantly, the percentage changes in the cortical thickness showed a negative correlation with that of the surface area. These results manifested at the short-timescale agree well with the long-timescale normal aging study (Storsve et al., 2014). Collectively, these results reinforce that cortical thickness and surface area play distinct roles in brain anatomy and aging (Lemaitre et al., 2012; Storsve et al., 2014; Vijayakumar et al., 2016), which is sensible since they have distinct genetic roots (Panizzon et al., 2009).

**Head motion underestimates thickness and volume whereas it overestimates surface area**

Head motion can be indicated by the CJV and EFC values. Higher CJV and EFC values denote apparent head motion, which could deteriorate the SNR and CNR, and blur the image. After excluding the two sessions with excessive head motion in subject 3, the stability of all the T1w-derived phenotypes has been significantly improved evidenced by smaller CVs and percentage changes. However, the pattern is different. On one hand, the cortical thickness and cortical volume estimation were increased indicating the underestimation induced by head motion, which is in concordance with the previous study (Reuter et al., 2015). On the other hand, the surface area measurement is decreased indicating head motion overestimate the surface area. This result indicates that the head motion evaluation should be controlled especially when investigating participants who move a lot such as autistic children or Parkinson's patients (Reuter et al., 2015).

**No time-of-day and daylight effects**

We found no time-of-day effect on cortical thickness, surface area, and brain volumes. However, previous studies reported lower cortical thickness (Trefler et al., 2016) and brain volume (Karch et al., 2019; Nakamura et al., 2015; Trefler et al., 2016) in the afternoon compared to the morning. But there are some limitations or shortcomings of the three studies that could have influenced their results. First, two of the studies were intersubjective design from which the intersubjective variation could interfere with the reported time-of-day effect (Nakamura et al., 2015; Trefler et al., 2016). Second, one study pooled data from other open datasets and separated the data according to the data collection time, in which each subject

generally only had one data point (Nakamura et al., 2015). Third, the other study only had two data points for morning and afternoon sessions for each subject in their primary and secondary datasets (Trefler et al., 2016). Fourth, although one is within-subject design (Karch et al., 2019), they did not use the longitudinal pipeline as recommended by the FreeSurfer team (Reuter et al., 2012) which will provide more robust results as also evidenced by this study. Fifth, none of the studies showed the data distributions in the morning and afternoon sessions. On the contrary, our study was a within-subject design which contained at least 10 sessions and up to 23 sessions for each subject in the morning or afternoon sessions. What's more, we used the longitudinal pipeline in our study and illustrated the data distributions (**Figs. S1-5**). Given that, we speculate that the time-of-day effect in the previous studies (Karch et al., 2019; Nakamura et al., 2015; Trefler et al., 2016) could be interfered with intersubjective variations and variations induced by the cross-sectional pipeline.

Regarding surface area, however, there are no studies on the time-of-day effect on the surface area. But, we did find that in the temporal pole, sub1 manifested a lower percentage change in the morning compared to the afternoon sessions, whereas sub2 showed the opposite pattern. The contradictory findings can be explained by one or two outlier sessions (**Fig. 9**) in the dataset highlighting the necessity of displaying the distribution instead of only reporting summary statistics (Allen, Erhardt, & Calhoun, 2012; Hintze & Nelson, 1998). Moreover, sub2 showed a higher percentage change in the posterior cingulate cortex during morning than afternoon sessions, which neither sub1 nor sub3 illustrated. Therefore, this result is part of the individual characters in sub2 instead of generalizing into the general population. In short, we infer that the time-of-day could not sway the results of the T1w-derived phenotypes.

Similarly, we found no daylight effect on cortical thickness, surface area, and brain volumes. Previous studies have focused on the seasonal effect (Zhang et al., 2023) on functional brain organization and stated that daylight could reconfigure the resting state brain networks (M. Y. Wang et al., 2023b). However, to the best of our knowledge, this study could be the first study directly investigating the daylight effect on T1w-derived phenotypes. Although sub3 manifested decreasing patterns regarding the cortical thickness and brain volume in the pericalcarine as the daylight length increases, this pattern did not emerge in the other two participants which impedes its generalization. Given the smallest data points (23 after QC) and head motion (largest CJV) within sub3 compared to 40 sessions within sub2 and 38 sessions within sub1, it is suspected that the results in sub3 could rise from the large variation, however, further investigation should be warrant.

## Limitations

Several limitations should be articulated. First, lacking female subjects hampers the generalization of our results. It is known that major events such as pregnancy (Hoekzema et al., 2017) and hormone levels can affect brain structure (Rizor et al., 2023). Second, although using the same MRI scanner and the same data processing pipeline can alleviate the interference from heterogeneous ones, it could impede the application to other datasets that are collected on other MRI scanners or data processing pipelines. Third, the timing of data collection was not fixed for each subject, for example, sub1 was not scanned between 10 and 12 on Mondays, instead, each subject was scanned based on the availability of the scanning time and their available time.

## Conclusion

In summary, the stability of T1w-derived phenotypes across one year during adulthood is fairly high with CVs within 2% while percentage changes within 5%. Moreover, quality

checking of the T1w images especially head motion should be reported and distinct become a common practice since it leads to the underestimation of the cortical thickness and volume while overestimation of the surface area. Lastly, data distribution reporting should also be a common practice since even one or two data outliers could sway the final results.

**Funding**

This study was financed by the Research Council of Norway (Project number: 276044: When default is not default: Solutions to the replication crisis and beyond).

**Data and code**

Data and code used in this manuscript can be found here on GitHub https://github.com/MengYunWang/BBSC/tree/main/fMRI/T1_analysis.

**Acknowledgments**

We appreciate the technical support and data collection support from our radiologists (Christel Jansen, Eva Øksnes, Roger Barndon, Trond Øvreaas, Tor Fjørtoft, and Turid Randa) at the Haukeland University Hospital.

**Conflicts of interest**

No declared conflicts of interest.

**References**

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. doi:10.1016/j.neuroimage.2017.10.034

Allen, E. A., Erhardt, E. B., & Calhoun, V. D. (2012). Data visualization in the neurosciences: overcoming the curse of dimensionality. *Neuron, 74*(4), 603-608. doi:10.1016/j.neuron.2012.05.001

Atkinson, D., Hill, D. L., Stoyle, P. N., Summers, P. E., & Keevil, S. F. (1997). Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Trans Med Imaging, 16*(6), 903-910. doi:10.1109/42.650886

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:DOI 10.18637/jss.v067.i01

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., . . . Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature, 604*(7906), 525-533. doi:10.1038/s41586-022-04554-y

Borga, M., Ahlgren, A., Romu, T., Widholm, P., Leinhard, O. D., & West, J. (2020). Reproducibility and repeatability of MRI-based body composition analysis. *Magnetic Resonance in Medicine, 84*(6), 3146-3156. doi:10.1002/mrm.28360

Carbonell, G., Kennedy, P., Bane, O., Kirmani, A., El Homsi, M., Stocker, D., . . . Taouli, B. (2022). Precision of MRI radiomics features in the liver and hepatocellular carcinoma. *European Radiology, 32*(3), 2030-2040. doi:10.1007/s00330-021-08282-1

Chen, C. H., Gutierrez, E. D., Thompson, W., Panizzon, M. S., Jernigan, T. L., Eyler, L. T., . . . Dale, A. M. (2012). Hierarchical genetic organization of human cortical surface area. *Science, 335*(6076), 1634-1636. doi:10.1126/science.1215330

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage, 9*(2), 179-194. doi:10.1006/nimg.1998.0395

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., . . . Killiany, R. J. (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. doi:10.1016/j.neuroimage.2006.01.021

Di, X., Woelfer, M., Kühn, S., Zhang, Z. G., & Biswal, B. B. (2022). Estimations of the weather effects on brain functions using functional MRI: A cautionary note. *Human Brain Mapping, 43*(11), 3346-3356. doi:10.1002/hbm.25576

Ducharme, S., Albaugh, M. D., Nguyen, T. V., Hudziak, J. J., Mateos-Perez, J. M., Labbe, A., . . . Brain Development Cooperative, G. (2016). Trajectories of cortical thickness maturation in normal brain development--The importance of quality control procedures. *Neuroimage, 125*, 267-279. doi:10.1016/j.neuroimage.2015.10.010

Elliott, L. T., Sharp, K., Alfaro-Almagro, F., Shi, S., Miller, K. L., Douaud, G., . . . Smith, S. M. (2018). Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature, 562*(7726), 210-216. doi:10.1038/s41586-018-0571-7

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *Plos One, 12*(9), e0184661. doi:10.1371/journal.pone.0184661

Finn, E. S., Poldrack, R. A., & Shine, J. M. (2023). Functional neuroimaging as a catalyst for integrated neuroscience. *Nature, 623*(7986), 263-273. doi:10.1038/s41586-023-06670-9

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A, 97*(20), 11050-11055. doi:10.1073/pnas.200033797

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage, 9*(2), 195-207. doi:10.1006/nimg.1998.0396

Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med, 33*(5), 636-647. doi:10.1002/mrm.1910330508

Frisoni, G. B., Fox, N. C., Jack, C. R., Jr., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology, 6*(2), 67-77. doi:10.1038/nrneurol.2009.215

Ganzetti, M., Wenderoth, N., & Mantini, D. (2016). Intensity Inhomogeneity Correction of Structural MR Images: A Data-Driven Approach to Define Input Algorithm Parameters. *Frontiers in Neuroinformatics, 10*, 10. doi:10.3389/fninf.2016.00010

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

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience, 19*(3), 123-137. doi:10.1038/nrn.2018.1

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., . . . Consortium, W. U.-M. H. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage, 80*, 105-124. doi:10.1016/j.neuroimage.2013.04.127

Grasby, K. L., Jahanshad, N., Painter, J. N., Colodro-Conde, L., Bralten, J., Hibar, D. P., . . . Enhancing NeuroImaging Genetics through Meta-Analysis Consortium -Genetics working, g. (2020). The genetic architecture of the human cerebral cortex. *Science, 367*(6484), eaay6690. doi:10.1126/science.aay6690

Gronenschild, E. H., Habets, P., Jacobs, H. I., 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*(6), e38234. doi:10.1371/journal.pone.0038234

Hagler, D. J., Jr., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., . . . Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage, 202*, 116091. doi:10.1016/j.neuroimage.2019.116091

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., . . . Fischl, B. (2006). Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage, 32*(1), 180-194. doi:10.1016/j.neuroimage.2006.02.051

Hintze, J. L., & Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *American Statistician, 52*(2), 181-184. doi:Doi 10.2307/2685478

Hoekzema, E., Barba-Müller, E., Pozzobon, C., Picado, M., Lucco, F., García-García, D., . . . Vilarroya, O. (2017). Pregnancy leads to long-lasting changes in human brain structure. *Nature Neuroscience, 20*(2), 287-296. doi:10.1038/nn.4458

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. Retrieved from <Go to ISI>://WOS:A1979JY78700003

Holz, N. E., Zabihi, M., Kia, S. M., Monninger, M., Aggensteiner, P. M., Siehl, S., . . . Marquand, A. F. (2023). A stable and replicable neural signature of lifespan adversity in the adult brain. *Nature Neuroscience, 26*(9), 1603-1612. doi:10.1038/s41593-023-01410-8

Hutton, C., De Vita, E., Ashburner, J., Deichmann, R., & Turner, R. (2008). Voxel-based cortical thickness measurements in MRI. *Neuroimage, 40*(4), 1701-1710. doi:10.1016/j.neuroimage.2008.01.027

Iscan, Z., Jin, T. B., Kendrick, A., Szeglin, B., Lu, H., Trivedi, M., . . . DeLorenzo, C. (2015). Test-retest reliability of freesurfer measurements within and between sites: Effects of visual approval process. *Human Brain Mapping, 36*(9), 3472-3485. doi:10.1002/hbm.22856

Karch, J. D., Filevich, E., Wenger, E., Lisofsky, N., Becker, M., Butler, O., . . . Kuhn, S. (2019). Identifying predictors of within-person variance in MRI-based brain volume estimates. *Neuroimage, 200*, 575-589. doi:10.1016/j.neuroimage.2019.05.030

Korbmacher, M., Wang, M. Y., Eikeland, R., Buchert, R., Andreassen, O. A., Espeseth, T., . . . Specht, K. (2023). Considerations on brain age predictions from repeatedly sampled data across time. *Brain and Behavior*. doi:10.1002/brb3.3219

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1-26. doi:DOI 10.18637/jss.v082.i13

Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., & Mattay, V. S. (2012). Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiology of Aging, 33*(3), 617 e611-619. doi:10.1016/j.neurobiolaging.2010.07.013

Lerch, J. P., van der Kouwe, A. J., Raznahan, A., Paus, T., Johansen-Berg, H., Miller, K. L., . . . Sotiropoulos, S. N. (2017). Studying neuroanatomy using MRI. *Nature Neuroscience, 20*(3), 314-326. doi:10.1038/nn.4501

Magnotta, V. A., Friedman, L., & First, B. (2006). Measurement of Signal-to-Noise and Contrast-to-Noise in the fBIRN Multicenter Imaging Study. *Journal of Digital Imaging, 19*(2), 140-147. doi:10.1007/s10278-006-0264-x

Malone, I. B., Leung, K. K., Clegg, S., Barnes, J., Whitwell, J. L., Ashburner, J., . . . Ridgway, G. R. (2015). Accurate automatic estimation of total intracranial volume: A nuisance variable with less nuisance. *Neuroimage, 104*, 366-372. doi:10.1016/j.neuroimage.2014.09.034

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. doi:10.1177/2515245920928009

Nakamura, K., Brown, R. A., Narayanan, S., Collins, D. L., Arnold, D. L., & Alzheimer's Disease Neuroimaging, I. (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

Orban, C., Kong, R., Li, J. W., Chee, M. W. L., & Yeo, B. T. T. (2020). Time of day is associated with paradoxical reductions in global signal fluctuation and functional connectivity. *Plos Biology, 18*(2). doi:ARTN e3000602

10.1371/journal.pbio.3000602

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. doi:10.1093/cercor/bhp026

Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature, 526*(7573), 371-379. doi:10.1038/nature15692

R Core Team. (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: a robust approach. *Neuroimage, 53*(4), 1181-1196. doi:10.1016/j.neuroimage.2010.07.020

Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage, 61*(4), 1402-1418. doi:10.1016/j.neuroimage.2012.02.084

Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., & 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

Rizor, E. J., Babenko, V., Dundon, N. M., Beverly-Aylwin, R., Stump, A., Hayes, M., . . . Grafton, S. T. (2023). Menstrual cycle-driven hormone concentrations co-fluctuate with white and grey matter architecture changes across the whole brain. *bioRxiv*, 2023.2010.2009.561616. doi:10.1101/2023.10.09.561616

Sele, S., Liem, F., Merillat, S., & Jancke, L. (2021). Age-related decline in the brain: a longitudinal study on inter-individual variability of cortical thickness, area, volume, and cognition. *Neuroimage, 240*, 118370. doi:10.1016/j.neuroimage.2021.118370

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., & Toga, A. W. (2003). Mapping cortical change across the human life span. *Nature Neuroscience, 6*(3), 309-315. doi:10.1038/nn1008

Sowell, E. R., Thompson, P. M., Leonard, C. M., Welcome, S. E., Kan, E., & Toga, A. W. (2004). Longitudinal mapping of cortical thickness and brain growth in normal children. *Journal of Neuroscience, 24*(38), 8223-8231. doi:10.1523/JNEUROSCI.1798-04.2004

Storsve, A. B., Fjell, A. M., Tamnes, C. K., Westlye, L. T., Overbye, K., Aasland, H. W., & Walhovd, K. B. (2014). Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: regions of accelerating and decelerating change. *Journal of Neuroscience, 34*(25), 8488-8498. doi:10.1523/JNEUROSCI.0391-14.2014

Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., . . . Consortium, E. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl Psychiatry, 10*(1), 100. doi:10.1038/s41398-020-0705-1

Trefler, A., Sadeghi, N., Thomas, A. G., Pierpaoli, C., Baker, C. I., & 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

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging, 29*(6), 1310-1320. doi:10.1109/TMI.2010.2046908

Vijayakumar, N., Allen, N. B., Youssef, G., Dennison, M., Yucel, M., Simmons, J. G., & Whittle, S. (2016). Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Human Brain Mapping, 37*(6), 2027-2038. doi:10.1002/hbm.23154

Wang, M. Y., Korbmacher, M., Eikeland, R., Craven, A. R., & Specht, K. (2024). The intra-individual reliability of 1H-MRS measurement in the anterior cingulate cortex across 1 year. *Human Brain Mapping, 45*(1), e26531. doi:10.1002/hbm.26531

Wang, M. Y., Korbmacher, M., Eikeland, R., & Specht, K. (2022). Deep brain imaging of three participants across 1 year: The Bergen breakfast scanning club project. *Frontiers in Human Neuroscience, 16*, 1021503. doi:10.3389/fnhum.2022.1021503

Wang, M. Y., Korbmacher, M., Eikeland, R., & Specht, K. (2023a). The Bergen Breakfast Scanning Club dataset: a deep brain imaging dataset. *bioRxiv*. doi:10.1101/2023.05.30.542072

Wang, M. Y., Korbmacher, M., Eikeland, R., & Specht, K. (2023b). *The long night effect on the brain functional organization*. Paper presented at the The Organization for Human Brain Mapping 2023 Annual Conference, Montreal, Canada. https://hvlopen.brage.unit.no/hvlopen-xmlui/bitstream/handle/11250/3088274/OHBM_Poster_861.pdf?sequence=1

Wang, Y., Tadimalla, S., Rai, R., Goodwin, J., Foster, S., Liney, G., . . . Haworth, A. (2021). Quantitative MRI: Defining repeatability, reproducibility and accuracy for prostate cancer imaging biomarker development. *Magnetic Resonance Imaging, 77*, 169-179. doi:10.1016/j.mri.2020.12.018

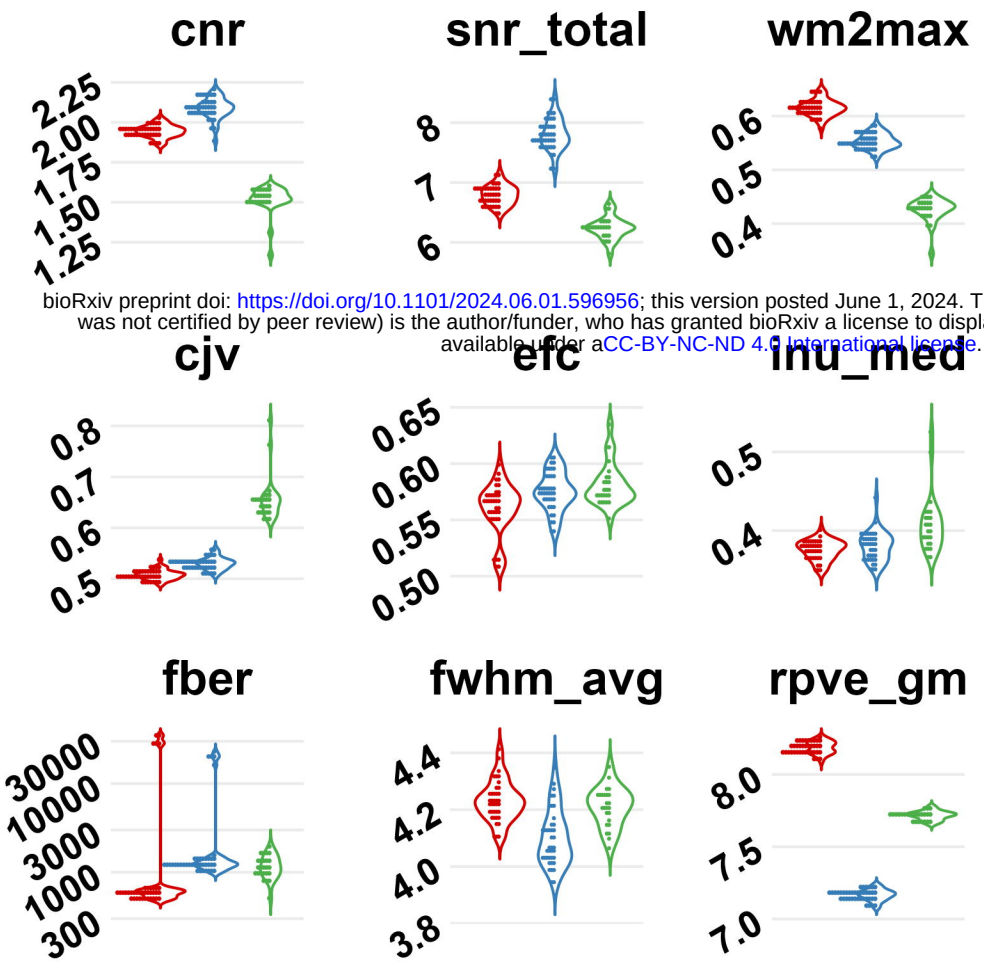
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*: Springer-Verlag New York.

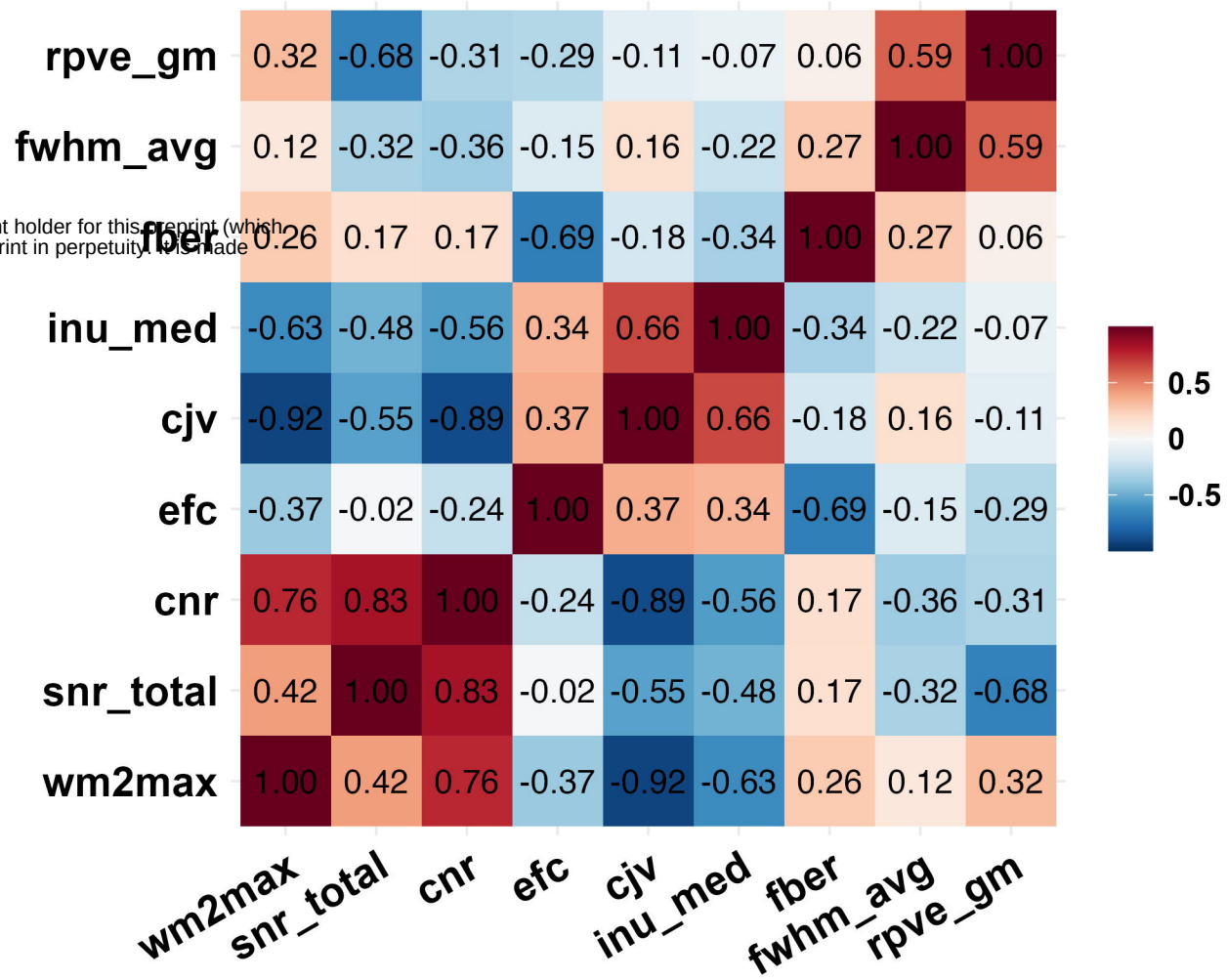
Wu, J., Li, J., Eickhoff, S. B., Scheinost, D., & Genon, S. (2023). The challenges and prospects of brain-based prediction of behaviour. *Nature Human Behaviour, 7*(8), 1255-1264. doi:10.1038/s41562-023-01670-1

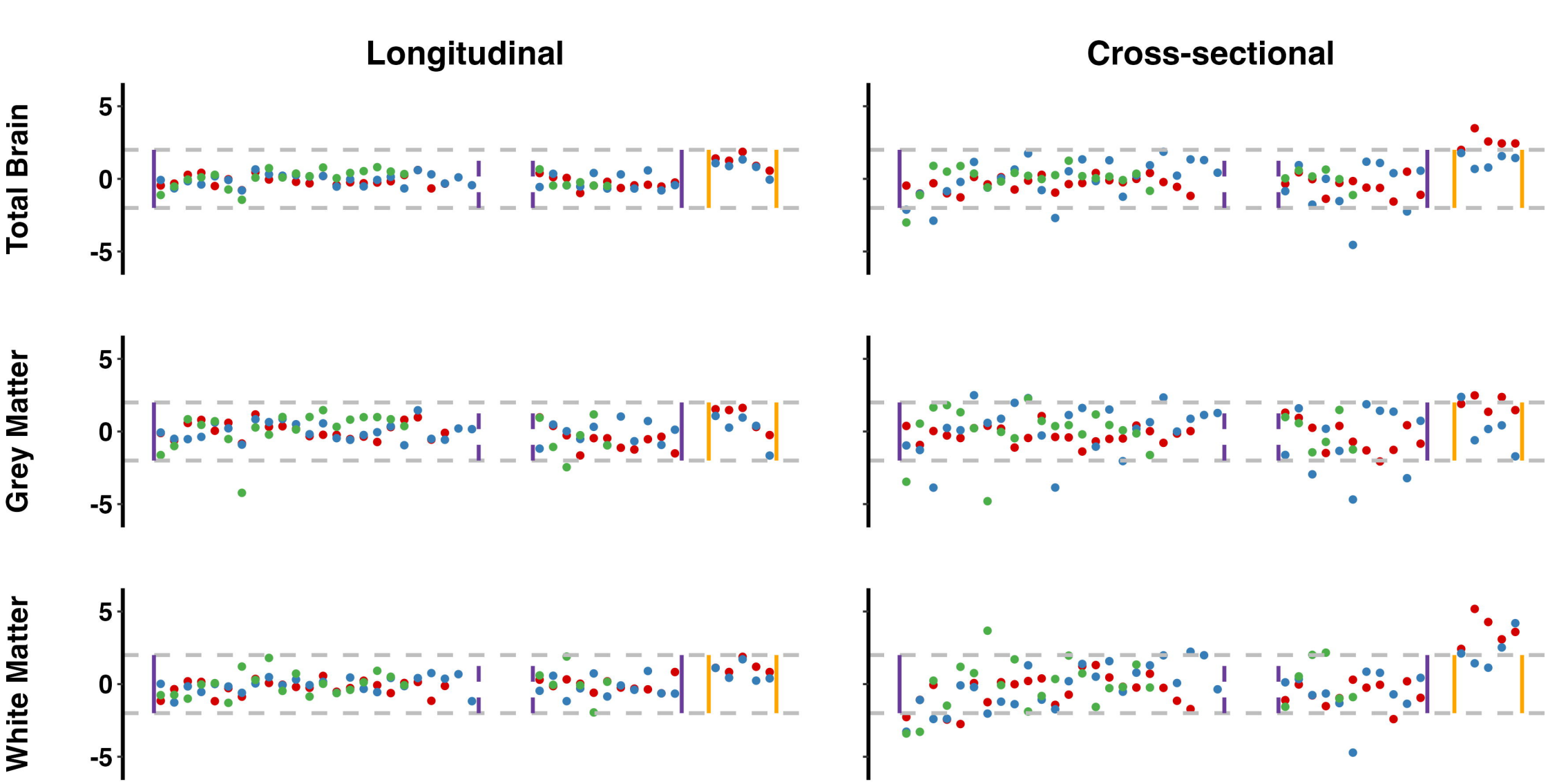
Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., . . . Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by

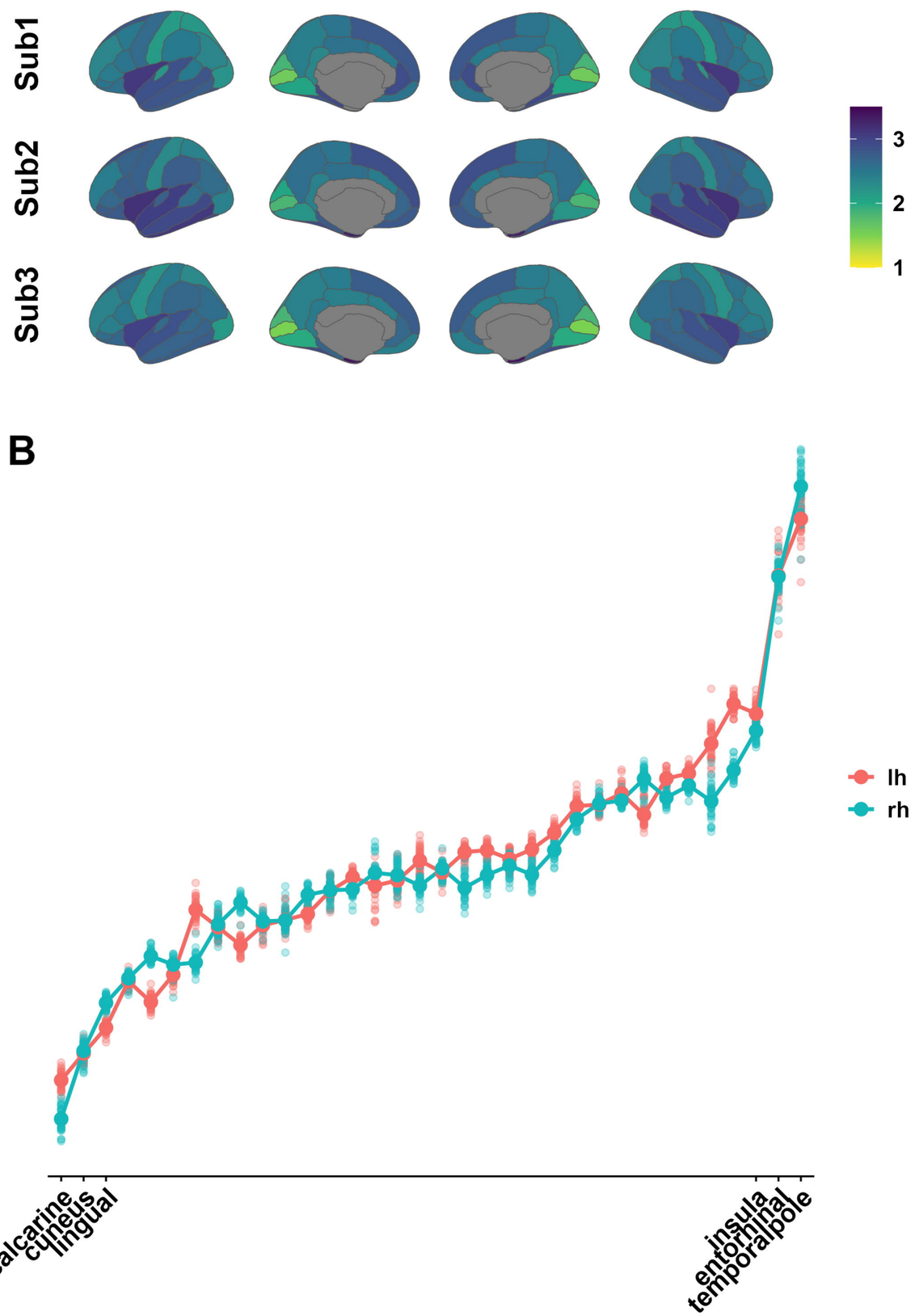
intrinsic functional connectivity. *Journal of Neurophysiology, 106*(3), 1125-1165. doi:10.1152/jn.00338.2011

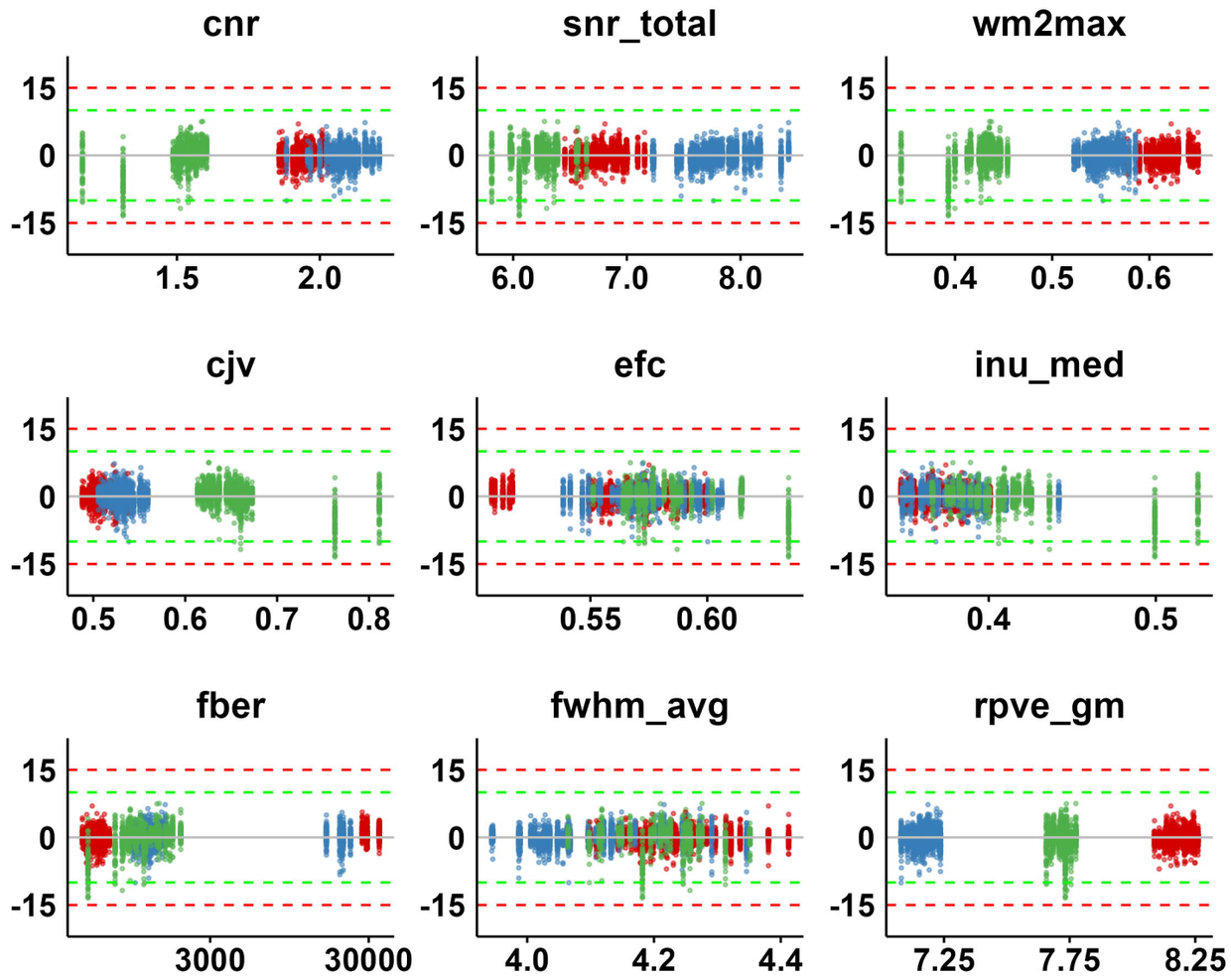
Zhang, R., Shokri-Kojori, E., & Volkow, N. D. (2023). Seasonal effect-an overlooked factor in neuroimaging research. *Translational Psychiatry, 13*(1). doi:ARTN 238 10.1038/s41398-023-02530-2

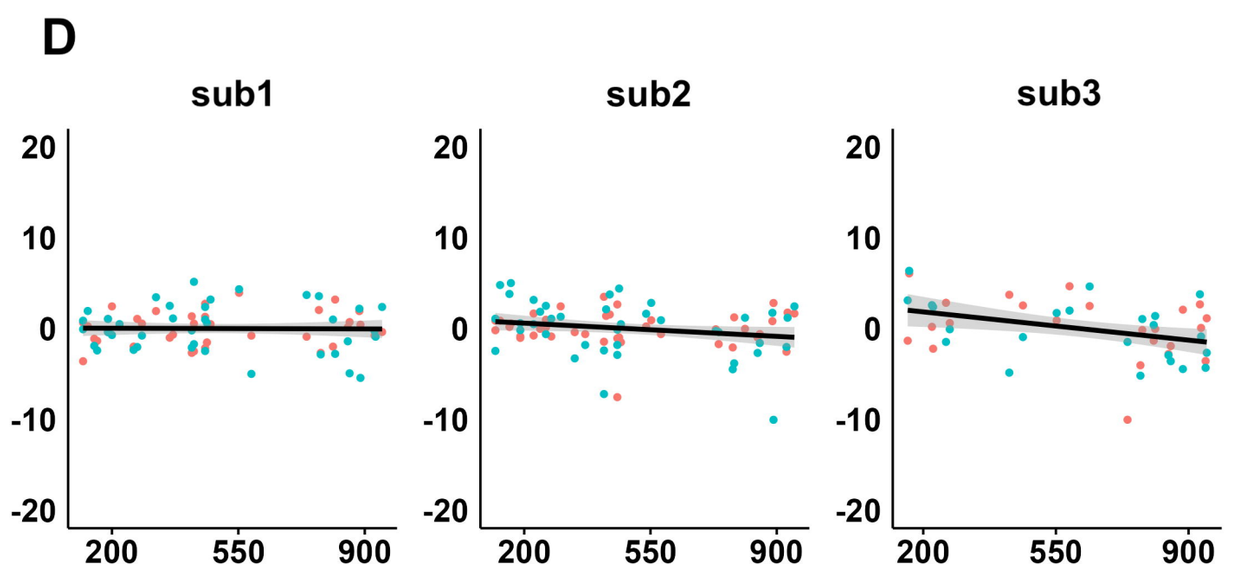


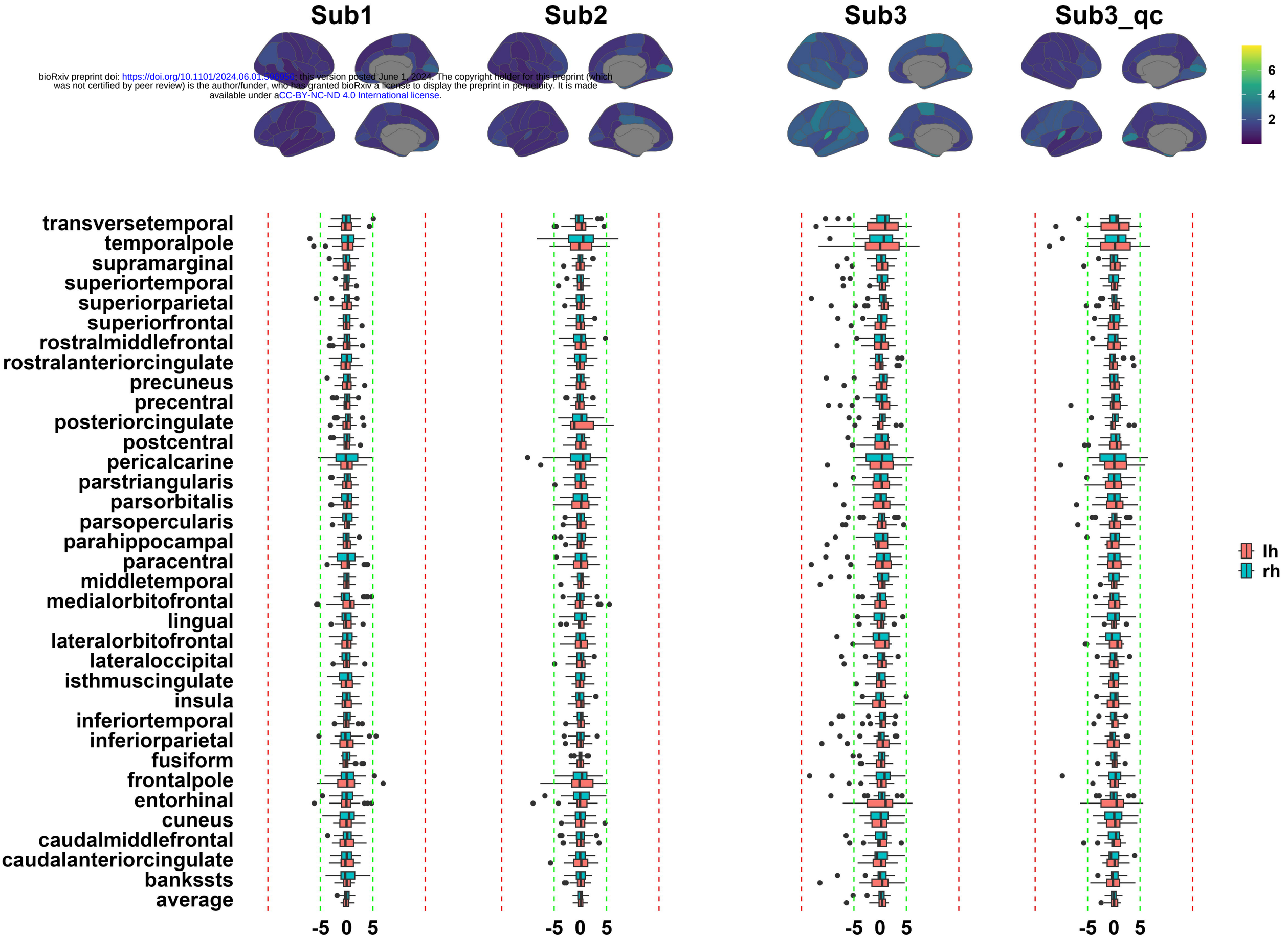



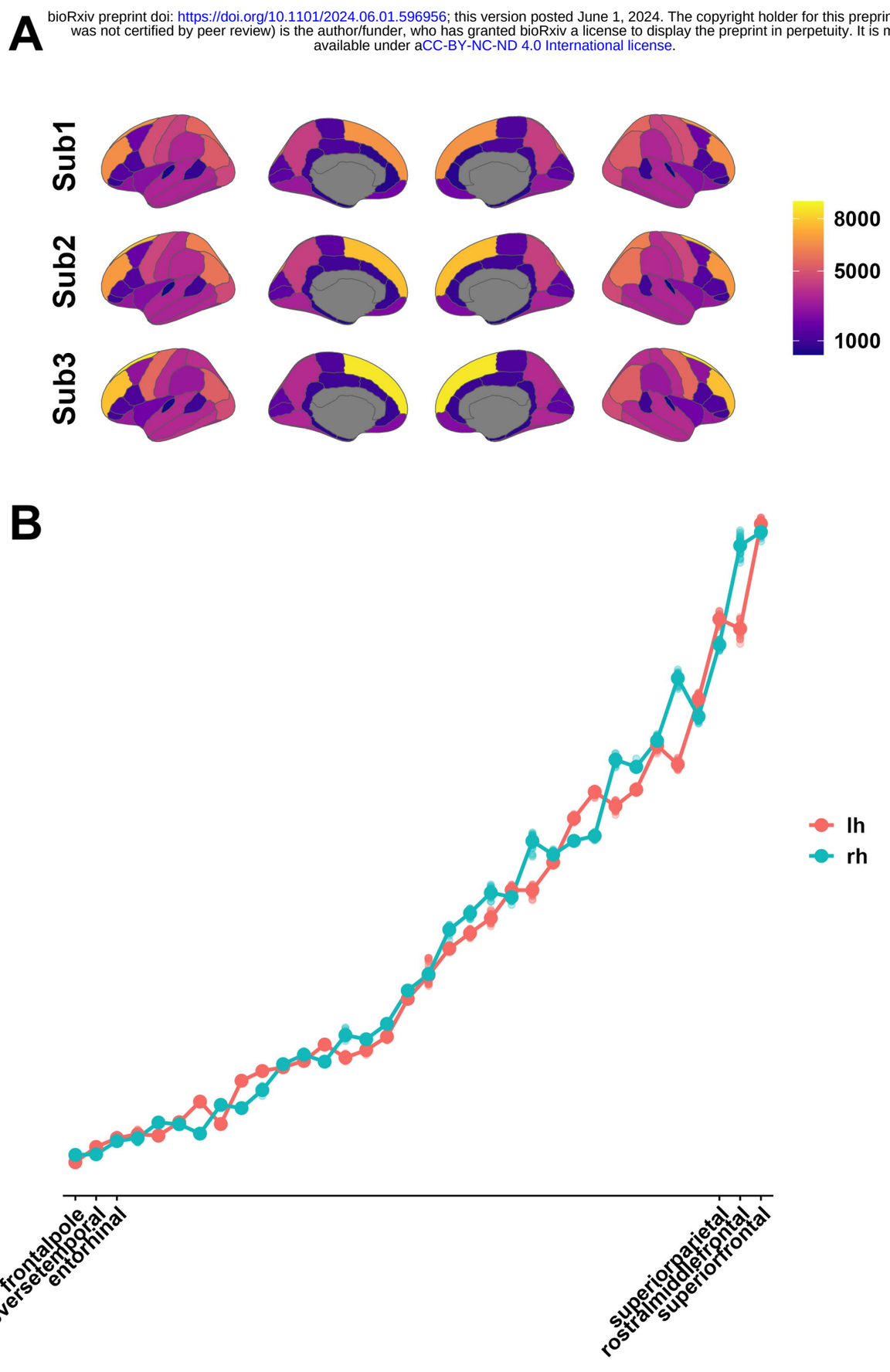



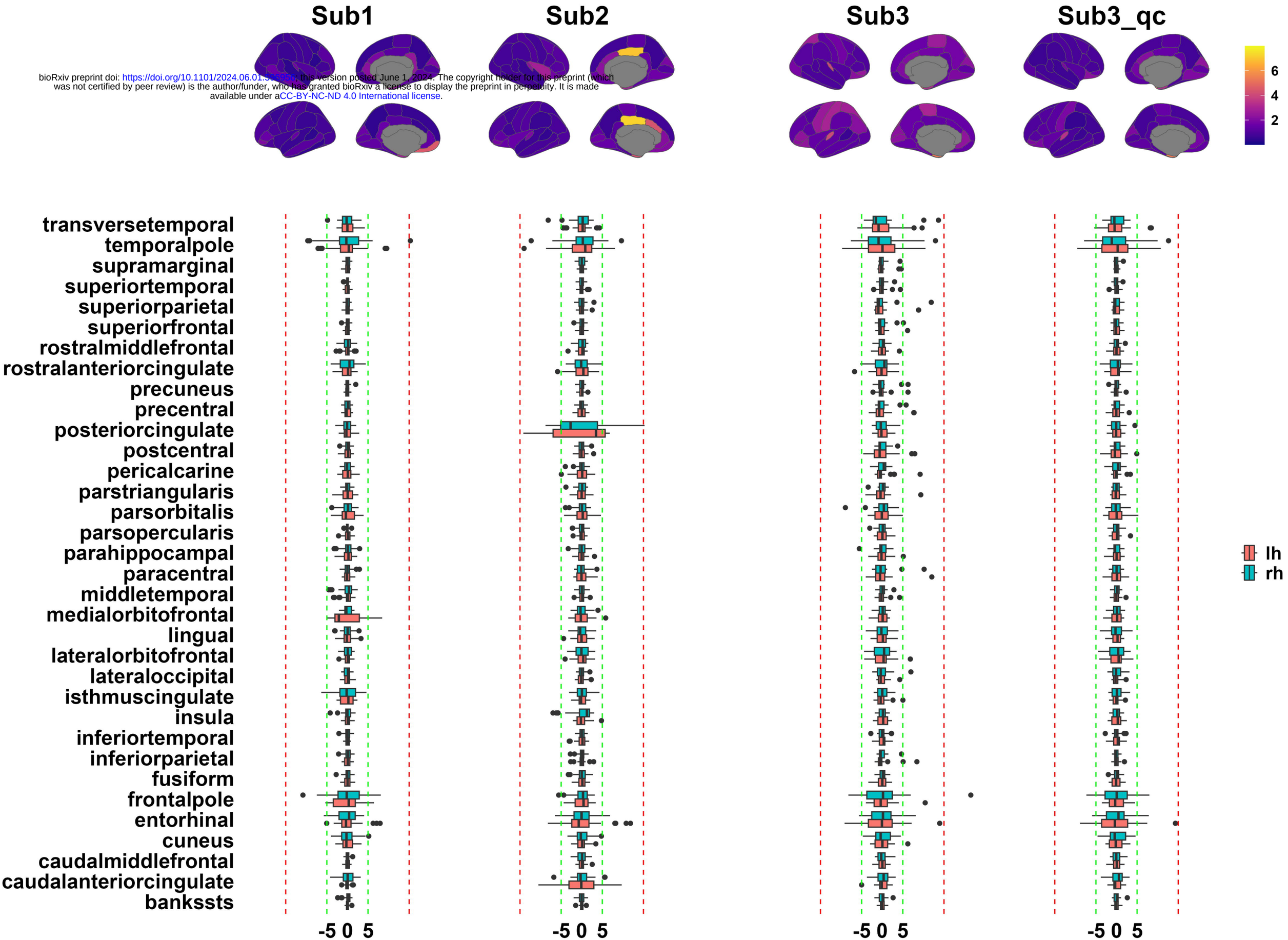


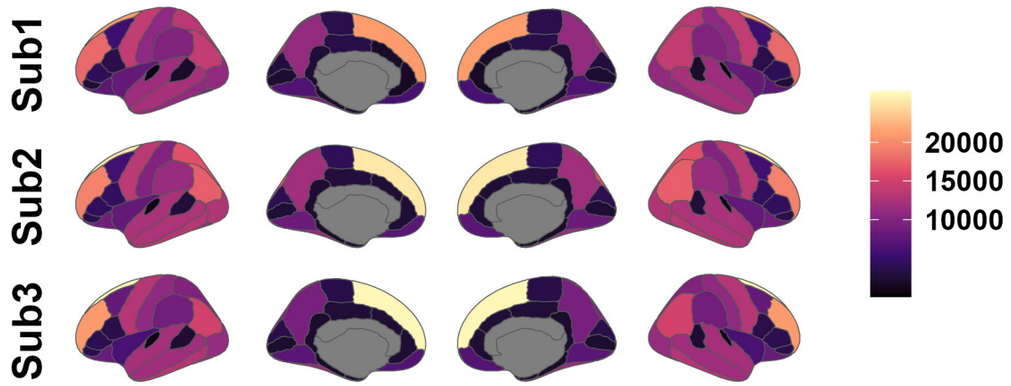





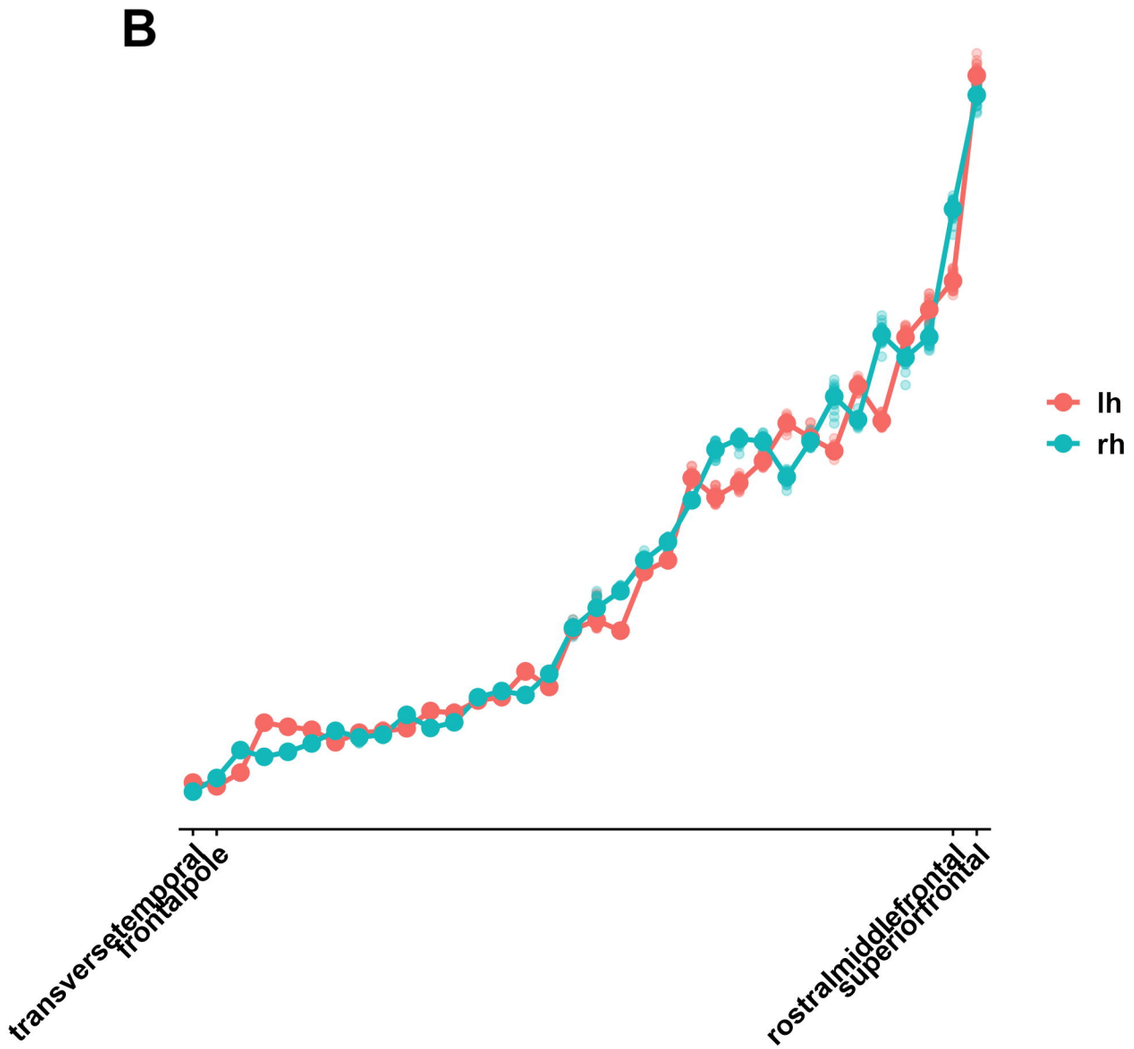




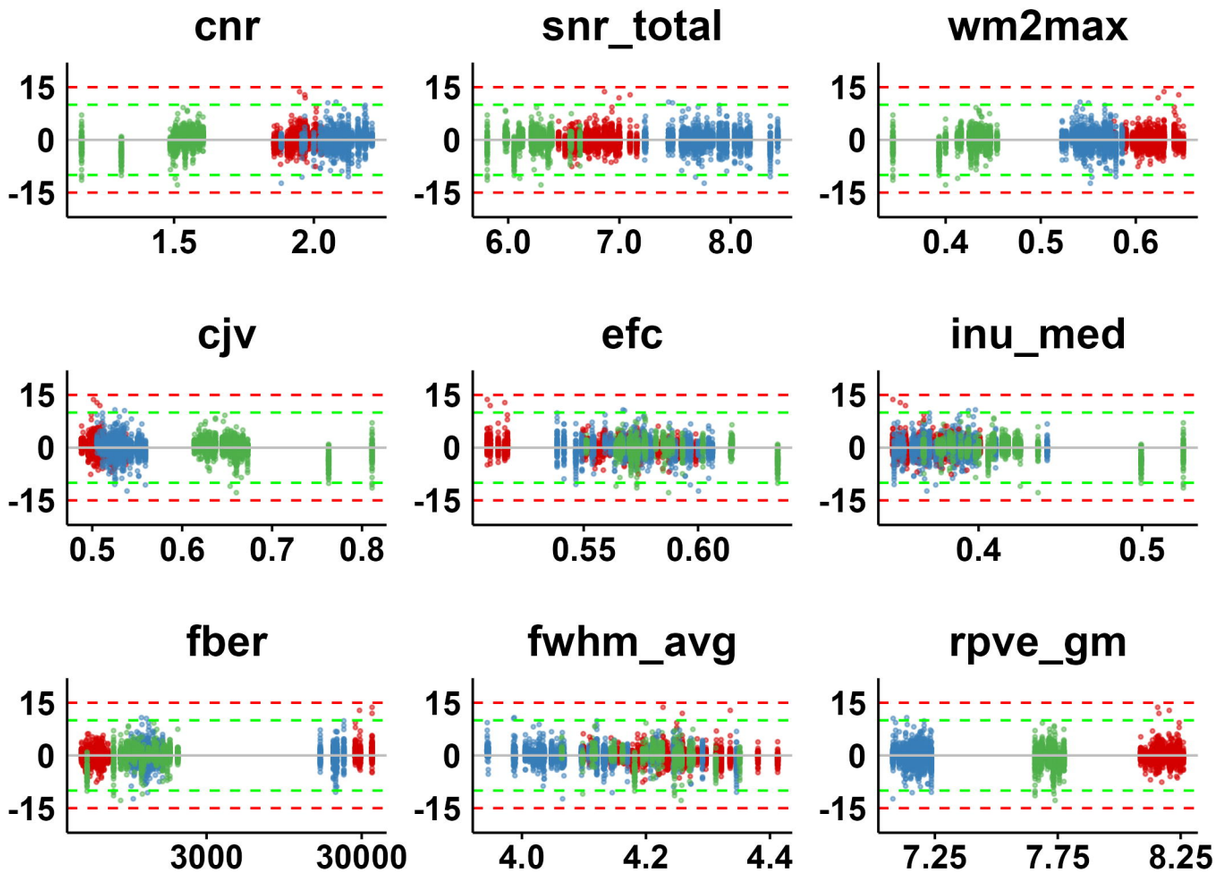
A


Sub1
Sub2
Sub3
20000
15000
10000

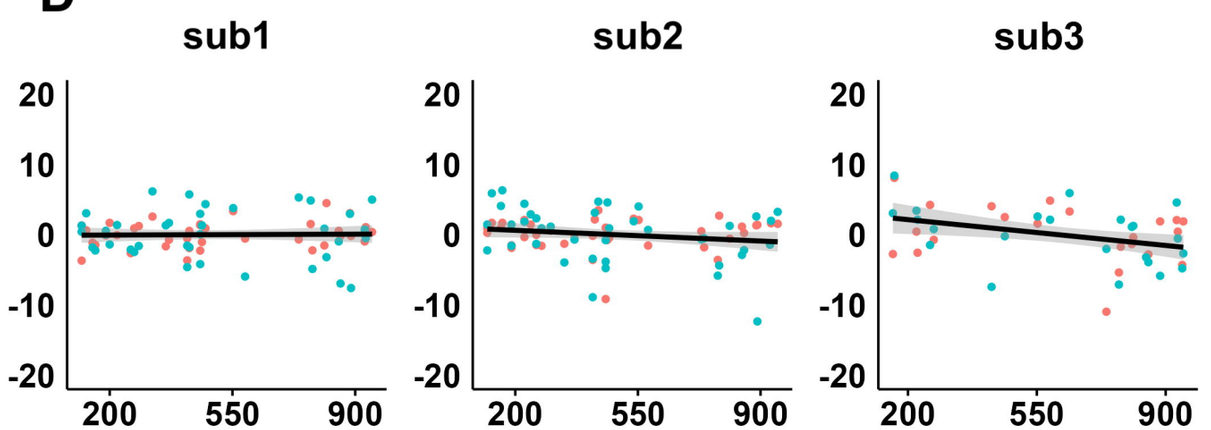
B

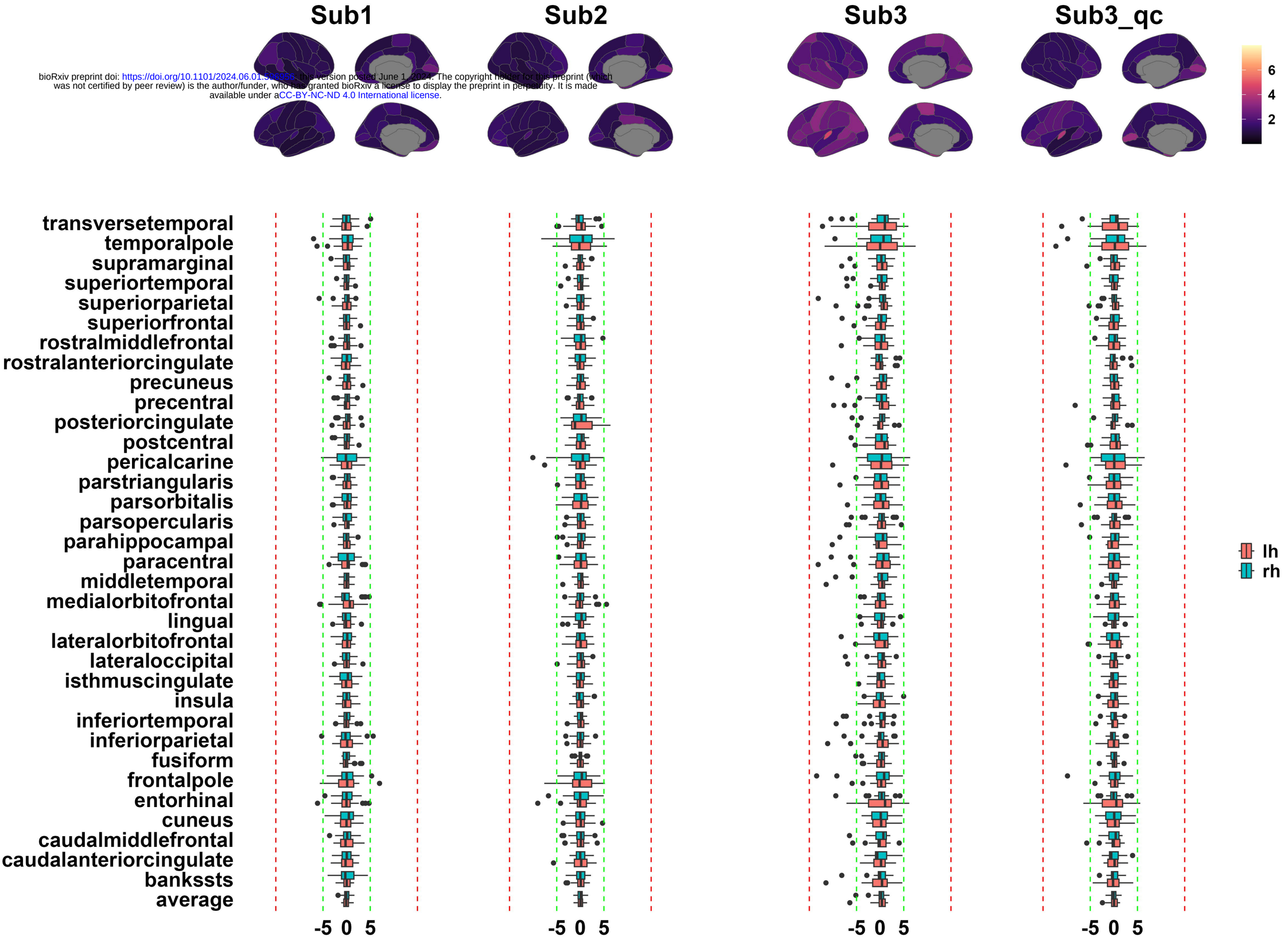

lh
rh
transversetemporal
frontalpole
rostralmiddlefrontal
superiorfrontal

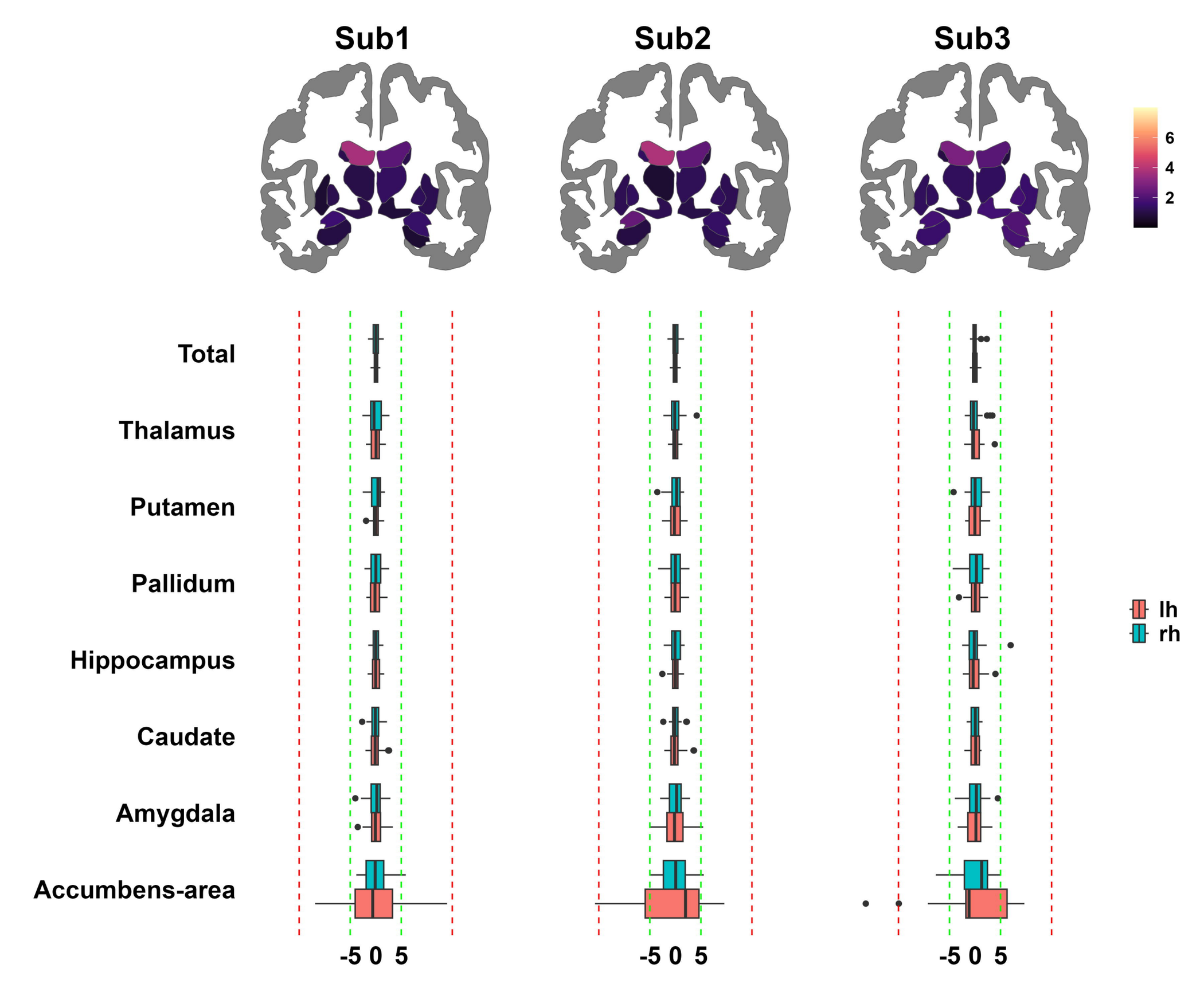
C

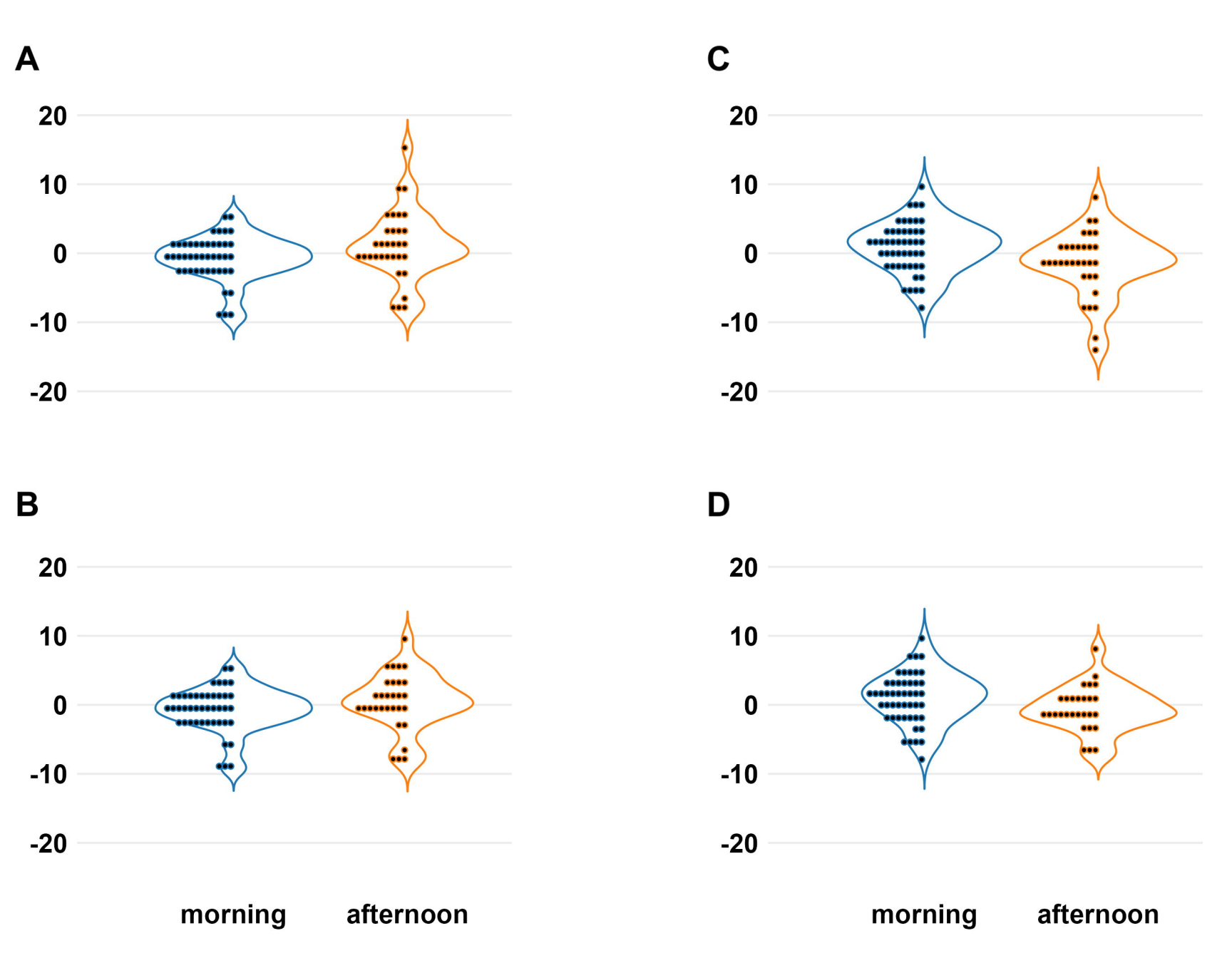

cnr
snr_total
wm2max
cjv
efc
inu_med
fber
fwhm_avg
rpve_gm

D


sub1
sub2
sub3




Sub1
Sub2
Sub3
6
4
2
Total
Thalamus
Putamen
Pallidum
Hippocampus
Caudate
Amygdala
Accumbens-area
lh
rh
-5 0 5
-5 0 5
-5 0 5


A
20
10
0
-10
-20
B
20
10
0
-10
-20
morning
afternoon
C
20
10
0
-10
-20
D
20
10
0
-10
-20
morning
afternoon