



## RESEARCH ARTICLE

## Predicting Age Groups using Brain Imaging Quality Data

Kadir Uludag, MS<sup>1,2\*</sup>

<sup>1</sup>CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, China

<sup>2</sup>Department of Psychology, University of Chinese Academy of Sciences, China

\*Corresponding author: Kadir Uludag, MS, CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Chaoyang District, Beijing 100101, China; Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, Tel: +86-1840-165-3162



### Abstract

Brain imaging quality of data is important to confirm their liability of brain imaging studies. Previous literature confirmed that some confounding factors such as movement, age, and gender may impact brain imaging quality. Automatic quality control (QC) applications may not be able to properly calculate their reliability due to confounding factors. There are a few studies on brain imaging quality data and relevant confounding factors such as age or gender.

**Methods:** Open data from a previous study was used to conduct this study. In total 26 participants were recruited. Random Forest (RF) and Neural Networks (NN) machine learning (ML) methods were used to predict age groups (cut-off age: 16). Patients were grouped by age groups. Then, the age group was predicted with RF and NN machine learning (ML) models.

**Goal of study:** The goal of the study was to predict age groups using brain imaging quality data.

**Results:** We found that according to NNs, the age group was predicted with an accuracy of over 60% (accuracy: 64%, sensitivity: 50%, specificity: 71%, area under curve (AUC): 55%). Furthermore, the RFML model found that the age group was predicted with an accuracy of 64% (sensitivity: 50%, specificity: 71%, AUC: 86.6%).

**Conclusion:** Our study showed that age groups can be predicted using the brain imaging quality of the data. Further studies should investigate the relationship between other brain imaging parameters related to the quality of data and age.

### Keywords

Schizophrenia, Brain imaging, Brain imaging quality, Prediction, Age, Brain age, Visual inspection

### Abbreviations

SCZ: Schizophrenia; NN: Neural Networks; RF: Random Forest; QC: Quality Control; MRI: Magnetic Resonance Imaging; HC: Healthy Controls; fMRI: Functional MRI; ML: Machine Learning

### Introduction

Neuroimages must be checked for potential distortion after processing [1]. Previously, the impact of data quality on the localization of brain activation in functional magnetic resonance imaging (fMRI) was investigated in-depth [2]. Also, some studies concluded that low-quality imaging data may cause false associations. Many techniques such as deep learning [3] and random forest were used to calculate the image quality of the human brain data whilst age is related to in-scanner motion and data quality [4]. Furthermore, several factors are related to MR image quality assessments such as geometric accuracy, high-contrast resolution, slice thickness accuracy, and slice position accuracy [5]. It is also important to analyze the association of parameters with age groups since it is one of the potential confounding factors.

Taken together, this study aimed to predict age groups using brain imaging quality data created using MRI QC.

### Methods

To provide measures describing the quality of this dataset, the data were analyzed using MRIQC [3]. The open fMRI Data from the previous study was used to conduct this study.

### Participants

This data was obtained from the Open fMRI database. The fMRI data is available at (<https://exhibits.stanford.edu/data/catalog/xg798vw8719>) [6].

## Measuring quality of machine learning image

MRIQC was used to measure the quality of brain imaging.

## Machine learning results

In this study, Neural Networks (NN) and Random Forest (RF) were used to predict age groups.

### Neural networks

Neural networks (NN) predicted age groups. Scikit-learn default parameters were used to create prediction models [7].

**Random forest:** Random Forest (RF) [8] is one of the popular ML algorithms used to predict age groups in this study. Scikit-learn default parameters were used to create prediction models [7].

### Data analysis

Participants were selected from the Open fMRI data data base. There was no missing value to report based on the shared data. The threshold of 0.05 was considered significant. Independent Sample T-test and Mann Whitney tests were used according to the normality of the study. Scikit-learn package of Python was used to create ML model [7].

## Results

In this study, the following brain imaging quality parameters were used to predict age groups: summary\_mean\_bg, summary\_mean\_csf, summary\_mean\_gm, summary\_mean\_wm, summary\_p05\_bg, summary\_p05\_csf, summary\_p05\_gm.

summary\_p05\_wm, summary\_p95\_bg, summary\_p95\_csf, summary\_p95\_gm, summary\_p95\_wm, summary\_stdv\_bg, summary\_stdv\_csf, summary\_stdv\_g, and summary\_stdv\_wm.

### Descriptive statistics

The mean age of the all participants was "16.89". The standard deviation was "4.26". The minimum age was "8.8" while the maximum age was "25.6". Most of the patients who participated in this study were female (n = 14) while 13 of the patients were male.

### Output of machine learning analysis

In this study, random forest (RF) and neural networks (NN) were implemented to predict age groups.

**Random forest:** The RF ML model found that the age group was predicted with an accuracy of 64% (sensitivity: 50%, specificity: 71%, AUC:86.6%). The most important predictive features were Summary\_p05\_wm, summary\_p05\_csf, summary\_stdv\_csf, summary\_mean\_wm, and summary\_p05\_gm respectively. For the RF model mean cross-val accuracy was  $61 \pm 0.12$ .

**Neural networks:** Based on the NN model, the accuracy of ML model was 64% (sensitivity: 50%,

specificity: 71%, AUC: 55%). The most important predictive features were Summary\_p95\_csf, summary\_mean\_wm, summary\_p95\_bg, summary\_mean\_csf, and summary\_stdv\_bg respectively. For the NN model mean cross-val accuracy was  $58.6 \pm 0.26$ .

## Discussion

Based on our main results of the study, we found that age groups can be predicted with both Random Forest (RF) and Neural Networks (NN) with the same accuracy over chance level.

Brain imaging data is not usable for extracting meaningful information without processing [9] and it should have some quality standards to be used.

The performance of the quality control (QC) strategies depended on the morphological measure [10]. Automated QC is important where visual QC is not practical [11]. Manual quality control strategies are reliable for QC of brain segmentation [10] however, manually controlling may take time. Furthermore, QC protocols allow different laboratories to search the impact of QC on the relationship between the brain and phenotypes [12].

Besides, The degree of anonymization of the data is important [13]. Revealing brain imaging quality data may allow researchers to predict their age groups. Furthermore, the age-dependent increase was found in brain because of the changes related to myelination.

Consistent with the findings that activity from almost all areas of the brain became less predictable with increased age [14]. Moreover, the ability to evaluate MRIs for disease characterization is hampered by the artifacts [15].

### Age-related potential changes in brain

Studies on such "brain age prediction" vary widely by their methods [16]. Developmental processes that occur during the first two decades of human development impact brain development (e.g. body growth and puberty) [17]. In addition, Age can be a confounding factor in some studies [18]. Further, the percent signal change associated with the BOLD effect increases with age in children ages [19]. Taken together, age is one of the important confounding factors that potentially impact brain imaging quality data.

## Conclusion

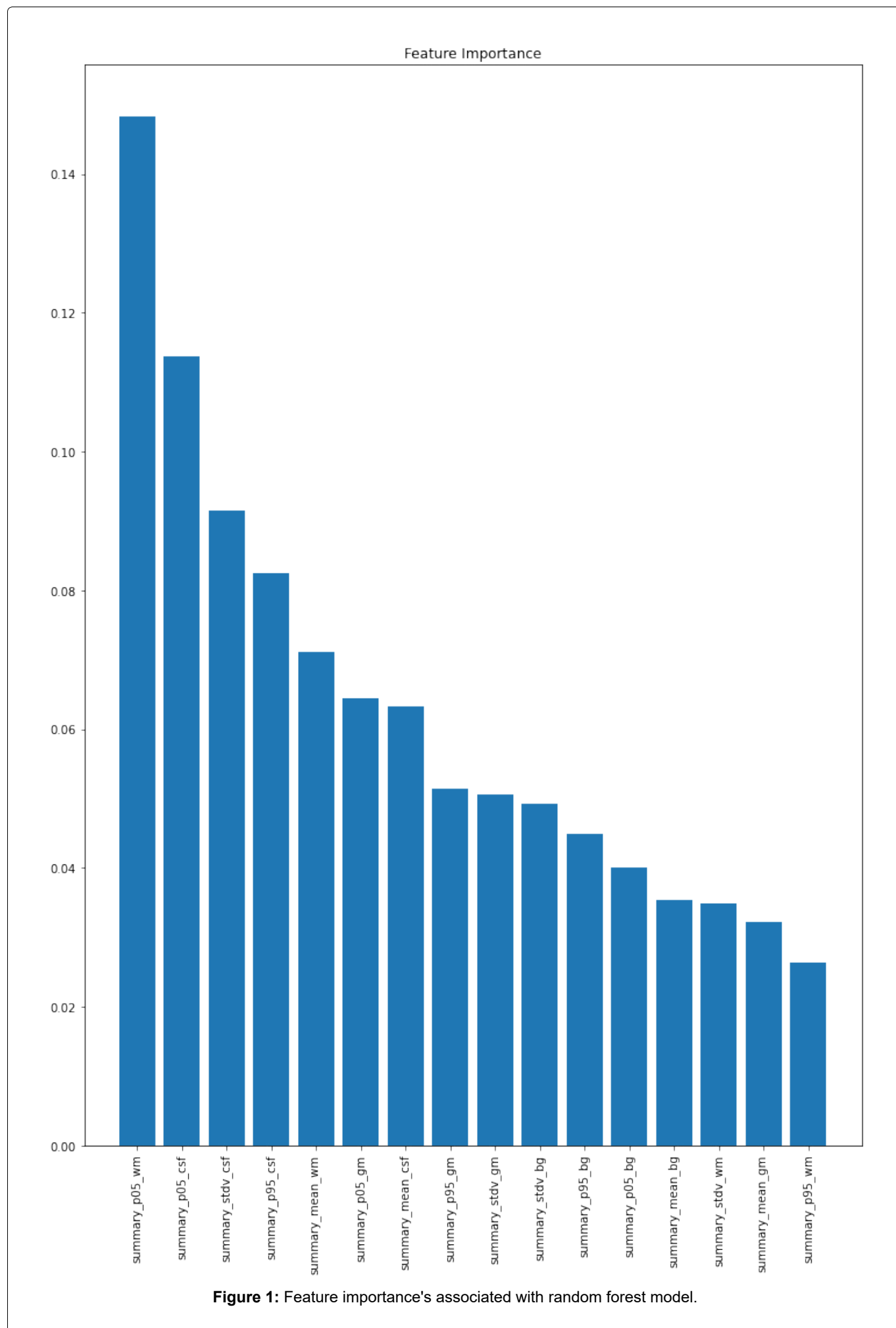
This study concluded that age groups can be predicted with high accuracy with RF and NN ML algorithms using brain imaging quality data.

## Suggestions for further studies

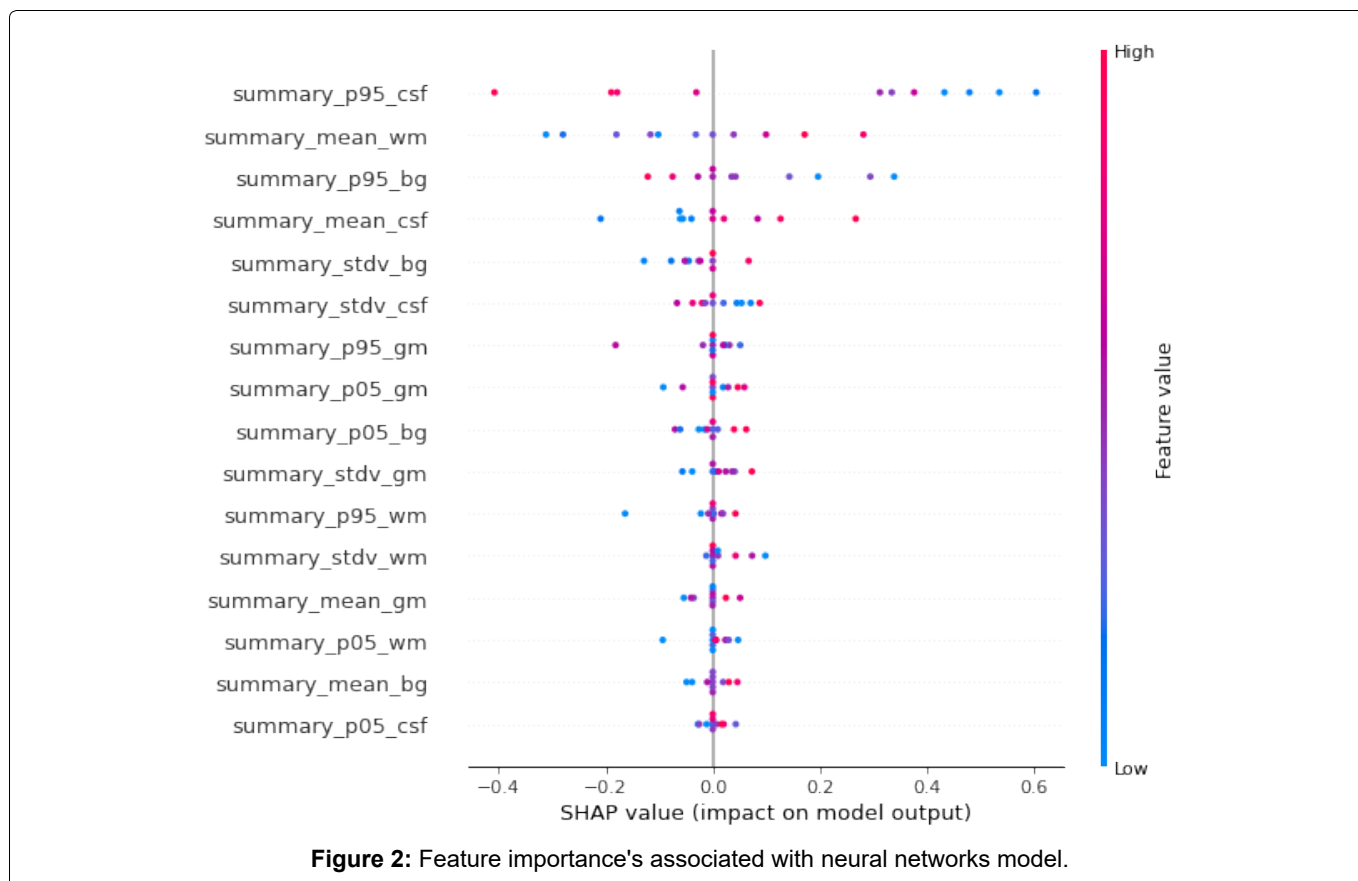
Further studies should investigate other confounding factors that may affect brain imaging quality.

## Limitations

There were several limitations in the current study. Gender was one of the important confounding factors.



**Figure 1:** Feature importance's associated with random forest model.



**Table 1:** Confusion matrix associated with random forest model.

|  |        |
|--|--------|
|  | [2, 2] |
|  | [2, 5] |

**Table 2:** Confusion matrix associated with neural networks model.

|  |        |
|--|--------|
|  | [2, 2] |
|  | [2, 5] |

**Table 3:** Accuracy scores (Random forest).

|              | Precision | Recall | f1-Score |
|--------------|-----------|--------|----------|
| 0            | 0.50      | 0.50   | 0.50     |
| 1            | 0.71      | 0.71   | 0.71     |
| accuracy     |           |        | 0.64     |
| macro avg    | 0.61      | 0.61   | 0.61     |
| weighted avg | 0.64      | 0.64   | 0.64     |

**Table 4:** Accuracy scores (Neural networks).

|           | Precision | Recall | f1-Score |
|-----------|-----------|--------|----------|
| 0         | 0.50      | 0.50   | 0.50     |
| 1         | 0.71      | 0.71   | 0.71     |
| accuracy  |           |        | 0.64     |
| macro avg | 0.61      | 0.61   | 0.61     |

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Acknowledgments

Thanks to the participants who made this study possible.

## Funding

There is no funding associated with this study. KU was supported by Chinese Government Scholarship provided for PhD students.

## Code availability

The MRIQC Web API is available under the Apache-2.0 license. The source code is accessible through GitHub (<https://github.com/poldracklab/mriqcwebapi>).

The fMRI data is available at (<https://exhibits.stanford.edu/data/catalog/xg798vw8719>) [6] (Table 1, Table 2, Table 3, Table 4) and (Figure 1, Figure 2).

## Funding

Kadir Uludag is funded by Chinese Government Scholarship Type-A.

## Competing Interests

No potential competing interest was reported by the authors.

## References

- Woodard JP, Carley-Spencer MP (2006) No-Reference image quality metrics for structural MRI. *Neuroinformatics* 4: 243-262.
- Geissler A, Gartus A, Foki T, Tahamtan AR, Beisteiner R, et al. (2007) Contrast-to-noise ratio (CNR) as a quality parameter in fMRI. *J Magn Reson Imaging* 25: 1263-1270.
- Esteban O, Birman D, Schaer M, Koyejo OO, Poldrack RA, et al. (2017) MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* 12: e0184661.

4. Rosen AFG, Roalf DR, Ruparel K, Blake J, Seelaus K, et al. (2018) Quantitative assessment of structural image quality. *Neuroimage* 169: 407-418.
5. Chen CC, Wan YL, Wai YY, Liu HL (2004) Quality assurance of clinical MRI scanners using ACR MRI phantom: Preliminary results. *J Digit Imaging* 17: 279-284.
6. Padmanabhan A, Geier CF, Ordaz SJ, Teslovich T, Luna B (2011) Developmental changes in brain function underlying the influence of reward processing on inhibitory control. *Dev Cogn Neurosci* 1: 517-529.
7. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. (2011) Scikit-learn: Machine learning in Python. *Journal of machine Learning research* 12: 2825-2830.
8. Breiman L (2001) Random forests. *Machine Learning*, 45: 5-32.
9. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, et al. (2018) Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* 166: 400-424.
10. Monereo-Sánchez J, de Jong JJA, Drenthen GS, Beran M, Backes WH, et al. (2021) Quality control strategies for brain MRI segmentation and parcellation: Practical approaches and recommendations - insights from the Maastricht study. *Neuroimage* 237: 118174.
11. Bastiani M, Cottaar M, Fitzgibbon SP, Suri S, Alfaro-Almagro F, et al. (2019) Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *Neuroimage*, 184: 801-812.
12. Benhajali Y, Badhwar A, Spiers H, Urchs S, Armoza J, et al. (2020) A standardized protocol for efficient and reliable quality control of brain registration in functional MRI studies. *Front Neuroinform* 14: 7.
13. White T, Blok E, Calhoun VD (2022) Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Hum Brain Mapp* 43: 278-291.
14. Mišić B, Mills T, Taylor MJ, McIntosh AR (2010) Brain noise is task dependent and region specific. *J Neurophysiol* 104: 2667-2676.
15. Gupta A, Sadri AR, Viswanath SE, Tiwari P (2020) Quality assessment of brain MRI scans using a dense neural network model and image metrics. Paper presented at the Medical Imaging 2020: Physics of Medical Imaging.
16. Baecker L, Dafflon J, da Costa PF, Garcia-Dias R, Vieira S, et al. (2021) Brain age prediction: A comparison between machine learning models using region- and voxel-based morphometric data. *Hum Brain Mapp* 42: 2332-2346.
17. Mills KL, Tamnes CK (2014) Methods and considerations for longitudinal structural brain imaging analysis across development. *Dev Cogn Neurosci* 9: 172-190.
18. Alfaro-Almagro F, McCarthy P, Afyouni S, Andersson JLR, Bastiani M, Miller, et al. (2021) Confound modelling in UK Biobank brain imaging. *Neuroimage* 224: 117002.
19. Schapiro MB, Schmithorst VJ, Wilke M, Byars AW, Strawsburg RH, et al. (2004). BOLD-fMRI signal increases with age in selected brain regions in children. *Neuroreport* 15: 2575-2578.