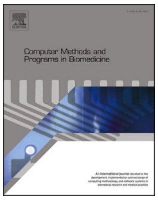
Computer Methods and Programs in Biomedicine 242 (2023) 107825

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb



The effect of processing pipelines, input images and age on automatic cortical morphology estimates

Giulia Debiasi [a,1], Ilaria Mazzonetto [a,1], Alessandra Bertoldo [a,b,*]

[a] *Department of Information Engineering, University of Padova, via Gradenigo 6/b, Padova 35131, Italy*
[b] *Padova Neuroscience Center (PNC), University of Padova, via Orus 2/b, Padova 35131, Italy*

ARTICLE INFO

Keywords:
Cortical thickness
FreeSurfer
Advanced Normalization Tools
Human Connectome Project
Ageing

ABSTRACT

Background and Objective: Magnetic resonance imaging of the brain allows to enrich the study of the relationship between cortical morphology, healthy ageing, diseases and cognition. Since manual segmentation of the cerebral cortex is time consuming and subjective, many software packages have been developed. FreeSurfer (FS) and Advanced Normalization Tools (ANTs) are the most used and allow as inputs a T1-weighted (T1w) image or its combination with a T2-weighted (T2w) image. In this study we evaluated the impact of different software and input images on cortical estimates. Additionally, we investigated whether the variation of the results depending on software and inputs is also influenced by age.

Methods: For 240 healthy subjects, cortical thickness was computed with ANTs and FreeSurfer. Estimates were derived using both the T1w image and adding the T2w image. Significant effects due to software, input images and age range were investigated with ANOVA statistical analysis. Moreover, the accuracy of the cortical thickness estimates was assessed based on their age-prediction precision.

Results: Using FreeSurfer and ANTs with T1w or T1w-T2w images resulted in significant differences in the cortical thickness estimates. These differences change with the age range of the subjects. Regardless of the images used, the more recent FS version tested exhibited the best performances in terms of age prediction.

Conclusions: Our study points out the importance of i) consistently processing data using the same tool; ii) considering the software, input images and the age range of the subjects when comparing multiple studies.

1. Introduction

Magnetic resonance imaging (MRI) based structural information, such as those obtainable with T1-weighted (T1w) and T2-weighted (T2w) images, are essential for tissue segmentation, anatomical delineation, and lesion location. This information has proved to be a valid tool to study the healthy ageing as well as to follow the progression of many neurological disorders [1–3]. Therefore, cortical thickness, cortical volume and cortical surface area measurements need to be as reliable as possible. The use of automated methods for their estimation has improved quantitative neuroimaging analysis, by introducing less subjectiveness and fastest processing of data. Cortical thickness computation can be performed by several tools which are mainly divided in two types, namely surface- and volume-based methods. Surface-based modelling reconstructs the cortical surfaces by means of polygonal meshes [4,5]. Instead, volume-based methods deform the White Matter (WM)/Grey Matter (GM) boundary towards the GM/cerebrospinal fluid (CSF) boundary, and derive a cortical thickness map through a diffeomorphic registration [6]. FreeSurfer (FS) [7] is a representative of the former typology, while the Advanced Normalization Tools (ANTs) cortical thickness pipeline [8] is a representative of the latter. Both FreeSurfer and ANTs have been used to analyse cortical

---

Abbreviations: ANTs, (Advanced Normalization Tools); CSF, (Cerebrospinal Fluid); DKT, (Desikan-Killany-Tourville); eTIV, (estimated Total Intracranial Volume); FS, (FreeSurfer); FS6, (FreeSurfer version 6.0); FS7, (FreeSurfer version 7.1.1); FS6 T1, (FreeSurfer version 6.0 with T1w image as input); FS6 T1T2, (FreeSurfer version 6.0 with T1w and T2w images as input); FS7 T1, (FreeSurfer version 7.1.1 with T1w image as input); FS7 T1T2, (FreeSurfer version 7.1.1 with T1w and T2w images as input); GM, (Grey Matter); HCP-A, (Human Connectome Project in Aging); MRI, (Magnetic Resonance Imaging); ROI, (Region Of Interest); RMSE, (Root Mean Square Error); T1w, (T1-weighted); T2w, (T2-weighted); WM, (White Matter).

* Corresponding author at: Department of Information Engineering, University of Padova, via Gradenigo 6/b, Padova 35131, Italy
E-mail address: bertoldo@dei.unipd.it (A. Bertoldo).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cmpb.2023.107825
Received 4 June 2023; Received in revised form 1 September 2023; Accepted 20 September 2023

thickness in both healthy and pathologic states [9–13]. Regarding FreeSurfer, the computed volumes of specific brain structures have been compared to manually derived volumes [14,15], as well as cortical thickness estimates have been validated with manual measurements [16,17] and histological analysis [18]. Also post-mortem data [4] was used for validation, resulting in good agreement between the two measurements with a maximum discrepancy of slightly more than 0.5 mm. As reported by the authors, differences may be accounted for by several factors such as individual variability, fixation effects, the precise location of the measurements, as well as MRI artifacts. Moreover, several studies have demonstrated the reliability of cortical measurements between scanner systems, albeit with sensitivity to scanner manufacturer and field strengths [19–21]. In this context, FreeSurfer has proven to be reliable at following morphological and pathological changes in the brain [22]. Over time, multiple software releases have been made available, due to ongoing improvements in the implementation of the processing steps. The issue that rises with a new release is that it could compromise the reproducibility of previous studies and could affect clinical translation. Gronenschild et al. [23] addressed this issue with the aim of testing the reliability of FreeSurfer across software versions (v4.3.1, v4.5.0, and v5.0.0), workstation types and operating systems: they found significant differences in volume and cortical thickness estimates. Similar findings were reported by Bigler et al. [24] that observed significant total volumes differences, and minimal significant differences between cortical thickness measurements. In another study performed by Chepkoech et al. [25], authors were interested in understanding whether classification accuracy, based on cortical thickness, changes depending on the version used rather than finding absolute differences that are somehow expected. They found that, even though there were absolute thickness estimation differences across versions, those differences did not affect classification accuracy. Recently, Haddad et al. [26] reported a good compatibility of cortical surface area from version 7.1 with that of previous versions, and moderate compatibility for the cortical thickness. Despite the well-known issue that different cortical thickness estimation approaches can yield different results [8,27–29], to the best of our knowledge there are no studies that compare recent FreeSurfer versions with ANTs. The main aim of our study is to verify whether there are significant differences in cortical thickness estimates between FreeSurfer and ANTs, and to determine if these differences equally affect different brain regions. If so, it is important to consider this aspect when comparing studies that exploit one of the two tools. Such comparison has already been performed by Tustison et al. [8] with a previous version of FreeSurfer and the native ANTs cortical thickness pipeline, resulting in good scan-rescan repeatability for both methods, and higher performance for ANTs in predicting age and sex. Both methods require as input of the main pipeline a T1w image, but they also provide the option to include a T2w image, that is used in specific processing steps. Several studies [30,31] report that the inclusion of a T2w image, due to the different contrast, can improve the reliability of hippocampal subfield measurements. Based on these results, we hypothesized that a similar improvement might also be achievable in the cortical estimates. Hence, we quantified for the first time the differences in the cortical estimates when using a T2w image as additional input. Finally, considering that progression of some diseases (i.e., neurodegenerative diseases) occurs over a certain period and at different ages, we investigated the influence of age on cortical estimates obtained from the different software.

2. Materials and methods

For this study, 240 healthy subjects from the Human Connectome Project in Aging (HCP-A, Release 2.0) [32] were chosen. The selection of the subjects was performed to create three different age groups of 80 participants each, matched for sex (40/40 M/F) and Quality Control score of the structural images available in the HCP-A release documentation. Imaging data of the HCP-A dataset were acquired on a Siemens 3T Prisma. T1w and T2w images were acquired using a multi-echo MPRAGE (TR/TI = 2500/1000 ms, TE = 1.8/3.6/5.4/7.2 ms, flip angle = 8°), and a variable-flip-angle turbo-spin-echo (TSE) sequence (TR/TE = 3200/564 ms, turbo factor = 314), respectively. Both structural scans had a sagittal FOV of 256 × 240 × 166 mm^3 with a matrix size of 320 × 300 × 208 slices) [33]. Subjects were divided in mature (from 36 to 64 years old), old (from 65 to 79 years old) and oldest old (from 80 years old) groups [32]. T1w and T2w images of each participant have been minimally preprocessed according to Glasser et al. [34] and Harms et al. [33]. All subjects were analysed with two FreeSurfer versions (6.0 and 7.1.1) and ANTs (version 2.3.0). Each software was run with different inputs: both a T1w and T2w images, as well as solely a T1w image (considered the standard input). Cortical features can be extracted using the FreeSurfer software through its recon-all pipeline. To improve the segmentation of the structural image, a low-frequency intensity nonuniformity correction is applied. Up to FreeSurfer version 6.0 (FS6), this step relied on the nonparametric nonuniform normalization (N3) approach [35]. In subsequent versions, the algorithm was enhanced with a robust B-spline approximation algorithm and a modified optimization strategy. This improved version, known as N4, has been shown to exhibit greater noise robustness, and improved convergence due to its iterative refinement scheme [36]. Following the segmentation of cortical and subcortical structures, white and pial surfaces are generated [37]. T2w image can be used at this step, contributing to the refinement of the pial surface thanks to its different contrast. For the analysis we extracted the mean cortical thickness, cortical surface area and volume for each region of interest (ROI) of the Desikan-Killany-Tourville (DKT) atlas [38]. Cortical thickness is computed as the distance between two corresponding points in the white matter and pial surface [4], whereas surface area is the average of areas of all faces that meet at a given vertex on the white matter surface [39]. In the case of volume computation, it is important to note that it does not simply involve the multiplication of area by thickness at each vertex. Briefly, the vertex coordinates of the matching faces of white matter and pial surface are used to define an oblique truncated triangular pyramid which can be decomposed into three tetrahedra. Volume is computed as the sum of the volumes of these polyhedrons [40]. Cortical surface area and volume were normalized by the total white surface area and by the estimated Total Intracranial Volume (eTIV), respectively. With ANTs, only the native cortical thickness pipeline (version 2.3.0) was tested. Briefly, the ANTs cortical thickness pipeline (antsCorticalThickness.sh) starts with N4 bias field correction [36] and then performs brain extraction [41]. The T2w image can be used in the brain segmentation process which is based on spatial tissue priors [42]. Cortical thickness is computed based on diffeomorphic mappings between the grey/white matter and the exterior cortical surface [6]. To parcellate the brain according to the DKT atlas, we applied the joint label fusion method [43] using 20 manually labelled subjects of the MindBoggle-101 dataset [38]. Due to the volume-based framework of ANTs, which does not include the cortical surfaces reconstruction, for each ROI, we only derived the cortical thickness, computed as the mean cortical thickness across the voxels of that region.

The workflow of the present study is shown in Fig. 1. A total of six approaches were implemented: FS6 with T1w (FS6 T1), FS6 with T1w and T2w (FS6 T1T2), FS7 with T1w (FS7 T1), FS7 with T1w and T2w (FS7 T1T2), ANTs with T1w (ANTS T1) and ANTs with T1w and T2w (ANTS T1T2). Four different comparisons were investigated: i) FS6 vs FS7 (both T1 and T1T2 approaches); ii) ANTS T1 vs ANTS T1T2; iii) FS6 T1 vs FS7 T1 vs ANTS T1; iv) FS6 T1T2 vs FS7 T1T2 vs ANTS T1T2. The analyses were performed with Matlab 2021b (MathWorks, MA, USA). In comparison (i) we conducted two three-way repeated measure ANOVA to investigate the impact of software version, input images and age range on cortical thickness as well as on volume, and a two-way repeated measure ANOVA to investigate the impact of software version and age range on cortical surface area. We did not investigate the impact of the input images on the white matter surface area as the

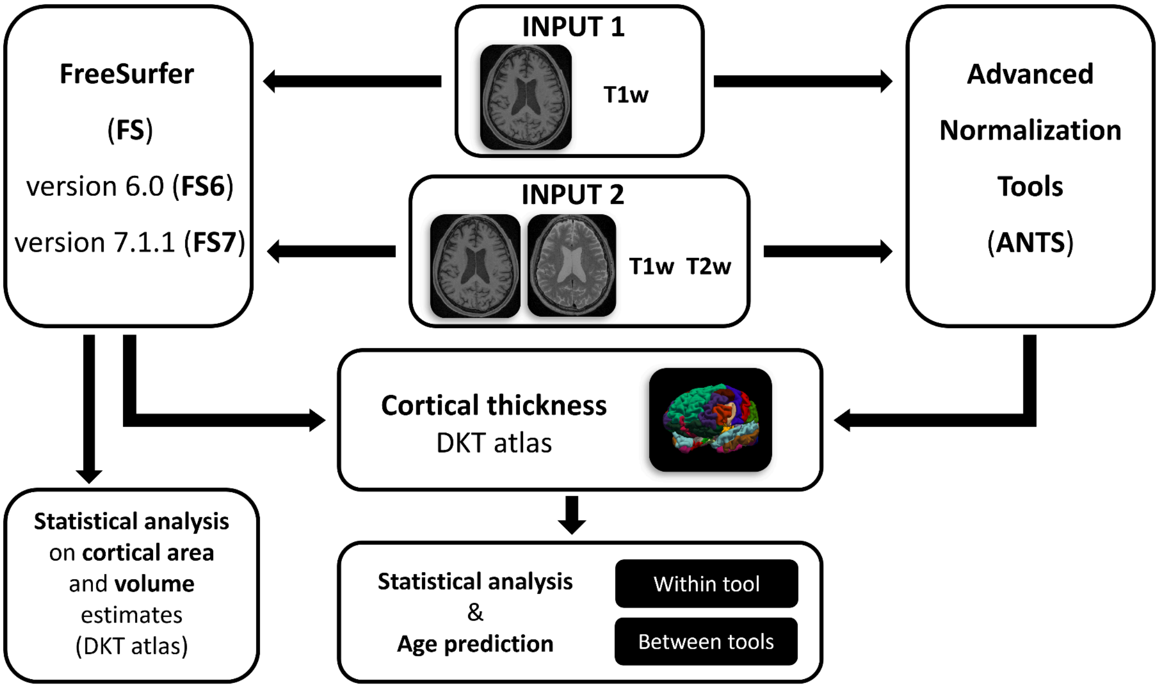


Fig. 1. Workflow of the study: T1w and T2w images were used as inputs of the software pipelines. Cortical thickness was computed using both software, whereas cortical surface area and volume were exclusively obtained through FreeSurfer.

T2w images were introduced into the process after its computation. In comparison (ii) we conducted a two-way repeated measure ANOVA to investigate the effects of input images and age range on cortical thickness. In comparisons (iii) and (iv) we conducted a two-way repeated measure ANOVA to investigate the effects of software and age range on cortical thickness. It is worth noting that area and volume were exclusively considered in the first comparison, as it was the only one focused solely on FreeSurfer. Following each ANOVA analysis, we performed post-hoc tests (specifically, the Tukey Kramer test) to infer the significative differences among the levels of tested factors. For comparisons iii) and iv) we also computed the mean cortical thickness differences (ANTS - FS) across ROIs. Moreover, to infer whether differences in the cortical thickness estimation are local or global, for each ROIs we computed correlations between cortical thickness measurements derived for each couple of two distinct approaches (i.e., same software with different input images or the same input images but with different software). Furthermore, given that there is no ground truth for cortical thickness data, an indirect method is required to assess the quality of the estimation. Since it is known that there is a progressive thinning of the cortex as age increases [17], we gauged the accuracy of cortical thickness estimations through their capability to predict age. To achieve this, we fitted the linear model previously used in the study by Tustison et al. [8]:

$$Age \sim eTIV + Gender + \sum_{i=1}^{n\,ROI} \overline{CT_i}$$

where $eTIV$ is the estimated Total Intracranial Volume, $\overline{CT_i}$ is the mean cortical thickness of the ROI i of the DKT atlas. For each comparison, cortical thickness data from the involved approaches were partitioned into training and test sets using varying proportions of training data (ranging from 10 % to 90 %). This procedure was repeated for $n = 1000$ permutations in order to generate a performance distribution using the Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum \left(AGE_{true} - AGE_{predicted}\right)^2}{N}}$$

Finally, a ROI-level analysis of cortical thickness measurements was carried out for the T1w-based approach of FS7 and ANTs, owing to their more extensive utilization compared to the T1w-T2w-based approach. Consequently, a paired Student's t-test with False Discovery Rate (FDR) correction was applied to the mean cortical thickness of the subjects' ROI of the three age groups with FS7 T1 and ANTS T1 approaches.

3. Results

3.1. Comparison (i): FreeSurfer-based approaches

The effect of FreeSurfer version and age on the area estimates were assessed with a two-way repeated measures ANOVA. The analysis revealed a significant effect due to age (F(2)=72.0, $p<0.05$, where F refers to the F-test, and 2 is equal to the degrees of freedom). Post-hoc test revealed that cortical surface area decreases with age ($p<0.05$), regardless of the version used. Similar analysis, also testing the effect of the input image, were conducted for volume and cortical thickness estimates. Regarding volume, ANOVA revealed significant effects due to software version ($F(1)=11.4$, $p<0.05$), image used ($F(1)=34.4$, $p<0.05$) and age ($F(2)=82.0$, $p<0.05$). Post-hoc tests indicated that in the old and oldest old groups, values derived from FS6 were lower than those from FS7 ($p<0.05$). Additionally, regardless of age, measurements derived from T1w images were lower than those derived from T1w-T2w images ($p<0.05$).

Lastly, cortical volume decreased with age regardless of the version and image used ($p<0.05$). In the supplementary materials (Fig. S1), we depicted the mean volume of the ROIs computed using different approaches for the three age groups. It could be seen that the spatial patterns of the volume measurements were consistent across age groups and, as previously mentioned, there was a reduction in volume with increasing age. The ROIs with the most pronounced differences among the approaches included the fusiform, inferior parietal, inferior temporal, lateral occipital, lateral orbito-frontal and lingual regions. The spider plots in Fig. 2 show the mean cortical thickness of the atlas ROIs estimated using FS6 T1 (blue), FS6 T1T2 (orange), FS7 T1 (yellow) and FS7 T1T2 (violet) for each age group. Significant effects were found for

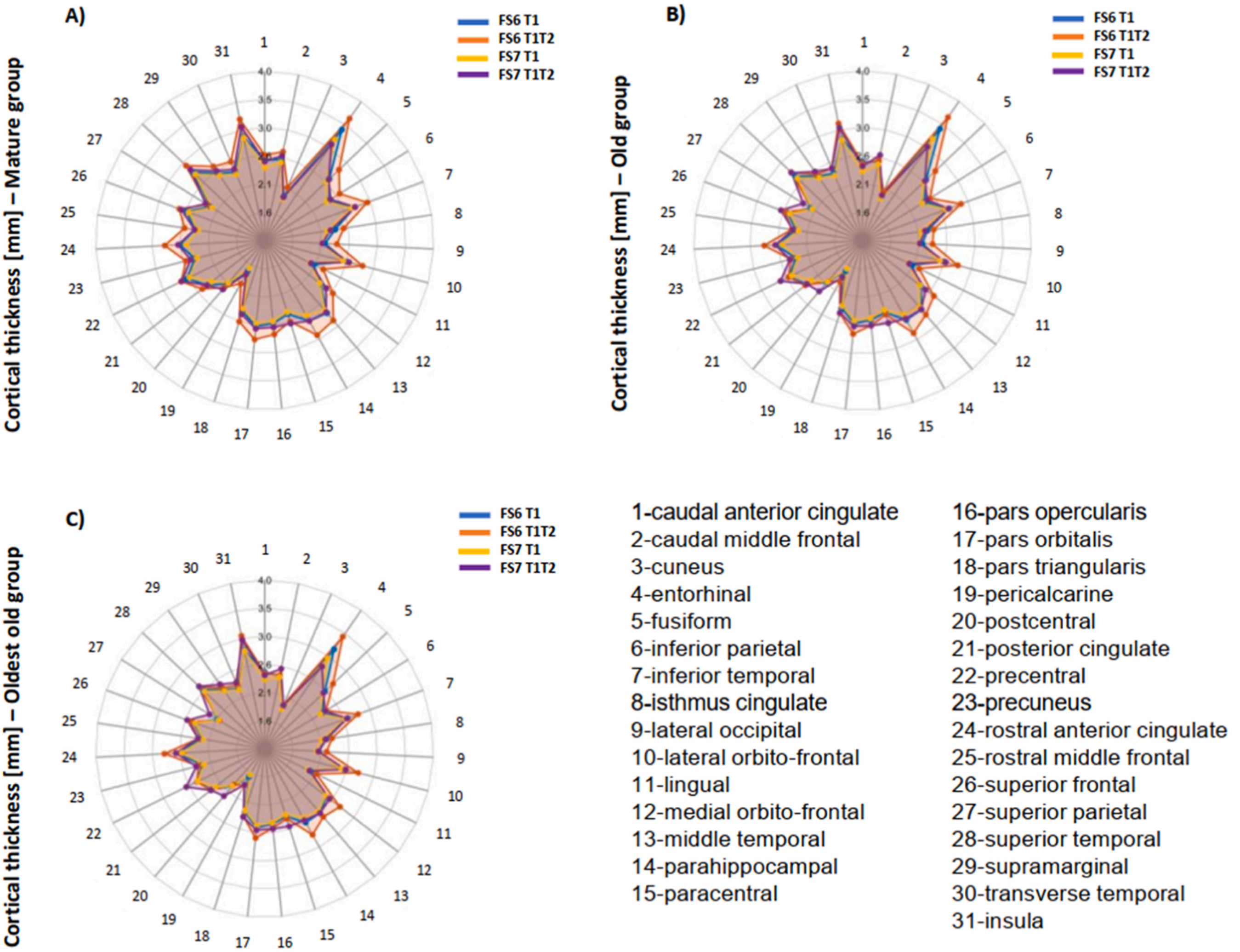


Fig. 2. Mean cortical thickness estimates derived from FreeSurfer approaches (comparison i)) and divided by age group. Panel A, B and C report mature, old and oldest old group mean cortical thickness values, respectively. Each point corresponds to a ROI of the DKT atlas as listed in the bottom right panel.

version ($F(1)=41.5$, $p<0.05$), image used ($F(1)=406.8$, $p<0.05$) and age ($F(2)=135.8$, $p<0.05$). Post-hoc tests reported that regardless of age, cortical thickness values derived from FS6 were greater than those from FS7 ($p<0.05$), and similar differences were observed between T1T2-derived values with respect to T1-derived ($p<0.05$). Furthermore, cortical thickness decreased with age, regardless of the version and image used ($p<0.05$). Mean values and standard deviations of correlations between the ROI cortical thickness estimates obtained using various combinations of software and input images are listed in Table 1. All the averaged correlations had $p<0.05$, except for the pericalcarine, medial orbito-frontal cortex, cuneus and isthmus cingulate, which did not reach statistical significance in four of the comparisons.

The most highly correlated approaches were FS7 T1-FS7 T1T2 ($r = 0.94$, $p<0.05$) while the least correlated were FS6 T1T2-FS7 T1T2 ($r = 0.69$, $p<0.05$). The results of the age prediction based on cortical thickness with a training portion equal to 50 % are summarized in Table 2 (rows: 1–4). FS7 had better predictive performance ($RMSE_{mean}\pm RMSE_{sd}$ FS7 T1: 15.13±0.69, FS7 T1T2: 15.11±0.69) than FS6 (FS6 T1: 15.93±0.69, FS6 T1T2: 16.00±0.71), and the inclusion of the T2w image into the pipeline did not improve this performance. The RMSE distributions of the different approaches at the various training portions are illustrated in the supplementary materials, specifically in Figs S2–S5.

Table 1
Mean ROIs cortical thickness correlations between various combinations of software/input images. Standard deviation is reported between parentheses.

	Correlation FS6 T1	FS7 T1	FS6 T1T2	FS7 T1T2	ANTS T1	ANTS T1T2
FS6 T1	1	0.93 (0.04)	0.84 (0.06)	0.87 (0.06)	0.50 (0.16)	–
FS7 T1	0.93 (0.04)	1	0.77 (0.07)	0.94 (0.03)	0.47 (0.16)	–
FS6 T1T2	0.84 (0.06)	0.77 (0.07)	1	0.69 (0.12)	–	0.59 (0.20)
FS7 T1T2	0.87 (0.06	0.94 (0.03)	0.69 (0.12)	1	–	0.36 (0.19)
ANTS T1	0.50 (0.16)	0.47 (0.16)	–	–	1	0.93 (0.05)
ANTS T1T2	–	–	0.59 (0.20)	0.36 (0.19)	0.93 (0.05)	1

3.2. *Comparison (ii): ANTs-based approaches*

The mean cortical thickness values obtained with ANTS T1 (blue) and ANTS T1T2 (orange) for the three age groups are shown in Fig. 3. The two-way repeated measures ANOVA on cortical thickness measurements revealed significant effects due to image used ($F(1)=1061.7$,

Table 2
RMSE of the six investigated approaches relative to a training portion of 50 %.

	RMSE mean [years]	RMSE sd [years]
FS6 T1	15.93	0.69
FS6 T1T2	16.00	0.71
FS7 T1	15.13	0.69
FS7 T1T2	15.11	0.69
ANTS T1	16.66	0.78
ANTS T1T2	15.60	0.76

$p<0.05$) and age ($F(2)=394.8$, $p<0.05$).

Post-hoc tests revealed that regardless of the image used, cortical thickness decreased with age ($p<0.05$). In the old and oldest old group, T1-derived values were higher than T1T2-derived values ($p<0.05$), while this distinction was not observed in the mature group. The mean correlation between ANTS-based approaches is reported in Table 1. Results of age prediction from cortical thickness with ANTs are listed in the last two rows of Table 2. The best predictive performance in terms of mean RMSE was achieved by ANTS T1T2 ($\text{RMSE}_{\text{mean}}\pm\ \text{RMSE}_{\text{sd}}$ 15.60 ±0.76).

3.3. Comparison (iii): T1w-based approaches

Mean cortical thickness measurements for FS6 T1 (blue), FS7 T1 (orange) and ANTS T1 (yellow) in the three age groups are presented in Fig. 4. A two-way repeated measures ANOVA revealed significant effects due to software ($F(2)=27.5$, $p<0.05$) and age ($F(2)=304.8$, $p<0.05$) on cortical thickness values obtained from the three approaches. Post-hoc tests indicated that cortical thickness decreased with age ($p<0.05$) regardless of the software used. Additionally, there were significant differences ($p<0.05$) among the values derived from the different approaches. Particularly, in the mature and old groups, the order was ANTS T1 > FS6 T1 > FS7 T1, whereas in the oldest old group there were not significant differences between FS6 T1 and ANTS T1 approaches. The mean differences in cortical thickness across the dataset are illustrated in Fig. 5. Cortical thickness estimates derived using ANTS were approximately 1 mm greater than those obtained with both FreeSurfer versions in the inferior temporal gyrus, the entorhinal region, the fusiform gyrus, the insula region and the rostral anterior cingulate cortex. Mean ROIs correlations for comparison iii) are reported in Table 1: weak correlation values were associated with FS6 T1-ANTS T1 (r = 0.50, $p<0.05$) and FS7 T1-ANTS T1 (r = 0.47, $p<0.05$) combinations. The results of the age prediction based on cortical thickness are summarized in Table 2 (rows: 1, 3, 6): the best predictive performance was achieved using the FS7 T1 approach ($\text{RMSE}_{\text{mean}}\pm\ \text{RMSE}_{\text{sd}}$: 15.13±0.69), whereas the least favourable results were obtained with the ANTS T1 approach ($\text{RMSE}_{\text{mean}}\pm\ \text{RMSE}_{\text{sd}}$: 16.66±0.78).

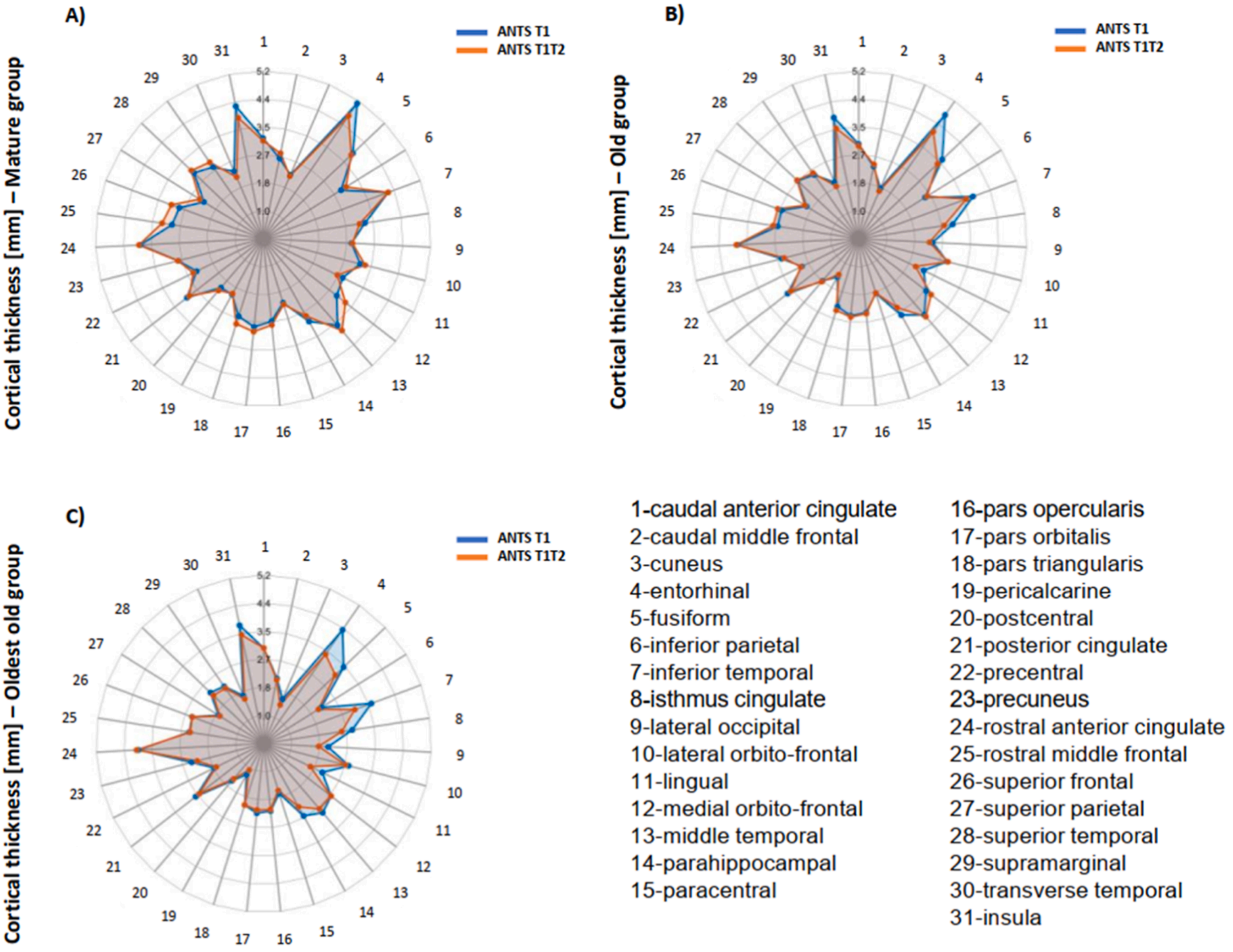
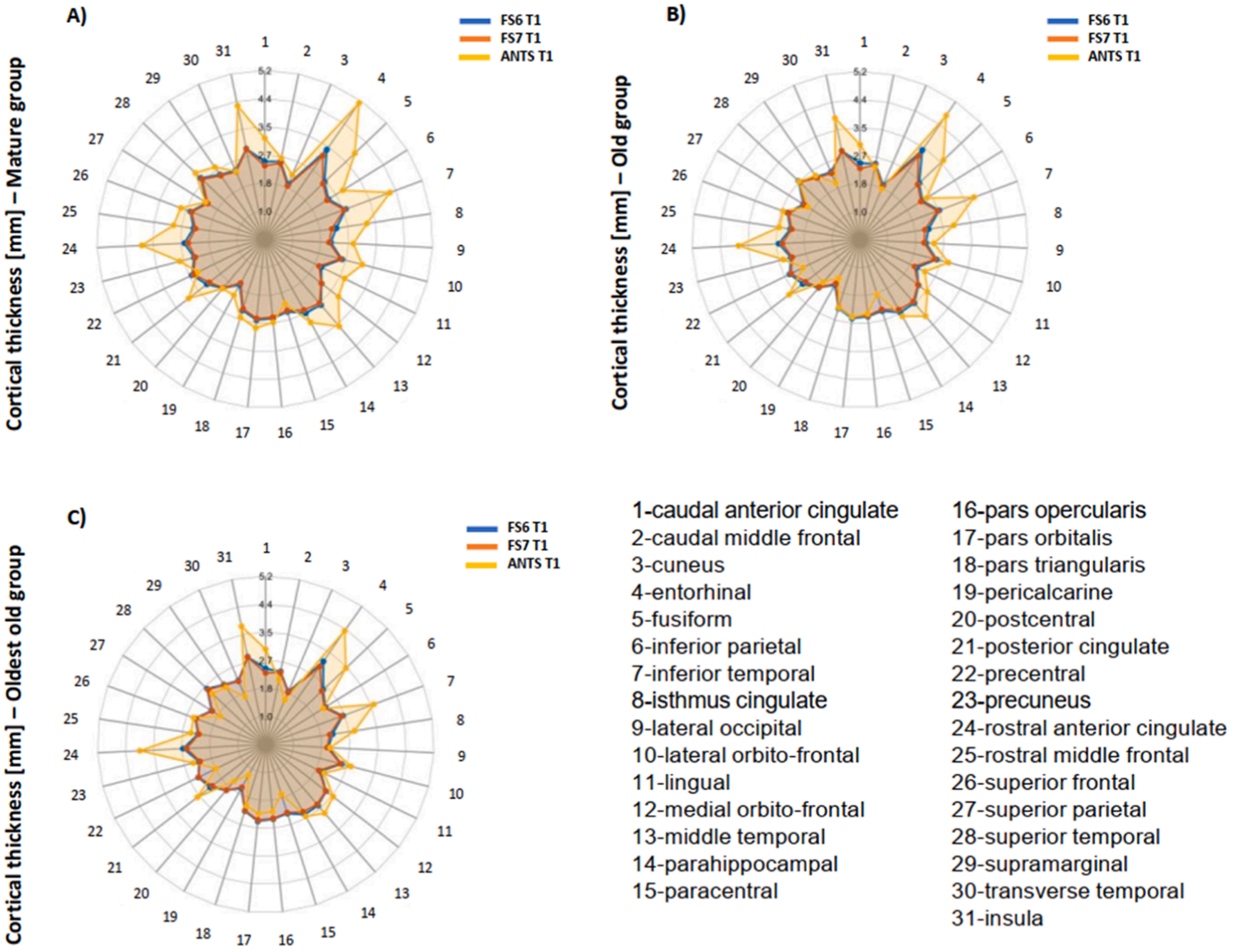


Fig. 3. Mean cortical thickness estimates derived from ANTS approaches (comparison ii)) and divided by age group. Panel A, B and C report mature, old and oldest old group mean cortical thickness values, respectively. Each point corresponds to a ROI of the DKT atlas as listed in the bottom right panel.



1-caudal anterior cingulate
2-caudal middle frontal
3-cuneus
4-entorhinal
5-fusiform
6-inferior parietal
7-inferior temporal
8-isthmus cingulate
9-lateral occipital
10-lateral orbito-frontal
11-lingual
12-medial orbito-frontal
13-middle temporal
14-parahippocampal
15-paracentral
16-pars opercularis
17-pars orbitalis
18-pars triangularis
19-pericalcarine
20-postcentral
21-posterior cingulate
22-precentral
23-precuneus
24-rostral anterior cingulate
25-rostral middle frontal
26-superior frontal
27-superior parietal
28-superior temporal
29-supramarginal
30-transverse temporal
31-insula

Fig. 4. Mean cortical thickness estimates derived from FS T1 and ANTS T1 approaches (comparison iii)) and divided by age group. Panel A, B and C report mature, old and oldest old group mean cortical thickness values, respectively. Each point corresponds to a ROI of the DKT atlas as listed in the bottom right panel.

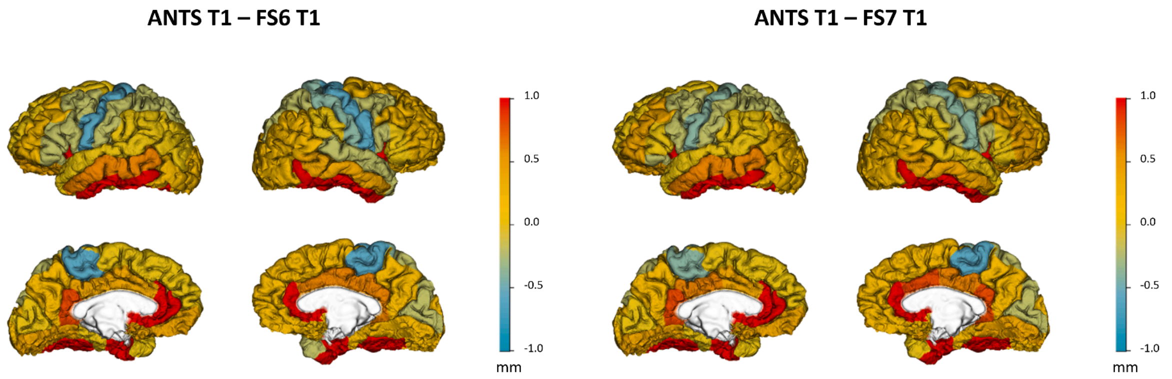


Fig. 5. Mean cortical thickness differences computed by subtracting FS T1 from ANTS T1. Images were generated as described in Schäfer et al. [44].

3.4. *Comparison (iv): T1w- and T2w-based approaches*

Mean cortical thickness estimates in the three age groups obtained using FS6 T1T2 (blue), FS7 T1T2 (orange) and ANTS T1T2 (yellow) are displayed in Fig. 6. A two-way repeated measures ANOVA revealed significant effects due to software ($F(2)=9.0$, $p<0.05$) and age ($F(2)=397.4$, $p<0.05$) in the cortical thickness measurements. Post-hoc tests indicated significant differences in the mature (ANTS T1T2 > FS6 T1T2 > FS7 T1T2, $p<0.05$), old (FS6 T1T2 > FS7 T1T2, $p<0.05$) and oldest old (FS6 T1T2 > ANTS T1T2, $p<0.05$) groups. Moreover, regardless of the software used, cortical thickness exhibited a decrease with age ($p<0.05$). Fig. 7 illustrates the mean cortical thickness differences of the T1w-T2w-based approaches. Both FS6 and FS7 estimates were 1 mm lower than those obtained with ANTS in the rostral anterior cingulate cortex. By focusing on the correlations presented in Table 1, it could be observed that there was poor correlation among the different software estimates

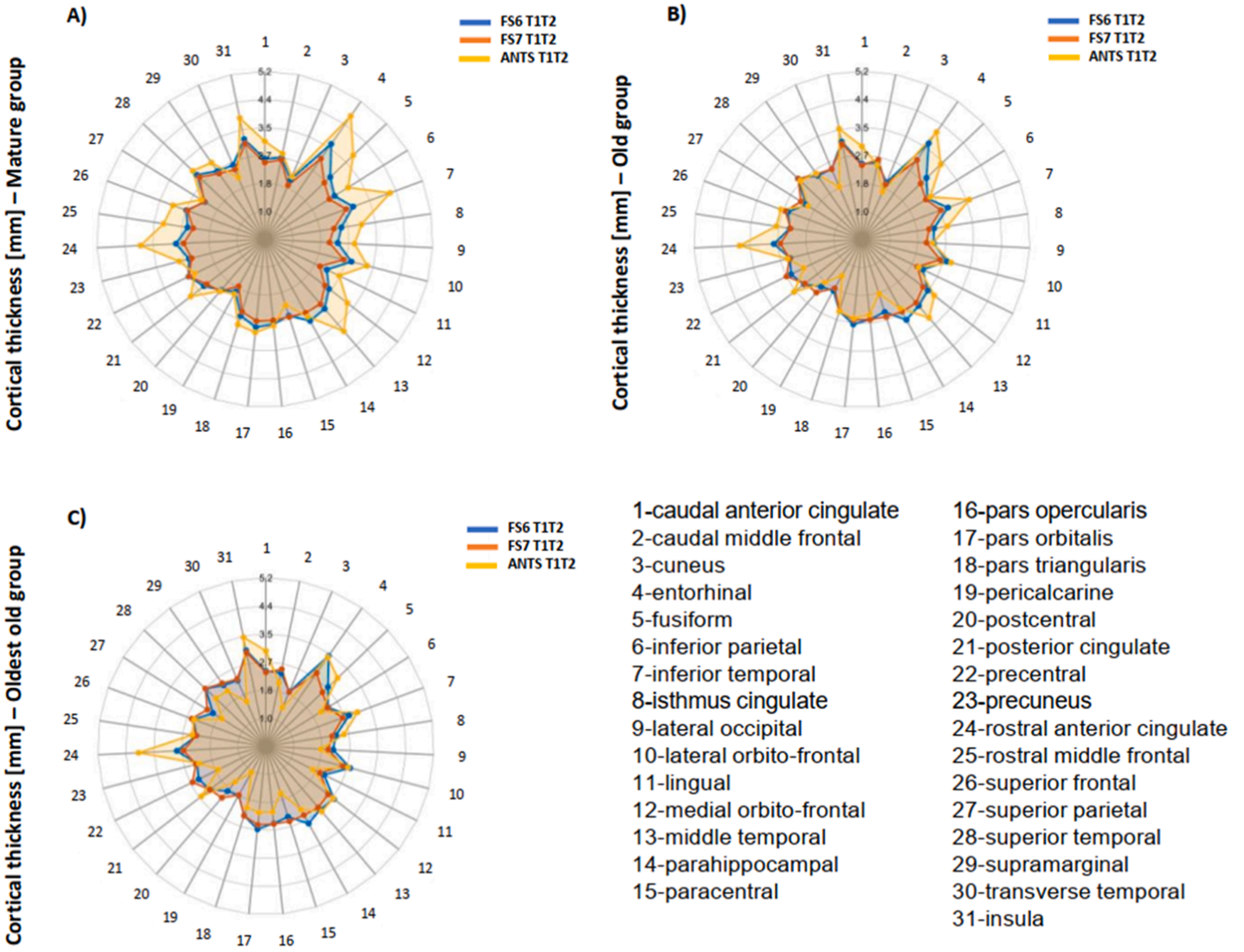


Fig. 6. Mean cortical thickness estimates derived from FS T1T2 and ANTS T1T2 approaches (comparison iv)) and divided by age group. Panel A, B and C report mature, old and oldest old group mean cortical thickness values, respectively. Each point corresponds to a ROI of the DKT atlas in the bottom right panel.

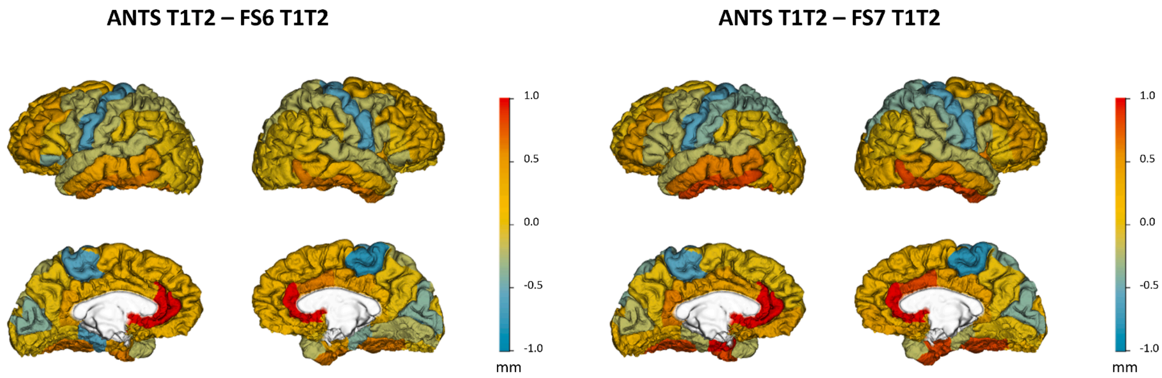


Fig. 7. Mean cortical thickness differences computed by subtracting FS T1T2 from ANTS T1T2. Images were generated as described in Schäfer et al. [44].

($r = 0.59$, $p<0.05$ for FS6 T1T2-ANTS T1T2, and $r = 0.36$, $p<0.05$ and FS7 T1T2-ANTS T1T2). However, when comparing different versions, the correlation value was slightly higher ($r = 0.69$, $p<0.05$). As indicated in Table 2 (rows: 2, 4, 6), FS7 T1T2 had the best predictive performance ($RMSE_{mean}\pm RMSE_{sd}$: FS7 T1T2: 15.11±0.69).

3.5. *ROI-level analysis*

Mean cortical thickness values from the 62 ROIs of the DKT atlas were included in a ROI-level comparison, involving FS7 T1 and ANTS T1 approaches. Within the mature group, the only non-significant differences were observed in the left and right postcentral and transverse temporal regions, as indicated in panel A of Fig. 8 (white cortical regions correspond to non-significant differences). In the old group, non-

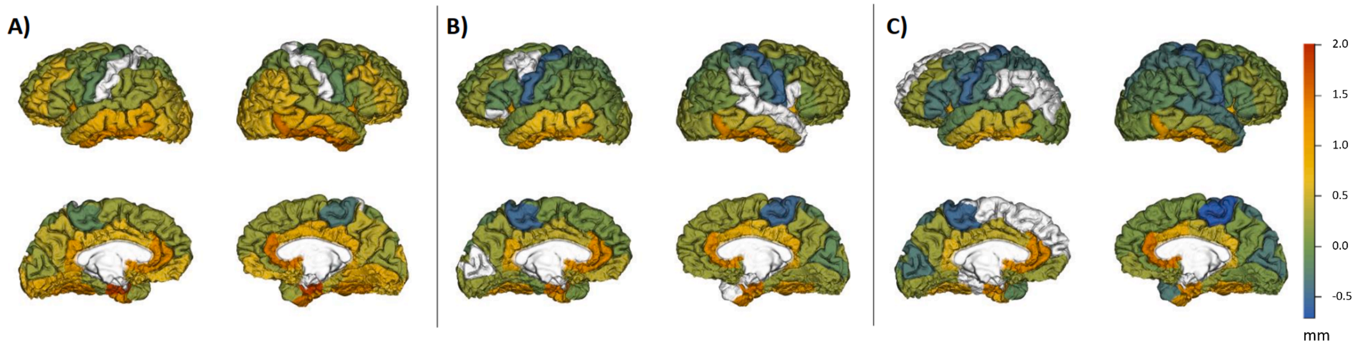


Fig. 8. Mean cortical thickness differences between ANTS T1 and FS7. White regions are associated to non-significant differences. The analysis was performed in the mature (panel A), old (panel B) and oldest old (panel C) groups separately. Images were generated as described in Schäfer et al. [44].

significant differences were observed in the left hemisphere in the caudal middle frontal, cuneus, pars orbitalis and pars triangularis regions, and in the right hemisphere in the pars opercularis, pars orbitalis, superior temporal and supramarginal regions (panel B of Fig. 8). In the oldest old group non-significant differences were found in the left inferior parietal, parahippocampal, superior frontal and supramarginal regions, as displayed in panel C of Fig. 8. The differences were attenuated as age increased, with the maximum differences being observed in the temporal and cingulate regions.

4. Discussion

In this study, we compared two of the most widely used toolboxes for computing cortical features. Additionally, we investigated whether incorporating an additional image alongside the standard T1w image could enhance the results. Our results highlighted that the cortical features obtained using different software or versions were significantly different. This agreed with previous works where significant differences were found in cortical thickness and volume measurements obtained with previous versions of FreeSurfer [23,24,26]. It is worth noting that differences emerged in all age groups, though not in the same way, indicating that the age range of the subjects played a non-negligible role. When utilizing T1w images, the correlation between cortical thickness estimations obtained from FreeSurfer versions was higher compared to when T2w images were included. Conversely, the cortical thickness computed using ANTs with the T1w image exhibited strong correlation with the results obtained by adding the T2w image. Low correlations were observed among different versions of FreeSurfer and ANTs, both utilizing the T1w image alone and incorporating the T2w image. The first result was in line with expectations, supported by various studies [8,27–29]. However, to the best of our knowledge, no previous studies had explored the effects of including T2w images at the whole-brain level. Given the absence of a ground truth for the cortical thickness estimates, we evaluated the accuracy of the results obtained with the different approaches based on their ability to predict age. This approach had previously been adopted in a study by Tustison et al. [8] involving a previous version of FreeSurfer. In our study, the best predictive performances were obtained with FreeSurfer version 7.1.1. Conversely, the weakest performances were obtained using ANTs. The inclusion of the T2w images did not influence these results. In the study by Tustison et al. [8], the most predictive estimates were obtained using ANTs. This discrepancy might arise from variations in the processed images, and due to the different FreeSurfer versions that have been utilized. Additionally, even though the mean RMSE values of our study are marginally elevated, it is essential to note that our analysis involved the complete dataset for age prediction, whereas in their study only a subset of subjects within a specific age range was considered. A ROI-level analysis between FS7 T1 and ANTS T1 was carried out to evaluate regional differences across the three age groups. Notably, significant differences were identified in all groups, albeit in varying regions. As age advanced, the differences became less pronounced, with the most substantial disparities noted in the temporal and cingulate regions. Interestingly, the temporal lobe and its neighbouring regions pose challenges for precise measurements [45], and furthermore, these areas exhibited differences across all age groups with consistent magnitude, reflecting their distinct characteristics [3]. We are aware that more recent FreeSurfer versions are now available (e.g., 7.2, 7.3 and 7.4) but, as reported by FreeSurfer developers, for version 7.1 and above recon-all will end up with the same results (https://surfer.nmr.mgh.harvard.edu/fswiki/ReleaseNotes). No manual correction was applied to the output of the cortical thickness pipelines since it could potentially introduce a bias, and this goes beyond the aims of the present study [45]. It is worth noting that all the results discussed in this study are derived from a cohort of healthy people and could not be translated to pathological subjects without a further study.

5. Conclusion

In the present study we demonstrate the importance of processing data with the same tool, particularly in longitudinal studies. Furthermore, when aligning one's findings with prior research, understanding the tool employed for estimations and the age range of the subjects became very important, given the pivotal role of cortical thickness in disease-related studies. Since the incorporation of the T2w image into the processing pipeline did not enhance FreeSurfer results, and only marginally affected the outcomes obtained using ANTs, its inclusion in the analysis is deemed nonessential.

Data availability

Data of the Human Connectome Project in Aging are available upon request at https://www.humanconnectome.org/study/hcp-lifespan-aging

CRediT authorship contribution statement

Giulia Debiasi: Conceptualization, Software, Formal analysis, Writing – original draft, Visualization. **Ilaria Mazzonetto:** Conceptualization, Software, Formal analysis, Writing – original draft, Visualization. **Alessandra Bertoldo:** Conceptualization, Writing – review & editing, Supervision, Project administration.

Declaration of Competing Interest

None.

Acknowledgments

Research reported in this publication was supported by the National Institute On Aging of the National Institutes of Health under Award Number U01AG052564 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. The HCP-Aging 2.0 Release data used in this report came from doi:10.15154/1520707.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cmpb.2023.107825.

References

[1] B. Dickerson, et al., Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area, Neurobiol. Aging 30 (3) (2009).
[2] Y. Takayanagi, et al., Reduced cortical thickness in schizophrenia and schizotypal disorder, Schizophr. Bull. 46 (2) (2020).
[3] S. Frangou, et al., Cortical thickness across the lifespan: data from 17,075 healthy individuals aged 3-90 years, Hum. Brain Mapp. 43 (1) (2022).
[4] B. Fischl, A. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images, in: Proceedings of the National Academy of Sciences of the United States of America 97, 2000.
[5] A. Dale, B. Fischl, M. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction, Neuroimage 9 (2) (1999).
[6] S. Das, et al., Registration based cortical thickness measurement, Neuroimage 45 (3) (2009).
[7] B. Fischl, FreeSurfer, Neuroimage 62 (2) (2012).
[8] N. Tustison, et al., Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements, Neuroimage 99 (2014).
[9] R. Savjani, et al., Accelerated changes in cortical thickness measurements with age in military service members with traumatic brain injury, J. Neurotrauma 34 (22) (2017).
[10] S. Das, et al., Short-Term memory depends on dissociable medial temporal lobe regions in amnestic mild cognitive impairment, Cereb. Cortex 26 (5) (2016) (New York, N.Y.: 1991).
[11] R. Seiger, et al., Cortical thickness estimations of FreeSurfer and the CAT12 toolbox in patients with Alzheimer's disease and healthy controls, J Neuroimaging 28 (5) (2018).
[12] L. Dimitrova, et al., A neurostructural biomarker of dissociative amnesia: a hippocampal study in dissociative identity disorder, Psychol. Med. 53 (3) (2023).
[13] S.S. Keller, et al., Volume estimation of the thalamus using freesurfer and stereology: consistency between methods, Neuroinformatics 10 (4) (2012) 341–350.
[14] W. Tae, et al., Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder, Neuroradiology 50 (7) (2008).
[15] R. Morey, et al., A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes, Neuroimage 45 (3) (2009).
[16] G. Kuperberg, et al., Regionally localized thinning of the cerebral cortex in schizophrenia, Arch. Gen. Psychiatry 60 (9) (2003).
[17] D. Salat, et al., Thinning of the cerebral cortex in aging, Cereb. Cortex 14 (7) (2004) (New York, N.Y.: 1991).
[18] H. Rosas, et al., Regional and progressive thinning of the cortical ribbon in Huntington's disease, Neurology 58 (5) (2002).
[19] X. Han, et al., Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer, Neuroimage 32 (1) (2006).
[20] B. Dickerson, et al., Detection of cortical thickness correlates of cognitive performance: reliability across MRI scan sessions, scanners, and field strengths, Neuroimage 39 (1) (2008).
[21] J. Jovicich, et al., MRI-derived measurements of human subcortical, ventricular and intracranial brain volumes: reliability effects of scan sessions, acquisition sequences, data analyses, scanner upgrade, scanner vendors and field strengths, Neuroimage 46 (1) (2009).
[22] D. Salat, et al., Regional white matter volume differences in nondemented aging and Alzheimer's disease, Neuroimage 44 (4) (2009).
[23] E. Gronenschild, et al., The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements, PLoS One 7 (6) (2012).
[24] E. Bigler, et al., FreeSurfer 5.3 versus 6.0: are volumes comparable? A chronic effects of neurotrauma consortium study, Brain Imaging Behav 14 (5) (2020).
[25] J. Chepkoech, et al., Effects of change in FreeSurfer version on classification accuracy of patients with Alzheimer's disease and mild cognitive impairment, Hum. Brain Mapp. 37 (5) (2016).
[26] E. Haddad, et al., Multisite test-retest reliability and compatibility of brain metrics derived from FreeSurfer versions 7.1, 6.0, and 5.3, Hum. Brain Mapp. 44 (4) (2023).
[27] M. Clarkson, et al., A comparison of voxel and surface based cortical thickness estimation methods, Neuroimage 57 (3) (2011).
[28] N. Bhagwat, et al., Understanding the impact of preprocessing pipelines on neuroimaging cortical surface analyses, Gigascience 10 (1) (2021).
[29] S. Kharabian Masouleh, et al., Influence of Processing Pipeline on Cortical Thickness Measurement, Cereb. Cortex 30 (9) (2020) (New York, N.Y.: 1991).
[30] P. Sämann, et al., FreeSurfer-based segmentation of hippocampal subfields: a review of methods and applications, with a novel quality control procedure for ENIGMA studies and other collaborative efforts, Hum. Brain Mapp. 43 (1) (2022).
[31] S. Mueller, et al., Systematic comparison of different techniques to measure hippocampal subfield volumes in ADNI2, Neuroimage Clin 17 (2017).
[32] S. Bookheimer, et al., The lifespan human connectome project in aging: an overview, Neuroimage 185 (2019).
[33] M. Harms, et al., Extending the human connectome project across ages: imaging protocols for the lifespan development and aging projects, Neuroimage 183 (2018).
[34] M. Glasser, et al., The minimal preprocessing pipelines for the Human Connectome Project, Neuroimage 80 (2013).
[35] J. Sled, A. Zijdenbos, A. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Trans. Med. Imaging 17 (1) (1998).
[36] N. Tustison, et al., N4ITK: improved N3 bias correction, IEEE Trans. Med. Imaging 29 (6) (2010).
[37] B. Fischl, et al., Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain, Neuron 33 (3) (2002).
[38] A. Klein, J. Tourville, 101 labeled brain images and a consistent human cortical labeling protocol, Front. Neurosci. 6 (2012).
[39] A. Winkler, et al., Measuring and comparing brain cortical surface area and other areal quantities, Neuroimage 61 (4) (2012).
[40] A. Winkler, et al., Joint analysis of cortical area and thickness as a replacement for the analysis of the volume of the cerebral cortex, Cereb. Cortex 28 (2) (2018) (New York, N.Y.: 1991).
[41] B. Avants, et al., Evaluation of open-access, automated brain extraction methods on multi-site multi-disorder data, in: 16th annual meeting for the organization of human brain mapping, 2010.
[42] B. Avants, et al., An open source multivariate framework for n-tissue segmentation with evaluation on public data, Neuroinformatics 9 (4) (2011).
[43] H. Wang, et al., Multi-atlas segmentation with joint label fusion, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013).
[44] Schäfer, T. and C. Ecker, fsbrain: an R package for the visualization of structural neuroimaging data. 2020.
[45] C. McCarthy, et al., A comparison of FreeSurfer-generated data with and without manual intervention, Front. Neurosci. 9 (2015).