



Investigating the Added Value of FreeSurfer's Manual Editing Procedure for the Study of the Reading Network in a Pediatric Population

Caroline Beelen^{1*}, Thanh Vân Phan², Jan Wouters³, Pol Ghesquière^{1†}
and Maaïke Vandermosten^{3†}

OPEN ACCESS

Edited by:

Dieter J. Meyerhoff,
University of California, San
Francisco, United States

Reviewed by:

Joseph O'Neill,
University of California, Los Angeles,
United States
Farshid Sepehrband,
University of Southern California,
United States

*Correspondence:

Caroline Beelen
caroline.beelen@kuleuven.be

[†]These authors share last authorship

Specialty section:

This article was submitted to *Brain Imaging and Stimulation*, a section of the journal *Frontiers in Human Neuroscience*

Received: 07 February 2020

Accepted: 30 March 2020

Published: 24 April 2020

Citation:

Beelen C, Phan TV, Wouters J, Ghesquière P and Vandermosten M (2020) Investigating the Added Value of FreeSurfer's Manual Editing Procedure for the Study of the Reading Network in a Pediatric Population. *Front. Hum. Neurosci.* 14:143. doi: 10.3389/fnhum.2020.00143

¹Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium, ²Icometrix Company, Leuven, Belgium, ³Research Group ExpORL, Department of Neuroscience, KU Leuven, Leuven, Belgium

Insights into brain anatomy are important for the early detection of neurodevelopmental disorders, such as dyslexia. FreeSurfer is one of the most frequently applied automatized software tools to study brain morphology. However, quality control of the outcomes provided by FreeSurfer is often ignored and could lead to wrong statistical inferences. Additional manual editing of the data may be a solution, although not without a cost in time and resources. Past research in adults on comparing the automatized method of FreeSurfer with and without additional manual editing indicated that although editing may lead to significant differences in morphological measures between the methods in some regions, it does not substantially change the sensitivity to detect clinical differences. Given that automated approaches are more likely to fail in pediatric—and inherently more noisy—data, we investigated in the current study whether FreeSurfer can be applied fully automatically or additional manual edits of T1-images are needed in a pediatric sample. Specifically, cortical thickness and surface area measures with and without additional manual edits were compared in six regions of interest (ROIs) of the reading network in 5- to 6-year-old children with and without dyslexia. Results revealed that additional editing leads to statistical differences in the morphological measures, but that these differences are consistent across subjects and that the sensitivity to reveal statistical differences in the morphological measures between children with and without dyslexia is not affected, even though conclusions of marginally significant findings can differ depending on the method used. Thereby, our results indicate that additional manual editing of reading-related regions in FreeSurfer has limited gain for pediatric samples.

Keywords: FreeSurfer, manual editing, automated processing, pediatric T1-weighted images, reading network, developmental neuroimaging

INTRODUCTION

The anatomy of the brain is considered to play a major role in neurodevelopmental disorders, and investigating anatomical features of the brain at pre-diagnostic age could offer us insights into the causal brain mechanisms and early neural markers of these disorders. Consequentially, it allows us to apply early intervention tools, which have demonstrated to be effective (Jones et al., 2007; Makrygianni and Reed, 2010; Peters-Scheffer et al., 2011; Estes et al., 2015; Lovett et al., 2017). In the case of dyslexia, starting intervention at the end of kindergarten has shown to be most efficient for closing the gap in reading performances with typically developing peers (Wanzek and Vaughn, 2007; Ozernov-Palchik and Gaab, 2016). Anatomical features of the brain can be studied non-invasively by using structural MRI (T1) scans, and this can lead to a better understanding of behavioral characteristics of developmental disorders, such as dyslexia (Durston et al., 2001). One of the most frequently applied software packages for structural brain imaging analysis world-wide is FreeSurfer, an open-access software program for the automated surface-based reconstruction of brain images (Dale et al., 1999; Fischl et al., 1999a). In the current study, we will examine whether FreeSurfer can be applied in a fully automated way or whether additional manual edits are needed to improve the automated analysis when studying regions of the reading network in a pediatric population of 5-to-6-year-old children with and without dyslexia.

FreeSurfer has been developed by the Martino Center for Biomedical Imaging with the purpose to be robust, accurate and easy to use (Fischl et al., 2002, 2004). The software tool has an automatic reconstruction pipeline for the processing of anatomical brain images, which involves several processing steps. The first step is skull stripping, motion artifact correction, and B1 bias field correction. The second step is gray-white matter segmentation based on a deformable surface template defined in MNI305 space. As an alternative, a template can be created from one's study sample (i.e., the average subject will form the template) with the advantage of the optimal formation of white and pial surfaces. The third step is region labeling on the cortical surface that is performed by non-linear registration of the cortical surface of the subject with the Desikan-Killiany/Destrieux atlas (Desikan et al., 2006; Destrieux et al., 2010). FreeSurfer adopts a probabilistic approach based on Markov random fields for automated labeling of brain regions by the implementation of a brain surface atlas that is generated by a training set of 40 manually labeled brains, created from 10 young adults (mean age = 21.5, age-range 19–24; six females, four males), 10 middle-aged adults (mean age = 49.8, age-range 41–57; seven females, three males), 10 typically developing elderly (mean age = 74.3, age-range 66–87; eight females, two males) and 10 elderly with Alzheimer's disease (mean age = 78.2, age-range 71–86; five females, five males; Desikan et al., 2006; Makowski et al., 2018). Outcome measures of the morphological analyses are, amongst other, surface area, cortical thickness and volume (Dale et al., 1999; Fischl et al., 1999a,b, 2001; Fischl and Dale, 2000). *The surface area* corresponds to the white surface (gray-white matter boundary), *cortical thickness* corresponds to

the distance between the white surface and the pial surface [gray matter-cerebrospinal fluid (CSF) boundary], and *volume* is the product of surface area and cortical thickness (Fischl and Dale, 2000). It is suggested that surface area and cortical thickness offer independent, but complementary information on brain anatomy, since they have different genetic sources (Panizzon et al., 2009; Winkler et al., 2010) and follow different developmental trajectories (Kapellou et al., 2006).

Over the past few decades, automated methods have increased in popularity for analyzing brain morphology because of increased efficiency (i.e., reduced time and costs) of the analyses (Eickhoff et al., 2015). Although past research indicated that cortical thickness obtained by the automated processing stream of FreeSurfer has good agreement with cortical thickness from histological and manual measurements (Fischl et al., 2002; Rosas et al., 2002; Cardinale et al., 2014), FreeSurfer recommends¹ to always visually check and, if observed necessary, manually adapt images in between the automatic processing stages, thereby optimizing the parcellation and segmentation of the brain images. The automatic reconstruction process can be interrupted for manual adjustments after specific processing stages, including skull stripping, intensity normalization, white matter segmentation and surface extraction (see Figure 1 of the **Supplementary Information**). Manual adjustments include fixing skull stripping errors, intensity normalization errors, topological errors, white matter errors, and pial errors (see the **Supplementary Information**). They might be necessary when an experimenter visually observes that the image is not well parcellated or segmented, as indicated by too little or too much skull of the brain left behind after skull stripping, incorrect placements of the pial surface or white surface, and the presence of small holes on the brain surface (visible on the inflated brain image). However, manually adapting images has disadvantages. First and foremost, it is a very time-consuming, labor-intensive task. In their guidelines, FreeSurfer suggests to take around 30 min per image to fix errors, yet in reality, such time window seems far too short as per image two-hundred slices, preferably in four different views (i.e., coronal, sagittal, axial and 3D-vision), need to be visually checked and possibly edited using five different fixing steps (for details on the editing process see the **Supplementary Information**). Also, manual editing requires a certain level of expertise, which can only be gained through practice and trial-and-error (Canna et al., 2018). The experimenter needs to get experience with how many edits at which locations are needed to obtain a desired change. To visualize the outcomes of the editing process, the image first has to undergo the automatic reconstruction process again. Another disadvantage of manual editing is that compared to the automated reconstruction of brain images, manual editing is more prone to inter-subject variation and rater drift (i.e., rater variation over time; Fischl et al., 2002; White et al., 2018). Especially in large databases, in which data are segmented over a prolonged period, intra-rater reliability will be difficult to maintain as rater drift becomes more significant over time (Spinks et al., 2002; Nugent et al., 2007). Due to the absence

¹<http://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/TroubleshootingData>

of a standardized detailed protocol on performing manual adjustments in FreeSurfer, inter-subject variation and rater drift might be considerably larger after manually editing the data.

FreeSurfer's fully automated approach and the automated approach with additional manual edits have been compared in former adult studies, more specifically in healthy adults (Canna et al., 2018; Waters et al., 2019), adults with a history of severe head injuries (Guenette et al., 2018) and in adults with 22q11.2 deletion syndrome (McCarthy et al., 2015). Each of these studies showed using paired samples *t*-tests that both methods did not significantly differ in their regions of interest (ROIs; McCarthy et al., 2015) or only in a minority of their ROIs (Canna et al., 2018; Waters et al., 2019). Furthermore, the intra-class correlations measuring consistency in differences between the methods across subjects were high for the ROIs included in these studies (McCarthy et al., 2015; Guenette et al., 2018; Waters et al., 2019), except for a few subcortical ROIs (Guenette et al., 2018). Finally, these studies showed that the automated method with additional manual edits does not have an increased sensitivity to detect differences when comparing a clinical group (McCarthy et al., 2015) or investigating individual brain-behavior relationships (Waters et al., 2019) between the methods. Hence, according to these studies, manual editing does not lead to differences in absolute values between the methods in most ROIs. Additionally, differences between the methods—in situations in which they occurred—are consistent, and manual editing does not lead to an increased sensitivity to detect individual or clinical group differences. However, since these studies were conducted in adults, their outcomes might not apply to pediatric populations. Manual adaptations are needed to a considerably larger extent in pediatric data, since non-linear and region-specific brain development leads to significant differences between adult and child brains (Phan et al., 2018b). In pediatric populations, segmentation and registration of MRI data is more prone to errors because of the use of an adult template (Muzik et al., 2000; Yoon et al., 2009; Phan et al., 2018a). Also, MRI images of pediatric populations are generally noisier and more prone to artifacts due to in-scanner motion that is higher in children (Brown et al., 2010; van Dijk et al., 2012). Head motion, in particular, will lead to severe ringing, blurring, and ghosting artifacts, hindering the determination of tissue boundaries, thereby risking the production of invalid outcome measures (Backhausen et al., 2016; Phan et al., 2018a,b). Even subtle motion, not easily detected by visual inspection, will lead to systematic biases in the automatic measurements of structural brain properties (Blumenthal et al., 2002; Reuter et al., 2015; Alexander-Bloch et al., 2016). Errors are comparable to yearly atrophy rates in neurodegenerative diseases (Barkhof et al., 2009; Rosas et al., 2011) or yearly growth rates of normal developing brain tissues (Hedman et al., 2012). Influences of motion on outcomes generated by the automated image processing pipeline of FreeSurfer (Blumenthal et al., 2002; Reuter et al., 2015) are significant; in adults, a small increase in motion led to 1.4%–2.0% gray matter volume reduction (Reuter et al., 2015), and in children, a small amount of motion led to 4% gray matter volume reduction, a moderate amount to 7% gray matter volume reduction and a large amount to 27% gray

matter volume reduction (Blumenthal et al., 2002). Despite these difficulties faced with implementing FreeSurfer's automated pipeline in pediatric samples, over the past few years there has been an increased use of FreeSurfer in pediatric samples, resulting from an increased interest in early-onset identification of neurodevelopmental disorders (Merkley et al., 2008; Wolosin et al., 2009; Fallucca et al., 2011; Widjaja et al., 2011; Overvliet et al., 2013; Wozniak et al., 2013; Mayer et al., 2015; Mahajan et al., 2016; Yang et al., 2016; Gold et al., 2017). Only a few of these studies mentioned whether they had performed manual adaptations (Merkley et al., 2008; Fallucca et al., 2011; Gold et al., 2017), and only one mentioned the actual editing process by briefly informing which type of edits at which locations had been made (Merkley et al., 2008). Currently, little is known about the influence of additional manual editing pediatric samples in FreeSurfer on morphological outcome measures.

In the current study, we will assess the added value of FreeSurfer's manual editing tool on outcome values of surface area and cortical thickness that are generated by FreeSurfer's automated reconstruction pipeline in a pediatric population of 5-to-6-year-old children with and without dyslexia. We take a similar approach as McCarthy et al. (2015) who investigated this in adults. First, we will investigate whether outcome measures obtained with the automated vs. semi-automated (i.e., additional manually edited data) method differ significantly, and whether these differences are consistent across subjects. We restrict our analyses to six pre-defined ROIs of the Desikan-Killiany atlas that belong to the reading network (Richlan et al., 2009; Beelen et al., 2019); the fusiform gyrus, the inferior parietal gyrus, the inferior temporal gyrus, the middle temporal gyrus, the pars opercularis of the inferior frontal gyrus and the superior temporal gyrus, as well as their right homologous counterparts. Then, we will examine whether the implementation of the fully automated approach (with no manual edits) affects the sensitivity to find differences between groups (in our case pre-readers with and without dyslexia diagnosis) by analyzing the difference in effect size between the fully automated method and the automated method with additional edits for the outcome measures surface area and cortical thickness. In our pediatric study, it is expected that the methods will significantly differ from one another, since pediatric data as opposed to adult data require more manual adjustments due to a non-matching template and increased in-scanner head motion. In our former pediatric study (Beelen et al., 2019), we observed for the automated method with manual edits significant group differences in the surface area of the bilateral fusiform gyrus between children with and without dyslexia. However, a comparison with the fully automated approach was not made. Hence, it remains to be seen whether manually editing pediatric images in FreeSurfer will result into an increased sensitivity to detect statistical differences between children with and without dyslexia.

MATERIALS AND METHODS

Participants

Participants were 54 Flemish children, of whom 31 children with (FRD⁺) and 23 children without (FRD⁻) a family risk for

dyslexia, defined as a first-degree relative (parent or sibling) with a clinical diagnosis of dyslexia. This sample is part of a larger longitudinal project (Vanvooren et al., 2014), of which the current sample is identical to the one described in Beelen et al. (2019). Participants underwent cognitive-behavioral tests once a year and EEG and MRI sessions alternately once every 2 years from the last year of kindergarten until the 5th grade of primary school. For the first MRI session, participants were trained with a child-friendly “submarine” protocol (Theys et al., 2014). This protocol was invented to make the participants familiar and at ease with the scanning procedure to reduce in-scanner motion artifacts. Originally, 71 participants underwent the first MRI session, but due to excessive motion in the scanner 17 participants were excluded from the study, resulting in the remaining sample of 54 participants. Images of the excluded participants had severe blurring, ringing or ghosting artifacts due to which they were unusable for analysis purposes according to the Blumenthal criteria (Blumenthal et al., 2002; Phan et al., 2018a). Images of the 54 participants that were included in the study had no, mild or moderate ringing, blurring or ghosting artifacts.

The cognitive-behavioral tests that were obtained from the participants from grade 2 onwards each year included a standardized word reading (Brus and Voeten, 1973) and pseudo-word reading (van den Bos et al., 1994) test. Reading tests were administered during the first semester of the 2nd to 4th grade, and the second semester of the 5th grade of primary school. Based on these tests, participants were retrospectively classified as typical readers (TR, $n = 38$) or readers with dyslexia (DR, $n = 16$). Children scoring below the 10th percentile on either the word reading or pseudo-word reading test at the three last time points were classified as readers with dyslexia ($n = 15$). If reading had only been assessed at two-time points, the child’s reading score had to be below the 10th percentile at both time points on either the word reading or pseudo-word reading test to be classified as a reader with dyslexia ($n = 1$). In our study, 45% of FRD⁺ children ($n = 14$) and 9% of FRD⁻ children ($n = 2$) fulfilled our dyslexia criteria ($n = 16$). Demographics and behavioral assessment scores of the study sample were mentioned in our former study (Beelen et al., 2019). The study was approved by the local ethical committee of the university hospital (UZ Leuven) and following ethical standards described within the declaration of Helsinki. The study has not been pre-registered. Informed consent had been obtained from the parents.

Image Acquisition

Scanning sessions of participants took place at the university hospital of Leuven (UZ Leuven). Total scanning time was nearly half an hour and T1-weighted brain images were acquired within 6 min. and 22 s. Scans were taken with a Philips 3T-scanner (Best, Netherlands) with 3D Turbo field echo and a 32-channel head coil. Per the scanning session, 182 contiguous coronal slices were collected with the following parameter settings: TR = 9.6 ms; TE = 4.6 ms; flip angle = 8°; FOV = 250 × 250 × 218 mm³; voxel size = 1 × 1 × 1.2 mm³.

Processing of T1-Weighted Images

The T1-weighted images were processed by the automated cross-sectional reconstruction processing stream in FreeSurfer version 5.3 on a Linux Ubuntu software system version 14.02. First, an isotropic brain image was created with all non-brain tissue removed by using a hybrid watershed/surface deformation procedure (Ségonne et al., 2004). Second, motion correction and b1-bias field correction were applied. Furthermore, images underwent segmentation of gray/white matter structures (Fischl et al., 2002, 2004), intensity normalization (Sied et al., 1998), gray/white matter boundary tessellation, automated topological correction (Fischl et al., 2001; Ségonne et al., 2007) and surface deformation following intensity gradients to optimally place gray matter/white matter/CSF borders (Dale and Sereno, 1993; Dale et al., 1999; Fischl and Dale, 2000). Thereafter, the Desikan-Killiany atlas was used to perform the cortical parcellation. The atlas automatically subdivides the human cortex into 34 gyral regions-of-interest based on anatomical markers of curvature and sulcal information on the inflated brain images (Desikan et al., 2006). For our study, we selected six ROIs that belonged to the reading network and their right homologous counterparts from the Desikan-Killiany atlas, which were bilaterally the fusiform gyrus, the inferior parietal gyrus, the inferior temporal gyrus, the middle temporal gyrus, the pars opercularis of the inferior frontal gyrus and the superior temporal gyrus (Beelen et al., 2019). Finally, signal-to-noise ratio (SNR) was calculated using the following formula (see Gedamu et al., 2008, p. 313):

$$\text{SNR} = \frac{\mu_{\text{signal intensity}}}{0.8 \cdot \sigma_{\text{noise intensity}}}$$

“... where signal intensity is the mean of the signal intensity distribution and noise intensity is the standard deviation of the noise intensity distribution of the image. Dividing by a factor of 0.8 is done to compensate for the Rayleigh distribution effect in the background noise.” Mean signal intensity corresponds to the mean intensity of white matter. In our sample, mean SNR = 977.36 dB; SD = 3324.54 dB.

Manual Intervention

The data that underwent the automated cross-sectional processing stream were copied. The copies underwent an additional manual editing procedure (i.e., automated data with additional edits), whereas the original processed images did not (i.e., fully automated data). The reconstructed brain images that underwent additional manual editing were adjusted in FreeSurfer’s supporting toolbox Freeview. Manual edits were performed according to a self-written manual editing protocol (see the **Supplementary Information** for a broad description of the editing process). In short, editing consisted of the following steps: (1) fixing skull stripping errors; (2) fixing intensity normalization errors; (3) fixing topological errors; (4) fixing white matter errors; and (5) fixing pial errors. After the manual adjustments were completed, the images underwent the automated reconstruction process again and were visually checked for remaining errors. If necessary, the same process was repeated.

Statistical Analyses

Statistical analyses were performed in IBM SPSS, version 25.0 (IBM Corp. 2017). Full factorial mixed-effect model analyses were run to compare surface area and cortical thickness of the ROIs between edited and unedited data. As a starting point, the following two models were tested: surface area/cortical thickness = method (2) + region (6) + hemisphere (2) + method*region (12) + method*hemisphere (12) + region*hemisphere (12) + method*region*hemisphere (24), and by using a backward selection procedure, the best models (i.e., with the lowest AIC) were selected for both outcome measures. The selected models were: surface area/cortical thickness = method (2) + region (6) + hemisphere (2) + region*hemisphere (12). All independent factors in the model were fixed, and the subject was the random intercept. Uncorrected results were presented. In the next step, we investigated the consistency of the outcome measures surface area and cortical thickness between the methods by calculating intra-class correlations (ICCs) per ROI and across subjects using a two-way mixed effect model. In a final step, to analyze whether manual edits increase the sensitivity to find statistical group differences between children with (DR) and without (TR) dyslexia, the p -value and effect size (Hedges' g) of the mean difference in cortical thickness and surface area between DR and TR were calculated per ROI for each method, and the effect sizes were compared across ROIs between the methods using paired samples t -tests.

RESULTS

Method comparisons were performed across ROIs and subjects (DR and TR), and ICCs were calculated per ROI and across subjects. Furthermore, sensitivity between the methods was tested by comparing the size of the group differences between DR and TR. For surface area, results revealed that there is a significant difference between the methods ($F_{(1,1242)} = 75.67$; $p < 0.001$). The automated method with additional edits ($\mu = 3,229.1 \text{ mm}^2$; $SE = 45.7 \text{ mm}^2$) has overall a significantly higher mean surface area than the fully automated method ($\mu = 3,077.0 \text{ mm}^2$; $SE = 45.7 \text{ mm}^2$; for an example of the difference in surface area between the methods in an individual subject, see **Figure 1A**). Also, there are no interaction effects between method and ROI or method, hemisphere and ROI, which suggests that results are similar across ROIs. **Table 1A** indicates that for each ROI the mean surface area is larger for the automated method with additional edits than the fully automated method. For cortical thickness, results also revealed that there is a significant difference between the methods ($F_{(1,1242)} = 7.03$; $p = 0.008$). The automated method with additional edits ($\mu = 2,890.9 \text{ }\mu\text{m}$; $SE = 25.8 \text{ }\mu\text{m}$) has overall a significantly lower mean thickness than the fully automated method ($\mu = 2,914.5 \text{ }\mu\text{m}$; $SE = 25.8 \text{ }\mu\text{m}$). Also, there are no interaction effects between method and ROI or method, hemisphere and ROI, which suggests that results are similar across ROIs. **Table 1B** indicates that for each ROI the mean cortical thickness is smaller for the automated method with

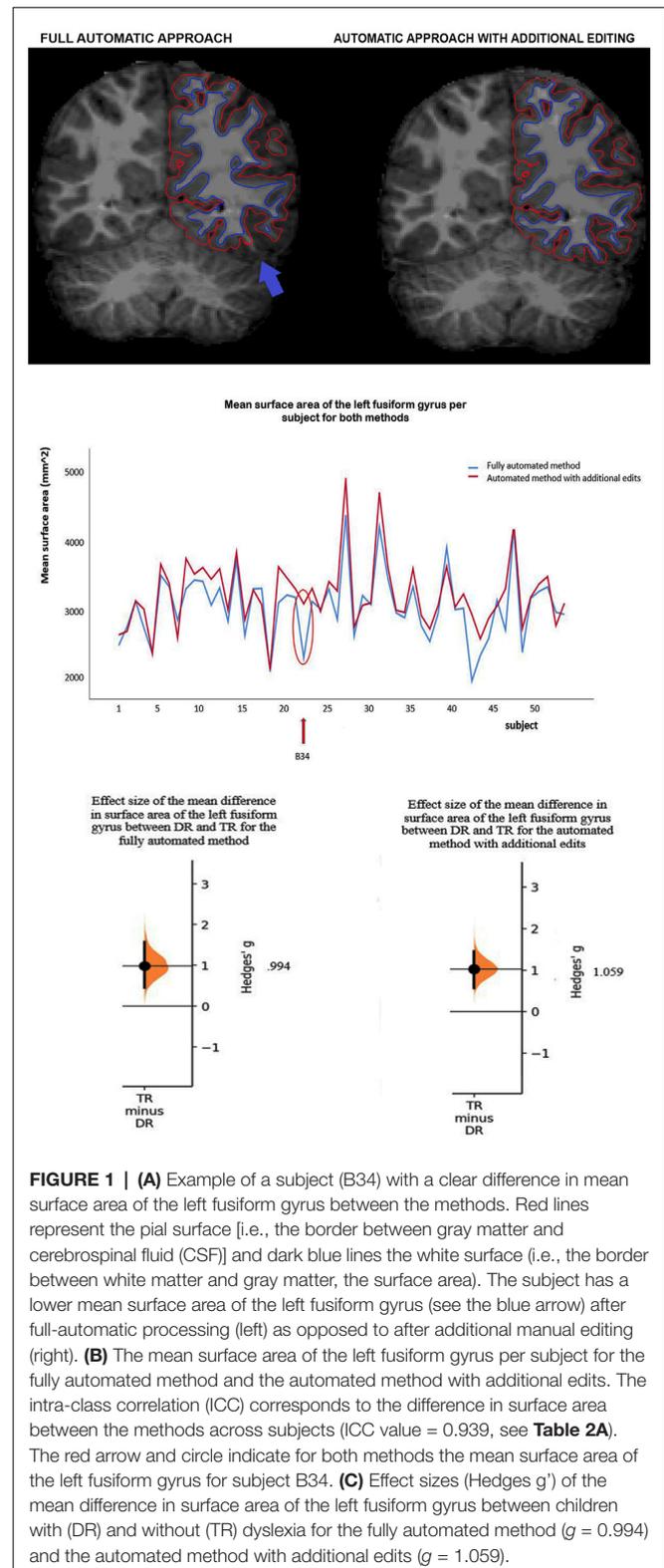


FIGURE 1 | (A) Example of a subject (B34) with a clear difference in mean surface area of the left fusiform gyrus between the methods. Red lines represent the pial surface [i.e., the border between gray matter and cerebrospinal fluid (CSF)] and dark blue lines the white surface (i.e., the border between white matter and gray matter, the surface area). The subject has a lower mean surface area of the left fusiform gyrus (see the blue arrow) after full-automatic processing (left) as opposed to after additional manual editing (right). **(B)** The mean surface area of the left fusiform gyrus per subject for the fully automated method and the automated method with additional edits. The intra-class correlation (ICC) corresponds to the difference in surface area between the methods across subjects (ICC value = 0.939, see **Table 2A**). The red arrow and circle indicate for both methods the mean surface area of the left fusiform gyrus for subject B34. **(C)** Effect sizes (Hedges' g) of the mean difference in surface area of the left fusiform gyrus between children with (DR) and without (TR) dyslexia for the fully automated method ($g = 0.994$) and the automated method with additional edits ($g = 1.059$).

additional edits than the fully automated method. For surface area and cortical thickness, ICC values indicated for each ROI excellent consistency between the methods (ICCs > 0.90; see

TABLE 1 | For both methods per ROI the mean and standard deviation.

ROI	Mean and SD for the fully automated method	Mean and SD for the automated method with additional manual edits
(A)		
Fusiform g. left	<i>M</i> = 3,090.93 <i>SD</i> = 485.47	<i>M</i> = 3,256.26 <i>SD</i> = 498.68
Fusiform g. right	<i>M</i> = 2,953.93 <i>SD</i> = 409.25	<i>M</i> = 3,129.02 <i>SD</i> = 447.76
Inferior parietal g. left	<i>M</i> = 4,276.39 <i>SD</i> = 532.61	<i>M</i> = 4,467.59 <i>SD</i> = 546.43
Inferior parietal g. right	<i>M</i> = 5,145.48 <i>SD</i> = 712.30	<i>M</i> = 5,338.54 <i>SD</i> = 729.91
Inferior temporal g. left	<i>M</i> = 2,800.80 <i>SD</i> = 419.59	<i>M</i> = 3,008.20 <i>SD</i> = 411.68
Inferior temporal g. right	<i>M</i> = 2,780.28 <i>SD</i> = 434.67	<i>M</i> = 2,977.48 <i>SD</i> = 431.43
Middle temporal g. left	<i>M</i> = 2,768.91 <i>SD</i> = 476.36	<i>M</i> = 2,913.91 <i>SD</i> = 470.51
Middle temporal g. right	<i>M</i> = 3,201.00 <i>SD</i> = 439.69	<i>M</i> = 3,351.52 <i>SD</i> = 431.69
Pars opercularis of the inferior frontal g. left	<i>M</i> = 1,532.00 <i>SD</i> = 261.91	<i>M</i> = 1,603.46 <i>SD</i> = 323.45
Pars opercularis of the inferior frontal g. right	<i>M</i> = 1,311.87 <i>SD</i> = 215.36	<i>M</i> = 1,349.24 <i>SD</i> = 211.44
Superior temporal g. left	<i>M</i> = 3,620.76 <i>SD</i> = 448.58	<i>M</i> = 3,781.39 <i>SD</i> = 410.60
Superior temporal g. right	<i>M</i> = 3,441.19 <i>SD</i> = 408.07	<i>M</i> = 3,572.22 <i>SD</i> = 378.57
(B)		
Fusiform g. left	<i>M</i> = 2,980.83 <i>SD</i> = 207.23	<i>M</i> = 2,939.43 <i>SD</i> = 190.28
Fusiform g. right	<i>M</i> = 3,094.72 <i>SD</i> = 202.26	<i>M</i> = 3,053.04 <i>SD</i> = 167.55
Inferior parietal g. left	<i>M</i> = 2,492.09 <i>SD</i> = 264.42	<i>M</i> = 2,466.02 <i>SD</i> = 263.78
Inferior parietal g. right	<i>M</i> = 2,631.57 <i>SD</i> = 256.81	<i>M</i> = 2,610.78 <i>SD</i> = 232.49
Inferior temporal g. left	<i>M</i> = 2,792.69 <i>SD</i> = 253.34	<i>M</i> = 2,774.57 <i>SD</i> = 265.17
Inferior temporal g. right	<i>M</i> = 2,912.50 <i>SD</i> = 269.33	<i>M</i> = 2,895.56 <i>SD</i> = 216.24
Middle temporal g. left	<i>M</i> = 2,975.56 <i>SD</i> = 330.09	<i>M</i> = 2,962.93 <i>SD</i> = 333.42
Middle temporal g. right	<i>M</i> = 3,105.20 <i>SD</i> = 308.95	<i>M</i> = 3,100.63 <i>SD</i> = 242.42
Pars opercularis of the inferior frontal g. left	<i>M</i> = 2,949.63 <i>SD</i> = 223.13	<i>M</i> = 2,912.61 <i>SD</i> = 240.76
Pars opercularis of the inferior frontal g. right	<i>M</i> = 2,860.28 <i>SD</i> = 241.91	<i>M</i> = 2,823.09 <i>SD</i> = 237.54
Superior temporal g. left	<i>M</i> = 3,037.09 <i>SD</i> = 241.35	<i>M</i> = 3,028.31 <i>SD</i> = 230.00
Superior temporal g. right	<i>M</i> = 3,142.09 <i>SD</i> = 214.90	<i>M</i> = 3,123.39 <i>SD</i> = 195.58

(A) For surface area, for each method per ROI and per hemisphere the mean and standard deviation (SD) in mm². (B) For cortical thickness, for each method per ROI and per hemisphere the mean and standard deviation (SD) in μ m.

Tables 2A,B and Figure 1B). Finally, the effect sizes measuring mean differences between DR and TR across ROIs were not significantly different between the methods for surface area ($t_{(11)} = -1.72$; $p = 0.113$) or cortical thickness ($t_{(11)} = 2.04$; $p = 0.066$). Significant group differences between DR and TR

TABLE 2 | Intra-class correlations (ICCs) per ROI and across subjects.

ROI	ICCs for the left hemisphere	ICCs for the right hemisphere
(A)		
Fusiform g.	0.939	0.941
Inferior parietal g.	0.944	0.959
Inferior temporal g.	0.911	0.943
Middle temporal g.	0.970	0.963
Pars opercularis of the inferior frontal g.	0.950	0.957
Superior temporal g.	0.962	0.965
(B)		
Fusiform g.	0.938	0.928
Inferior parietal g.	0.935	0.955
Inferior temporal g.	0.951	0.941
Middle temporal g.	0.966	0.944
Pars opercularis of the inferior frontal g.	0.968	0.948
Superior temporal g.	0.959	0.960

(A) For surface area, for both hemispheres per ROI and across subjects the intra-class correlations (ICCs) between the methods based on a two-way mixed-effects model.

(B) For cortical thickness, for both hemispheres per ROI and across subjects the intra-class correlations (ICCs) between the methods based on a two-way mixed-effects model.

were observed in the surface area of the bilateral fusiform gyrus for both the automated data with additional edits (see also Beelen et al., 2019) and the fully automated data (see Table 3A). Also, the accompanying effect sizes were large (see Table 3A and Figure 1C). Contrary, border significant results were observed in the surface area of the right inferior temporal gyrus for the automated method with additional edits as opposed to the fully automated method (Table 3A), and in the cortical thickness of the pars opercularis of the left inferior frontal gyrus for the fully automated method as opposed to the automated method with additional edits (Table 3B). Hence, depending on the choice of method border significant results lead to different conclusions on group differences between DR and TR.

DISCUSSION

In this study, we investigated for the first time in a pediatric population whether additional manual editing of FreeSurfer data generated with the automated reconstruction pipeline is of added value. First, we investigated in 5-to-6-year-old children in six ROIs belonging to the reading network and their right counterparts if there are statistical differences in cortical thickness or surface area when generated fully automatically in FreeSurfer vs. automatically with additional manual editing. Additionally, for both morphological measures, we checked if differences between the methods are consistent across subjects. Finally, we investigated whether manual editing leads to an increased sensitivity to detect statistical differences in surface area or cortical thickness between 5-to-6-year-old children with and without dyslexia. Results revealed that the methods differ significantly from each other. Specifically, the automated method with additional manual edits reported larger and thinner ROIs than the fully automated method. Additionally, the intra-class correlations between the methods were high for all ROIs, revealing that although the methods differ significantly, the

TABLE 3 | For both methods per ROI the *P*-values and effect sizes (Hedges' *g*) of reading group differences (DR vs. TR.)

ROI	<i>P</i> -values and effect sizes (Hedges' <i>g</i>) of reading group differences (DR vs. TR) for the fully automated method	<i>P</i> -values and effect sizes (Hedges' <i>g</i>) of reading group differences (DR vs. TR) for the automated method with additional manual edits
(A)		
Fusiform g. left	$t_{(52)} = 3.34$; $p = 0.002$; $g = 0.994$	$t_{(52)} = 3.55$; $p = 0.001$; $g = 1.059$
Fusiform g. right	$t_{(52)} = 2.83$; $p = 0.007$; $g = 0.842$	$t_{(52)} = 2.96$; $p = 0.005$; $g = 0.879$
Inferior parietal g. left	$t_{(52)} = -0.43$; $p = 0.671$; $g = 0.129$	$t_{(52)} = -0.49$; $p = 0.628$; $g = 0.145$
Inferior parietal g. right	$t_{(52)} = 0.01$; $p = 0.993$; $g = 0.003$	$t_{(52)} = 0.44$; $p = 0.663$; $g = 0.131$
Inferior temporal g. left	$t_{(52)} = 0.98$; $p = 0.333$; $g = 0.291$	$t_{(52)} = 1.48$; $p = 0.144$; $g = 0.442$
Inferior temporal g. right	$t_{(52)} = 1.62$; $p = 0.112$; $g = 0.481$	$t_{(52)} = 2.05$; $p = 0.046$; $g = 0.609$
Middle temporal g. left	$t_{(52)} = 1.15$; $p = 0.254$; $g = 0.343$	$t_{(52)} = 1.22$; $p = 0.229$; $g = 0.363$
Middle temporal g. right	$t_{(52)} = 1.12$; $p = 0.268$; $g = 0.333$	$t_{(52)} = 1.33$; $p = 0.190$; $g = 0.397$
Pars opercularis of the inferior frontal g. left	$t_{(52)} = 1.00$; $p = 0.333$; $g = 0.293$	$t_{(52)} = 0.62$; $p = 0.536$; $g = 0.185$
Pars opercularis of the inferior frontal g. right	$t_{(52)} = 0.69$; $p = 0.496$; $g = 0.204$	$t_{(52)} = 0.76$; $p = 0.451$; $g = 0.226$
Superior temporal g. left	$t_{(47)} = 0.81$; $p = 0.424$; $g = 0.195$	$t_{(52)} = 0.63$; $p = 0.533$; $g = 0.187$
Superior temporal g. right	$t_{(52)} = 1.28$; $p = 0.207$; $g = 0.379$	$t_{(52)} = 1.08$; $p = 0.286$; $g = 0.320$
(B)		
Fusiform g. left	$t_{(52)} = 0.44$; $p = 0.665$; $g = 0.129$	$t_{(52)} = 0.52$; $p = 0.606$; $g = 0.151$
Fusiform g. right	$t_{(52)} = 0.48$; $p = 0.636$; $g = 0.142$	$t_{(52)} = -0.52$; $p = 0.606$; $g = 0.154$
Inferior parietal g. left	$t_{(52)} = 0.26$; $p = 0.800$; $g = 0.075$	$t_{(52)} = 0.01$; $p = 0.989$; $g = 0.004$
Inferior parietal g. right	$t_{(52)} = 1.19$; $p = 0.238$; $g = 0.356$	$t_{(52)} = 0.75$; $p = 0.458$; $g = 0.223$
Inferior temporal g. left	$t_{(52)} = 0.61$; $p = 0.545$; $g = 0.180$	$t_{(52)} = 0.76$; $p = 0.452$; $g = 0.225$
Inferior temporal g. right	$t_{(52)} = 1.01$;	$t_{(52)} = 0.69$;

(Continued)

TABLE 3 | Continued

ROI	<i>P</i> -values and effect sizes (Hedges' <i>g</i>) of reading group differences (DR vs. TR) for the fully automated method	<i>P</i> -values and effect sizes (Hedges' <i>g</i>) of reading group differences (DR vs. TR) for the automated method with additional manual edits
	$p = 0.320$; $g = 0.297$	$p = 0.495$; $g = 0.207$
Middle temporal g. left	$t_{(52)} = 0.84$; $p = 0.405$; $g = 0.251$	$t_{(52)} = 0.72$; $p = 0.478$; $g = 0.212$
Middle temporal g. right	$t_{(52)} = 1.26$; $p = 0.214$; $g = 0.374$	$t_{(52)} = 1.44$; $p = 0.157$; $g = 0.429$
Pars opercularis of the inferior frontal g. left	$t_{(52)} = 2.11$; $p = 0.040$; $g = 0.629$	$t_{(52)} = 1.51$; $p = 0.137$; $g = 0.449$
Pars opercularis of the inferior frontal g. right	$t_{(52)} = 2.00$; $p = 0.051$; $g = 0.594$	$t_{(21)} = 0.99$; $p = 0.333$; $g = 0.342$
Superior temporal g. left	$t_{(52)} = 1.27$; $p = 0.210$; $g = 0.379$	$t_{(52)} = 0.72$; $p = 0.477$; $g = 0.212$
Superior temporal g. right	$t_{(52)} = 1.72$; $p = 0.091$; $g = 0.512$	$t_{(52)} = 1.92$; $p = 0.061$; $g = 0.571$

(A) For surface area, for each method per ROI and per hemisphere the *p*-value (significance at level $p < 0.05$) and the effect size (Hedges' *g*) of reading group differences (DR vs. TR). (B) For cortical thickness, for each method per ROI and per hemisphere the *p*-value (significance at level $p < 0.05$) and the effect size (Hedges' *g*) of reading group differences (DR vs. TR).

difference is consistent, and therefore leads to similar statistical inferences regarding outcome measures. Finally, effect sizes of differences in surface area or cortical thickness between children with and without dyslexia did not differ between the methods, indicating that manual editing does not lead to an increased sensitivity to detect dyslexia-related morphological brain differences in a pediatric sample, although conclusions of marginally significant findings can differ depending on the chosen method.

In line with our expectation, we observed that the fully automated method and the automated method with manual edits significantly differ in surface area and cortical thickness across ROIs. Compared to the automated method with manual edits, the fully automated method has a lower surface area and a higher cortical thickness across ROIs. Adult studies of McCarthy et al. (2015), Canna et al. (2018) and Waters et al. (2019) observed in only 0%–26.5% of the ROIs in their studies differences in brain morphometry between the methods. Also, they observed no differences in brain morphometry between the methods in ROIs of our study. Hence, it seems that pediatric data needs more manual adjustments than adult data. However, there were few methodological differences between these studies and our study. McCarthy et al. (2015) applied Bonferroni correction to 34 ROIs in their study, whereas in our study there were fewer ROIs to correct for. Additionally, whereas our study focused on

fixing all possible type of edits, the study of Waters et al. (2019) focused on fixing skull stripping errors only, and the study of Canna et al. (2018) on fixing intensity normalization errors only. If we would perform a similar amount of Bonferroni corrections as McCarthy et al. (2015), i.e., correct for 34 ROIs, we would also fail to observe differences in outcome measures between the methods. Yet, suggestively their correction has been too strong. As a consequence, it cannot be confirmed that pediatric data needs more manual adaptations than adult data. Future studies should implement similar paradigms and focus more on replicating findings.

Another observation in our study was the high intra-class correlations between the methods for all ROIs, which indicate that even though the methods may differ for particular ROIs, the differences are highly consistent across subjects, and the observed relation for cortical thickness and surface area is maintained across subjects. The high intra-class correlations suggest that the choice of method is less relevant, since they indicate that both methods lead to similar statistical inferences on group estimates. Nevertheless, it remains important to stick to a single method within a study, as for certain ROIs the methods may statistically differ from each other (see also Guenette et al., 2018). In line with our results, studies of McCarthy et al. (2015) and Waters et al. (2019) observed for all brain regions high intra-class correlations between the methods. On the contrary, Guenette et al. (2018) observed high intra-class correlations for few subcortical ROIs, and, for instance, not for the amygdala and hippocampus. Past research, however, demonstrated that areas such as the amygdala and hippocampus are troublesome areas to segment for FreeSurfer and that their volumes are often overestimated (Tae et al., 2008; Nugent et al., 2013; Schoemaker et al., 2016; Schmidt et al., 2018). Hence, for these subcortical regions, extra caution is warranted.

Finally, our study revealed no significant differences between the methods in effect sizes comparing children with and without dyslexia, indicating no increased sensitivity to detect clinical group differences. Note, however, that marginally significant results can switch from significant to non-significant depending on the method used (see the right inferior temporal gyrus in **Table 2A** and the pars opercularis of the left inferior frontal gyrus in **Table 2B**), whereas stronger group differences (as observed in the bilateral area of the fusiform gyrus, **Table 2A**), remain significant regardless of whether manual edits are performed or not. In sum, although it might seem worthwhile to edit pediatric data, as shown by outcome differences between the methods in the segmentation process and in the output matrices that FreeSurfer generates, results indicate that the additional time and costs required to manually adjust images do not result into an increased sensitivity to detect morphological differences between the reading groups. On the contrary, one should consider whether the benefits of performing manual adaptations to optimize the data outweigh the costs of the high amount of time required for editing and the accompanying costs involved. Especially in large data sets, due to the excessive time needed for editing, benefits might not outweigh the costs, even though some quality control of the data (e.g., dealing with outliers, severe artifacts and motion) would be necessary. Hence, the

results of our study and previous adult studies together indicate that manually adjusting data in FreeSurfer has a limited impact on statistical findings regarding clinical group estimates and individual neurocognitive measures. Possibly, there could be gains for pediatric images with severe motion, which had been excluded from our study. Furthermore, it should be mentioned that under certain circumstances one should always consider the adaptation of the brain image. For instance, in cases involving individual brain modeling or in cases where individual data is needed for accurate predictions of personal diagnosis or treatment outcomes. In such cases, one should aim for the most accurate result, which usually corresponds to some adaptation of the image.

Currently, we have no ground truth on whether automated brain imaging methods or human experts provide the best outcomes (i.e., outcomes that are closest to the true values) regarding brain measures such as surface area or cortical thickness. For a long time, manual editing has been considered to be the golden standard, but is slowly overtaken by automated methods that get more and more accurate while constantly being adapted and improved. In the future, automatic editing procedures or machine learning techniques may become available that can assist or even replace the manual editing procedure, significantly reducing the time needed for adjustments and improving the segmentation of the brain regions. For instance, Canna et al. (2018) developed in their study an automated control point search (ACPS) that improved surface reconstructions to a similar extent as manually editing (i.e., there was a high reproducibility between the method with manual and automated edits across different data sets). Likewise, Sta Cruz et al. (2020) developed in their study a machine learning random forest imputation technique to replace missing or incorrect average values across regions by imputed values that were computed based on available multivariate information. Sta Cruz et al. (2020) tested their newly developed technique on morphological child and adult data that had been automatically processed by FreeSurfer and additionally underwent manual editing. The edited data was compared to the data acquired with the machine learning random forest imputation technique. The newly developed technique proved to be equally effective as manual editing, and would especially be interesting for studies consisting of large datasets, containing more than 250 participants.

A drawback is that we cannot make any claims regarding the necessity of additional editing for images with severe motion ($n = 17$), as they had been excluded from our analyses. Since a significant part of our study sample showed severe motion, future studies on pediatric populations are strongly encouraged to focus on diminishing motion artifacts, for instance by using prospection techniques (Brown et al., 2010; Kuperman et al., 2011; Tisdall et al., 2012). An example would be the implementation of equipment that continuously localizes and follows head position during scanning (White et al., 2010; Tisdall et al., 2012). Furthermore, we recommend implementing a template specifically designed for children in pediatric imaging studies (Alexander et al., 2017; Phan et al., 2018a). Such an atlas will lead to fewer errors in the segmentation process, and

therefore more accurate outcome measures, as well as to an increased sensitivity to find clinical group differences (Phan et al., 2018a). When using an age-specific pediatric atlas as opposed to the adult atlas used by FreeSurfer, it has been shown that only half of the number of subjects is needed to detect significant morphological differences in gray matter volume between 5-to-6-year-old children with and without dyslexia. Recently developed atlases for pediatric populations are the neonatal M-CRIB atlas (Alexander et al., 2017), which is based on the Desikan-Killiany atlas, and could potentially be implemented within the FreeSurfer pipeline. Soon, it is expected that more and more age-specific pediatric atlases will become available.

To conclude, results show that although T1-weighted images fully automatically generated by FreeSurfer, and images with additional manually edits statistically differ in measures of cortical thickness and surface area in a pediatric sample of 5-to-6-year-old children with and without dyslexia, these differences are highly consistent, and additional editing does not result into an increased sensitivity to detect morphological differences between the reading groups.

DATA AVAILABILITY STATEMENT

The dataset for this manuscript is not publicly available, because the conditions of our ethics approval do not permit public archiving of anonymized study data and consent had only been obtained from participants for participation in the study and not to share data with third parties. Requests to access the dataset should be directed to Pol Ghesquière (pol.ghesquiere@kuleuven.be) explaining the purpose of their request. Following the EU general data protection regulation (GDPR), data will be released to requestors upon the following conditions: consent of the representative of the minor and a formal agreement between parties. Please note that the MRI data cannot be shared under any circumstance as MRI data are person-specific and therefore are not anonymous.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethische Commissie Onderzoek UZ/KU Leuven.

REFERENCES

- Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., et al. (2016). Subtle in-scanner motion biases automated measurement of brain anatomy from *in vivo* MRI. *Hum. Brain Mapp.* 37, 2385–2397. doi: 10.1002/hbm.23180
- Alexander, B., Murray, A. L., Loh, W. Y., Matthews, L. G., Adamson, C., Beare, R., et al. (2017). A new neonatal cortical and subcortical brain atlas: the melbourne children's regional infant brain (M-CRIB) atlas. *NeuroImage* 147, 841–851. doi: 10.1016/j.neuroimage.2016.09.068
- Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., and Vetter, N. C. (2016). Quality control of structural MRI images applied using freesurfer—a hands-on workflow to rate motion artifacts. *Front. Neurosci.* 10:558. doi: 10.3389/fnins.2016.00558
- Barkhof, F., Calabresi, P. A., Miller, D. H., and Reingold, S. C. (2009). Imaging outcomes for neuroprotection and repair in multiple

Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CB: conceptualization, methodology, validation, visualization, investigation, software, formal analysis, data curation, writing—original draft, writing—review and editing. TP: methodology, software, visualization, writing—review and editing. JW: methodology, validation, visualization, resources, supervision, project administration, and funding acquisition. MV: conceptualization, methodology, validation, investigation, data curation, writing—review and editing, supervision, project administration, and funding acquisition. PG: conceptualization, methodology, validation, resources, writing—review and editing, supervision, project administration, and funding acquisition.

FUNDING

The study was supported by the EU Horizon 2020 Marie Skłodowska-Curie Innovative Training Network (ITN) in 2014: “Advancing brain research in children's developmental neurocognitive disorders” (Childbrain, #641652), by the Research Council KU Leuven (Onderzoeksraad; 'OT/12/044) and by the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek; G0920.12 and 12T4818N).

ACKNOWLEDGMENTS

Thanks to Catherine Theys, Astrid de Vos and Sophie Vanvooren for participant selection and data collection, to Jolijn Vanderauwera for assisting in quality checks of the images, to Cleo Vandervost for assisting in the manual editing process and to all the involved children and their parents for participation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2020.00143/full#supplementary-material>.

- sclerosis trials. *Nat. Rev. Neurol.* 5, 256–266. doi: 10.1038/nrneurol.2009.41
- Beelen, C., Vanderauwera, J., Wouters, J., Vandermosten, M., and Ghesquière, P. (2019). Atypical gray matter in children with dyslexia before the onset of reading instruction. *Cortex* 121, 399–413. doi: 10.1016/j.cortex.2019.09.010
- Blumenthal, J. D., Zijdenbos, A., Molloy, E., and Giedd, J. N. (2002). Motion artifact in magnetic resonance imaging: implications for automated analysis. *NeuroImage* 16, 89–92. doi: 10.1006/nimg.2002.1076
- Brown, T. T., Kuperman, J. M., Erhart, M., White, N. S., Roddey, J. C., Shankaranarayanan, A., et al. (2010). Prospective motion correction of high-resolution magnetic resonance imaging data in children. *NeuroImage* 53, 139–145. doi: 10.1016/j.neuroimage.2010.06.017
- Brus, B. T., and Voeten, M. J. M. (1973). *Een Minuut Test, Vorm A En B*. Nijmegen, The Netherlands: Berkhout.
- Canna, A., Russo, A. G., Ponticorvo, S., Manara, R., Pepino, A., Sansone, M., et al. (2018). Automated search of control points in surface-based

- morphometry. *NeuroImage* 176, 56–70. doi: 10.1016/j.neuroimage.2018.04.035
- Cardinale, F., Chinnici, G., Bramerio, M., Mai, R., Sartori, I., Cossu, M., et al. (2014). Validation of freesurfer-estimated brain cortical thickness: comparison with histologic measurements. *Neuroinformatics* 12, 535–542. doi: 10.1007/s12021-014-9229-2
- Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. segmentation and surface reconstruction. *NeuroImage* 9, 179–194. doi: 10.1006/nimg.1998.0395
- Dale, A. M., and Sereno, M. I. (1993). Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. *J. Cogn. Neurosci.* 5, 162–176. doi: 10.1162/jocn.1993.5.2.162
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010
- Durston, S., Hulshoff Pol, H. E., Casey, B. J., Giedd, J. N., Buitelaar, J. K., and van Engeland, H. (2001). Anatomical MRI of the developing human brain: what have we learned? *J. Am. Acad. Child Adolesc. Psychiatry* 40, 1012–1020. doi: 10.1097/00004583-200109000-00009
- Eickhoff, S. B., Thirion, B., Varoquaux, G., and Bzdok, D. (2015). Connectivity-based parcellation: critique and implications. *Hum. Brain Mapp.* 36, 4771–4792. doi: 10.1002/hbm.22933
- Estes, A., Munson, J., Rogers, S. J., Greenon, J., Winter, J., and Dawson, G. (2015). Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. *J. Am. Acad. Child Adolesc. Psychiatry* 54, 580–587. doi: 10.1016/j.jaac.2015.04.005
- Fallucca, E., MacMaster, F. P., Haddad, J., Easter, P., Dick, R., May, G., et al. (2011). Distinguishing between major depressive disorder and obsessive-compulsive disorder in children by measuring regional cortical thickness. *Arch. Gen. Psychiatry* 68, 527–533. doi: 10.1001/archgenpsychiatry.2011.36
- Fischl, B., and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U S A* 97, 11050–11055. doi: 10.1073/pnas.200033797
- Fischl, B., Liu, A., and Dale, A. M. (2001). Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* 20, 70–80. doi: 10.1109/42.906426
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355. doi: 10.1016/s0896-6273(02)00569-x
- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., et al. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage* 23, 69–84. doi: 10.1016/j.neuroimage.2004.07.016
- Fischl, B., Sereno, M. I., and Dale, A. M. (1999a). Cortical surface-based analysis. II: inflation, flattening and a surface-based coordinate system. *NeuroImage* 9, 195–207. doi: 10.1006/nimg.1998.0396
- Fischl, B., Sereno, M. I., Tootell, R. B. H., and Dale, A. M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284. doi: 10.1002/(sici)1097-0193(1999)8:4<272::aid-hbm10>3.0.co;2-4
- Gedamu, E. L., Collins, D. L., and Arnold, D. L. (2008). Automated quality control of brain MR images. *J. Magn. Reson. Imaging* 28, 308–319. doi: 10.1002/jmri.21434
- Gold, A. L., Steuber, E. R., White, L. K., Pacheco, J., Sachs, J. F., Pagliaccio, D., et al. (2017). Cortical thickness and subcortical gray matter volume in pediatric anxiety disorders. *Neuropsychopharmacology* 42, 2423–2433. doi: 10.1038/npp.2017.83
- Guenette, J. P., Stern, R. A., Tripodis, Y., Chua, A. S., Schultz, V., Sydnor, V. J., et al. (2018). Automated versus manual segmentation of brain region volumes in former football players. *Neuroimage Clin.* 18, 888–896. doi: 10.1016/j.nicl.2018.03.026
- Hedman, A. M., van Haren, N. E. M., Schnack, H. G., Kahn, R. S., and Hulshoff Pol, H. E. (2012). Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. *Hum. Brain Mapp.* 33, 1987–2002. doi: 10.1002/hbm.21334
- Jones, K., Daley, D., Hutchings, J., Bywater, T., and Eames, C. (2007). Efficacy of the incredible years basic parent training programme as an early intervention for children with conduct problems and ADHD. *Child Care Health Dev.* 33, 749–756. doi: 10.1111/j.1365-2214.2007.00747.x
- Kapellou, O., Counsell, S. J., Kennea, N., Dyet, L., Saeed, N., Stark, J., et al. (2006). Abnormal cortical development after premature birth shown by altered allometric scaling of brain growth. *PLoS Med.* 3:e265. doi: 10.1371/journal.pmed.0030265
- Kuperman, J. M., Brown, T. T., Ahmadi, M. E., Erhart, M. J., White, N. S., Roddey, A., et al. (2011). Prospective motion correction improves diagnostic utility of pediatric MRI scans. *Pediatr. Radiol.* 41, 1578–1582. doi: 10.1007/s00247-011-2205-1
- Lovett, M. W., Frijters, J. C., Wolf, M., Steinbach, K. A., Sevcik, R. A., and Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: the impact of grade at intervention and individual differences on intervention outcomes. *J. Educ. Psychol.* 109, 889–914. doi: 10.1037/edu0000181
- Mahajan, R., Dirlikov, B., Crocetti, D., and Mostofsky, S. H. (2016). Motor circuit anatomy in children with autism spectrum disorder with or without attention deficit hyperactivity disorder. *Med. Hypotheses* 9, 67–81. doi: 10.1002/aur.1497
- Makowski, C., Béland, S., Kostopoulos, P., Bhagwat, N., Devenyi, G. A., Malla, A. K., et al. (2018). Evaluating accuracy of striatal, pallidal, and thalamic segmentation methods: comparing automated approaches to manual delineation. *NeuroImage* 170, 182–198. doi: 10.1016/j.neuroimage.2017.02.069
- Makrygianni, M. K., and Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autistic spectrum disorders. *Res. Autism Spectr. Disord.* 4, 577–593. doi: 10.1016/j.rasd.2010.01.014
- Mayer, A. R., Hanlon, F. M., and Ling, J. M. (2015). Gray matter abnormalities in pediatric mild traumatic brain injury. *J. Neurotrauma* 32, 723–730. doi: 10.1089/neu.2014.3534
- McCarthy, C. S., Ramprasad, A., Thompson, C., Botti, J. A., Coman, I. L., and Kates, W. R. (2015). A comparison of freesurfer-generated data with and without manual intervention. *Front. Neurosci.* 9:379. doi: 10.3389/fnins.2015.00379
- Merkley, T. L., Bigler, E. D., Wilde, E. A., McCauley, S. R., Hunter, J. V., and Levin, H. S. (2008). Short communication: diffuse changes in cortical thickness in pediatric moderate-to-severe traumatic brain injury. *J. Neurotrauma* 25, 1343–1345. doi: 10.1089/neu.2008.0615
- Muzik, O., Chugani, D. C., Juhász, C., Shen, C., and Chugani, H. T. (2000). Statistical parametric mapping: assessment of application in children. *NeuroImage* 12, 538–549. doi: 10.1006/nimg.2000.0651
- Nugent, T. F. III, Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., Lenane, M., et al. (2007). Dynamic mapping of hippocampal development in childhood onset schizophrenia. *Schizophr. Res.* 90, 62–70. doi: 10.1016/j.schres.2006.10.014
- Nugent, A. C., Luckenbaugh, D. A., Wood, S. E., Bogers, W., Zarate, C. A., and Drevets, W. C. (2013). Automated subcortical segmentation using FIRST: test-retest reliability, interscanner reliability, and comparison to manual segmentation. *Hum. Brain Mapp.* 34, 2313–2329. doi: 10.1002/hbm.22068
- Overvliet, G. M., Besseling, R. M. H., Jansen, J. F. A., van der Kruis, S. J. M., Vles, J. S. H., Hofman, P. A. M., et al. (2013). Early onset of cortical thinning in children with rolandic epilepsy. *Neuroimage Clin.* 2, 434–439. doi: 10.1016/j.nicl.2013.03.008
- Ozernov-Palchik, O., and Gaab, N. (2016). Tackling the ‘Dyslexia Paradox’: reading brain and behavior for early markers of developmental dyslexia. *Wiley Interdiscip. Rev. Cogn. Sci.* 7, 156–176. doi: 10.1002/wcs.1383
- Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., et al. (2009). Distinct genetic influences on cortical surface area and cortical thickness. *Cereb. Cortex* 19, 2728–2735. doi: 10.1093/cercor/bhp026

- Peters-Scheffer, N., Didden, R., Korzilius, H., and Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Res. Autism Spectr. Disord.* 5, 60–69. doi: 10.1016/j.rasd.2010.03.011
- Phan, T. V., Sima, D. M., Beelen, C., Vanderauwera, J., Smeets, D., and Vandermosten, M. (2018a). Evaluation of methods for volumetric analysis of pediatric brain data: the childmetrix pipeline versus adult-based approaches. *Neuroimage Clin.* 19, 734–744. doi: 10.1016/j.nicl.2018.05.030
- Phan, T. V., Smeets, D., Talcott, J. B., and Vandermosten, M. (2018b). Processing of structural neuroimaging data in young children: bridging the gap between current practice and state-of-the-art methods. *Dev. Cogn. Neurosci.* 33, 206–223. doi: 10.1016/j.dcn.2017.08.009
- Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., and 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
- Richlan, F., Kronbichler, M., and Wimmer, H. (2009). Functional abnormalities in the dyslexic brain: a quantitative meta-analysis of neuroimaging studies. *Hum. Brain Mapp.* 30, 3299–3308. doi: 10.1002/hbm.20752
- Rosas, H. D., Liu, A. K., Hersch, S., Glessner, M., Ferrante, R. J., Salat, D. H., et al. (2002). Regional and progressive thinning of the cortical ribbon in Huntington's disease. *Neurology* 58, 695–701. doi: 10.1212/wnl.58.5.695
- Rosas, H. D., Reuter, M., Doros, G., Lee, S. Y., Triggs, T., Malarick, K., et al. (2011). A tale of two factors: what determines the rate of progression in Huntington's disease? A longitudinal MRI study. *Mov. Disord.* 26, 1691–1697. doi: 10.1002/mds.23762
- Schmidt, M. F., Storrs, J. M., Freeman, K. B., Jack, C. R. Jr., Turner, S. T., Griswold, M. E., et al. (2018). A comparison of manual tracing and freesurfer for estimating hippocampal volume over the adult lifespan. *Hum. Brain Mapp.* 39, 2500–2513. doi: 10.1002/hbm.24017
- Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., et al. (2016). Hippocampus and amygdala volumes from magnetic resonance images in children: assessing accuracy of FreeSurfer and FSL against manual segmentation. *NeuroImage* 129, 1–14. doi: 10.1016/j.neuroimage.2016.01.038
- Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., et al. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 22, 1060–1075. doi: 10.1016/j.neuroimage.2004.03.032
- Ségonne, F., Pacheco, J., and Fischl, B. (2007). Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans. Med. Imaging* 26, 518–529. doi: 10.1109/tmi.2006.887364
- Sied, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97. doi: 10.1109/42.668698
- Spinks, R., Magnotta, V. A., Andreasen, N. C., Albright, K. C., Ziebell, S., Nopoulos, P., et al. (2002). Manual and automated measurement of the whole thalamus and mediodorsal nucleus using magnetic resonance imaging. *NeuroImage* 17, 631–642. doi: 10.1006/nimg.2002.1185
- Sta Cruz, S., Dinov, I. D., Herting, M. M., González-Zacarias, C., Kim, H., Toga, A. W., et al. (2020). Imputation strategy for reliable regional MRI morphological measurements. *Neuroinformatics* 18, 59–70. doi: 10.1007/s12021-019-09426-x
- Tae, W. S., Kim, S. S., Lee, K. U., Nam, E. C., and Kim, K. W. (2008). Validation of hippocampal volumes measured using a manual method and two automated methods (FreeSurfer and IBASPM) in chronic major depressive disorder. *Neuroradiology* 50, 569–581. doi: 10.1007/s00234-008-0383-9
- Theys, C., Wouters, J., and Ghesquière, P. (2014). Diffusion tensor imaging and resting-state functional MRI-scanning in 5- and 6-year-old children: training protocol and motion assessment. *PLoS One* 9:e94019. doi: 10.1371/journal.pone.0094019
- Tisdall, M. D., Hess, T. A., Reuter, M., Meintjes, E. M., Fischl, B., and van der Kouwe, A. J. W. (2012). Volumetric navigators (VNavs) for prospective motion correction and selective reacquisition in neuroanatomical MRI. *Magn. Reson. Med.* 68, 389–399. doi: 10.1002/mrm.23228
- van den Bos, K. P., Spelberg, H. C., Scheepstra, A. J. M., and de Vries, J. R. (1994). *De Klepel Vorm A En B: Verantwoording, Handleiding, Diagnostiek En Behandeling*. Nijmegen: Berkhout.
- van Dijk, K. R. A., Sabuncu, M. R., and Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage* 59, 431–438. doi: 10.1016/j.neuroimage.2011.07.044
- Vanvooren, S., Poelmans, H., Hofmann, M., Ghesquière, P., and Wouters, J. (2014). Hemispheric asymmetry in auditory processing of speech envelope modulations in prereading children. *J. Neurosci.* 34, 1523–1529. doi: 10.1523/JNEUROSCI.3209-13.2014
- Wanzek, J., and Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychol. Rev.* 36, 541–561. doi: 10.1080/02796015.2007.12087917
- Waters, A. B., Mace, R. A., Sawyer, K. S., and Gansler, D. A. (2019). Identifying errors in freesurfer automated skull stripping and the incremental utility of manual intervention. *Brain Imaging Behav.* 13, 1281–1291. doi: 10.1007/s11682-018-9951-8
- White, T., Jansen, P. R., Muetzel, R. L., Sudre, G., El Marroun, H., Tiemeier, H., et al. (2018). Automated quality assessment of structural magnetic resonance images in children: comparison with visual inspection and surface-based reconstruction. *Hum. Brain Mapp.* 39, 1218–1231. doi: 10.1002/hbm.23911
- White, N., Roddey, C., Shankaranarayanan, A., Han, E., Rettmann, D., Santos, J., et al. (2010). PROMO: real-time prospective motion correction in MRI using image-based tracking. *Magn. Reson. Med.* 63, 91–105. doi: 10.1002/mrm.22176
- Widjaja, E., Mahmoodabadi, S. Z., Snead, O. C. III, Almedhar, A., and Smith, M. L. (2011). Widespread cortical thinning in children with frontal lobe epilepsy. *Epilepsia* 52, 1685–1691. doi: 10.1111/j.1528-1167.2011.03085.x
- Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., et al. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage* 53, 1135–1146. doi: 10.1016/j.neuroimage.2009.12.028
- Wolosin, S. M., Richardson, M. E., Hennessey, J. G., Denckla, M. B., and Mostofsky, S. H. (2009). Abnormal cerebral cortex structure in children with ADHD. *Hum. Brain Mapp.* 30, 175–184. doi: 10.1002/hbm.20496
- Wozniak, J. R., Mueller, B. A., Bell, C. J., Muetzel, R. L., Hoecker, H. L., Boys, C. J., et al. (2013). Global functional connectivity abnormalities in children with fetal alcohol spectrum disorders. *Alcohol. Clin. Exp. Res.* 37, 748–756. doi: 10.1111/acer.12024
- Yang, D. Y. J., Beam, D., Pelphrey, K. A., Abdullahi, S., and Jou, R. J. (2016). Cortical morphological markers in children with autism: a structural magnetic resonance imaging study of thickness, area, volume, and gyrification. *Mol. Autism* 7:11. doi: 10.1186/s13229-016-0076-x
- Yoon, U., Fonov, V. S., Perusse, D., and Evans, A. C. (2009). The effect of template choice on morphometric analysis of pediatric brain data. *NeuroImage* 45, 769–777. doi: 10.1016/j.neuroimage.2008.12.046

Conflict of Interest: Author TP was employed by the company Icometrix. The remaining 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.

Copyright © 2020 Beelen, Phan, Wouters, Ghesquière and Vandermosten. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.