VOLUMETRIC AND VBM MEASURES IN 1.5-5 YEARS OLD CHILDREN OBTAINED USING ADULT SEGMENTATION TOOLS

Master's thesis

Elena Ukharova
Master's Degree Program in Human Neuroscience
Faculty of Medicine
University of Turku

Supervisor: Jetro J. Tuulari, M.D., Ph.D.

June 2020

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

UNIVERSITY OF TURKU
FACULTY OF MEDICINE

Elena Ukharova: VOLUMETRIC AND VBM MEASURES IN 1.5-5 YEARS OLD CHILDREN OBTAINED USING ADULT SEGMENTATION TOOLS

Master's thesis
June 2020

MRI-based volumetric and morphometric studies in a healthy pediatric population give a unique opportunity to investigate brain development, potentially leading to development of structural markers for neurological and psychiatric diseases. However, pediatric data analysis presents significant challenges for established processing tools, which were initially developed for adult population. This study aimed to investigate sexual dimorphism and age-related changes in neural tissues in healthy 1.5-5-years-old children and to critically assess the feasibility of the use of popular software such as CAT-12, FSL SIENAX and FSL VBM to obtain volumetric and VBM measures in this age group. Results showed inter-method inconsistency in estimations of total intracranial (TIV), grey (GM) and white matter (WM) volumes. Nonetheless, TIV and GM measures proved to be highly correlated with each other regardless of the chosen processing tool. As tissue segmentation is an essential part of the VBM analysis, quality of the GM and WM segmentations were assessed using Dice coefficients against manually corrected, curated FreeSurfer segmentations. Regardless of the used method, the quality of the segmentation was higher for the group of children of age 5 compared to 1.5-2-years-old group (toddlers); and for GM compared to WM. The amount of statistically significant voxels for FSL VBM results was noticeably higher than for CAT-12. FSL VBM analysis revealed higher GM volumes in females compared to males in the left auditory cortex, while CAT-12 showed no statistically significant difference. CAT-12 and FSL VBM agreed on increased GM volumes in toddlers compared to 5-year-olds in the frontal lobe, lingual gyri and cerebellum; and in putamina in 5-year-olds compared to toddlers. The results indicate that we need to be cautious when interpreting the neuroimaging findings in younger children as they may significantly vary due to the differences in used preprocessing methods and statistical analysis.

KEYWORDS: Magnetic resonance imaging, voxel-based morphometry, volumetric measures, neurodevelopment, segmentation

Table of Contents

List of abbreviations iii

1. Introduction 1

1.1 Challenges of pediatric MRI processing 1

1.2 Quality of pediatric MRI data 3

1.3 Volumetric measures 4

1.4 VBM 4

1.5 Sexual dimorphism in neurodevelopment 6

2. Aims and hypothesis 6

3. Methods 7

3.1 Participants 7

3.2 MRI Acquisition 7

3.3 Data analysis 7

3.3.1 Intermediary image manipulations 8

3.3.2 FSL VBM 8

3.3.3 FSL SIENAX 9

3.3.4 CAT-12 9

3.3.5 FreeSurfer 9

3.4 Statistical analysis 10

3.4.1 Volumetric measures 10

3.4.2 VBM analysis 10

3.4.3 Dice scores 10

4. Results 11

4.1 Brain extraction 11

4.2 Neural tissue volumes 11

4.3 VBM 15

4.3.1 FSL 15

4.3.2 CAT-12 19

4.4 Dice coefficients 19

5. Discussion 24

5.1 The challenges of image preprocessing 25

5.2 Volumetric measures 25

5.3 Discrepancy in VBM analysis 26

5.4 VBM: age and gender effect on GM 27

5.5 Segmentation quality assessment 29

6. Conclusions 30

References 31

List of abbreviations

AMAP	Adaptive maximum a posterior
CSF	Cerebrospinal fluid
DSC	Dice similarity coefficient
FOV	Field of view
FWHM	Full width at half maximum
GLM	General linear model
GM	Grey matter
HMRF	Hidden Markov random field
MRI	Magnetic resonance imaging
ROI	Region of interest
SD	Standard deviation
TFCE	Threshold-free cluster enhancement
TIV	Total intracranial volume
VBM	Voxel-based morphometry
WM	White matter

1. Introduction

In the last decade, developmental neuroscience has been one of the fast-progressing fields, and the accumulating neuroimaging findings have provided much insight into many facets of typical and atypical brain development. Recent results have shown that many processes, that ultimately lead to diseases, begin in childhood – much earlier than the clinical onset (Knickmeyer, et al., 2014). In the light of these discoveries understanding of the underlying processes of neurodevelopment and factors that may disrupt them during critical periods became a crucial step towards prevention and early medical intervention in neurological and neuropsychiatric disorders (Knickmeyer, et al., 2014).

Neurodevelopmental research owes its rapid progress to the advancements in neuroimaging and magnetic resonance imaging (MRI) in particular. With the development of MRI protocols, high-resolution pediatric imaging became possible *in vivo* (Vogel et al., 2016). However, despite having technological means to acquire high-quality brain images in the pediatric population, there is still a relative paucity of tools to process and analyze them as existing algorithms are developed for the adult population and not automatically suited for children (Phan et al., 2018).

Regarding already published data, there is a relative abundance of infant scans soon after birth. In contrast, however, the age range of 6 months to 4 years is poorly covered (Knickmeyer et al., 2008). Child scans including ages above 4 years are more established. This has led to the lack of information on developmental leaps within the central nervous system before, during and right after toddlerhood. A similar situation can be seen in methodology studies: while the adult processing methods have been adapted and widely used in older children (Ghosh et al., 2010; Schoemaker et al., 2016; Vijayakumar et al., 2018), the feasibility of their use for subjects of preschool age is yet to be thoroughly tested.

This study aims to help to fill in the existing gap in the literature and investigate the developmental changes in children between 18 months to 5 years of age. The second objective of the study is to assess the reliability of existing (pre)processing pipelines.

1.1 Challenges of pediatric MRI processing

First and foremost, it is important to note that pediatric brain cannot be seen as a scaled-down version of an adult brain. Neurodevelopment is a highly non-linear process as different areas follow their own maturation timeline and pace. It has been shown that, for example, age-related *apparent* cortical thinning is significantly more prominent in parietal and occipital lobes

compared to the frontal and temporal cortex (Ball et al., 2012). In the case of subcortical areas, the development trajectories also vary depending on the structure: relative volumes of the thalamus and caudate decrease with age, while, for example, hippocampal volume increases (Sussman et al., 2016).

It is important to keep in mind that we have to be critical while interpreting these findings. Changes in volumes of neural tissues during childhood and adolescence are commonly explained by synaptic pruning and myelination for grey and white matter, respectively. However, the underlying processes are likely to be less straightforward and include different mechanisms. Such decrease in cortical thickness seems to originate not only from the removal of inefficient synapses (a core component of pruning), but may also be explained by the myelination of intracortical axons, which increases voxel intensity in T1-weighted MR images, shifting gray-white matter boundary (Paus et al., 2008).

A major challenge for adult MRI processing software, when used in the pediatric population, occurs at the brain extraction, often the very first phase of image preprocessing. During this procedure, non-brain tissues are differentiated from the brain and excluded from the image before further processing. This step often yields erroneous results for children, one potential explanation being the significant age-related changes in distance between cerebral cortex surface and cranial bone (Beauchamp et al., 2011). Sharp intensity changes between the skull, CSF, and neural tissues are used for the determination of brain boundaries by a number of brain extraction tools such as FSL BET (Smith, 2002) and AFNI 3DSkullStrip (Cox, 1996). Compared to the adult population, the layer of non-brain tissues surrounding the cortex is narrower, which can create difficulties for algorithms in regard to finding the boundary between the brain and the skull (Phan et al., 2018).

Challenges often arise during segmentation procedure as well – in pediatric images neural tissues can often be mislabeled. Segmentation algorithms commonly utilize intensity probability distribution at some part of the estimation to differentiate between brain tissues (Despotović et al., 2015). However, tissue intensity can vary depending on the area: for example, grey matter in central structures has higher intensity compared to the cortex (Murgasova et al., 2006). To avoid misclassifications caused by intensity overlaps, many popular brain tissue segmentation toolboxes, including SPM and FSL, rely on probabilistic atlases to spatially constrain segmentation (Fonov et al., 2011), and provide initial estimates of the individual tissue maps. As the standard atlases are derived from the MR scans taken of the adult population, they likely do not

match tissue distribution in children. This often leads to errors in tissue labeling, especially in younger children.

There is a need to create children-specific processing protocols that would take into account the significant variability in brain structures depending on the age group. As current approaches tend to suffer from improper tissue segmentation (Phan et al., 2018), the accuracy of the measures obtained from neonates and young children is compromised. It can be especially critical in longitudinal studies, which generally aim to follow changes in the brain structure throughout the participants' lives. Preprocessing errors, especially if they cause random errors, can lead to inadequate representation of younger age groups, thus skewing the results.

1.2 Quality of pediatric MRI data

As we know, it is difficult for young children to remain still for long periods of time. This is especially challenging in the case of MR scans as the environment is unfamiliar and can be uncomfortable even for adults. While in clinical work patients are often sedated for the scan duration, in research it is very rarely utilized as sedation is still associated with some health risks and thus not, understandably, considered ethical. Additionally, the use of sedation can be costly as it typically requires pediatric anesthesiologists and experienced nurses to be present. Consequently, MRI data obtained from the pediatric population is highly susceptible to distortions and other poor data quality due to movement.

Subject motion during the data acquisition can lead to motion artifacts such as ghosting and blurring (Brown et al., 2010). Motion artifacts do not have to be severe to interfere with the analysis. Even subtle motion that cannot be detected by the human eye can cause significant errors in volumetric and morphometric measures (Alexander-Bloch et al., 2016; Reuter et al., 2015). While in some cases it is possible to correct the motion bias, it is still quite common for pediatric scans to be too severely distorted to be usable, reducing the amount of the data available for analysis.

Another challenge in pediatric neuroimaging arises from the myelination process and corresponding contrast changes. Neonatal brain shows reversed tissue contrast compared to adults on T1 images caused by low myelination: the intensity of white matter is lower than grey matter (Branson, 2013). During the time period between 6 and 12 months of age, the differentiation between grey and white matter becomes more difficult as both tissues can have similar intensities (Saunders et al., 2007). While the active myelination phase lasts longer and is most

prominent up to approximately until 2 years of age, after 12 months tissue intensities become adult-like, making segmentation processes easier compared to neonates.

1.3 Volumetric measures

One type of measurement that can be obtained from MRI is the volume of the brain structures, as well as total GM and WM volumes. It is a quite straightforward way to quantify the structural anatomy of the brain, which allows us to track the volumetric changes during healthy aging or to analyze the correlation of brain volumes with different conditions.

Volumetric measures have been actively investigated in connection to normal developmental processes (Gennatas et al., 2017; Gilmore et al., 2012; Gogtay et al., 2004; Hu et al., 2013). Further, abnormalities in adjusted total intracranial volume, as well as grey matter volume, have been shown to be present in children with attention-deficit/hyperactivity disorder, autism spectrum disorders and dyslexia (Batty et al., 2010; Nordahl et al., 2011; Ramus et al., 2018; Castellanos et al., 2002). Grey matter volume abnormalities were also detected in metabolism-associated conditions such as type 1 diabetes (Mazaika et al., 2016), childhood obesity (Bauer et al., 2015; Perlaki et al., 2018), and various psychiatric and neurological conditions such as bipolar disorder (DelBello et al., 2004), psychosis (Okada et al., 2018), epilepsy and anxiety (Jones et al., 2015), depression (Merz et al., 2018).

1.4 VBM

Voxel-based morphometry (VBM) is a neuroimaging method that involves a voxel-by-voxel analysis of grey matter differences. This method was originally developed (Ashburner & Friston, 2000) to investigate the differences in brain morphology between groups of subjects and has been actively utilized in neurodevelopmental research since.

VBM analysis starts with the segmentation of brain-extracted MR images into grey matter, white matter, and cerebrospinal fluid. Grey matter is further spatially normalized by non-linear transformation to a chosen template and smoothed, followed by voxel-wise statistical tests performed to identify regions that have a significant difference in grey matter structure.

Compared to the traditional morphometry approach, which involves the analysis of volumes of *a priori* chosen regions of interest (ROI), i.e., the hippocampus, VBM allows us to perform analysis on the whole-brain-level without introducing any spatial constraints.

Another important distinction between VBM and volumetric analysis is that the former's objective is significantly less straightforward in interpretation. While volumetric GM measures are easy to comprehend as they are exactly what the name states – volumes, measured in milliliters, VBM output can be interpreted in two different ways, depending on the optional step of the analysis called modulation, which corrects the values for the amount of warp from individual to template image yielding "proportional volumes". The choice of the terminology to describe what VBM results actually represent is, at times, not so consistent, which adds to the complexity of the matter. Depending on the source, GM volume, density, concentration, or mass can be used interchangeably, although none of these are absolutely correct.

GM segmentation procedure used in VBM does not give binary results, in case of which each voxel that is treated as GM would have an intensity equal to 1 and 0 for all other voxels. Instead, it assigns a relative value from 0 to 1 for each voxel, reflecting the relative amount of GM in the voxel compared to other tissues (Gennatas et al., 2017). As the GM segmentation is further spatially normalized to the template, it introduces volume changes. Consequently, in this case, VBM does not compare volumes, but rather regional GM concentrations or GM density (which should not be confused with the cell density) (Good et al., 2001; Mechelli et al., 2005). This approach is referred to as "non-modulated" VBM.

"Modulated" VBM, which is currently the default VBM method, accounts for volume changes due to spatial registration from the subject's native space to the template by adjusting the VBM maps using the Jacobian determinant calculated from the deformation matrix (Ashburner & Friston, 2003). In this case, VBM detects regional differences in volume, but in terms of the absolute amount of GM in the region (Good et al., 2001; Mechelli et al., 2005). Results of the modulated VBM in literature are often referred to as changes in GM (regional) volumes or changes in GM mass. However, it is quite common to see the use of the word "density" as well, making it indistinguishable from non-modulated VBM. This study uses the modulated version of the VBM analysis and will use the term "volume" for the description of the results.

Despite these shortcomings, VBM has frequently been used and allows us to detect subtle neuroanatomical differences and presents a valuable tool for MRI analysis that complements the classic volumetric comparisons.

VBM has been successfully used in the pediatric population to investigate brain abnormalities in subjects with various neurological and neurodevelopmental disorders (Carducci et al., 2013; McAlonan et al., 2005; Tondelli et al., 2018; Villagran et al., 2013; Xiao et al., 2014) as well

as to study regional neural tissue differences in obese (Kennedy et al., 2016; Perlaki et al., 2018) and healthy children (Gennatas et al., 2017; Ou et al., 2016; Peterson et al., 2003). The work of Gennatas et al. is of particular interest regarding the discussion on how to interpret VBM results. In this study, it was demonstrated that GM volumes, density, and modulated VBM volumes do not necessarily follow similar trajectories. VBM analysis performed on the young subjects ranging in age from 8 to 23 years showed that GM volume significantly decreases with age, while GM density increases and modulated VBM volumes show only a slight decline. Therefore, it is important to remember that volumetric measures and VBM-derived differences in volume are not directly comparable, but only as results that supplement each other.

1.5 Sexual dimorphism in neurodevelopment

Gender differences in total intracranial volume (TIV) have been reported for both adults and children (Gilmore et al., 2007). Over the life span, TIV is greater in males by approximately 10%, with grey matter showing a similar tendency.

Sex-related differences in developmental trajectories can be seen in grey matter development as well. Cortical grey matter growth follows an inverted U shape with peaks of volume happening earlier for females at approximately 10 years of age compared to 14 years for males (Giedd et al., 2012).

In contrast to grey matter, white matter growth continues throughout adolescence into adulthood. WM growth is more rapid for males, and the magnitude of the differences between sexes in WM volumes increases with age (Giedd et al., 2012).

2. Aims and hypothesis

This study aims to characterize the gross tissue changes that occur within the brain from toddlerhood (18-24 months) to 5 years of age. An additional objective is to perform exploratory analysis to find gender and age correlations with cortical volumes in children.

Previously reported sexual dimorphism is expected, including a relatively bigger GM volume for females compared to age-matched males after correction for total intracranial volume.

From the methodological standpoint, this study aims to provide children-tailored recommendations and assess the feasibility of the use of MRI processing software packages commonly used in the adult population as well as conclude on which tools seem to be best suited for the pediatric population of a given age.

3. Methods

3.1 Participants

The study utilized subjects belonging to the FinnBrain Birth Cohort Study (Karlsson et al., 2018) – longitudinal study investigating the combined influence of environmental and genetic factors on child development. MR scans of 38 healthy, normally developing children (see Table 1) were used for data analysis.

Age	Total number	Male	Female
18 months	1	0	1
24 months	7	3	4
5 years	30	15	15

Table 1. Study population demographic description.

For analysis purposes, participants have been further separated into two age groups: "toddlers" (18-24 months) and 5-year-olds.

3.2 MRI Acquisition

MR imaging was performed at the Turku University Hospital with Siemens Skyra 3T scanner. The T1-weighted images were acquired using MPRAGE sequence with $1mm^3$ spatial resolution, and with the following parameters: 172 slices, 256 mm FOV, TR 1900 ms, TE 3.26 ms, 9° flip angle.

3.3 Data analysis

For the study two different preprocessing pipelines were used: SPM12 (www.fil.ion.ucl.ac.uk) with use of the CAT-12 toolbox (Gaser & Dahnke, 2016) and FSL (Jenkinson et al., 2012).

Additionally, images were segmented using the FreeSurfer toolbox (www.surfer.nmr.mgh.harvard.edu) that has been chosen as the reference technique to estimate the performance of CAT-12 and FSL segmentations. FreeSurfer has been shown to provide good accuracy in tissue segmentations compared to other automated segmentation software packages in the pediatric

population (Ghosh et al., 2010; Mayer et al., 2016), however, being still inferior to current gold-standard – manual segmentation (Schoemaker et al., 2018). As manual segmentation is an extremely time consuming, it is not an optimal choice for studies with even relatively large samples. With that in mind, for purposes of this study, visually inspected and corrected FreeSurfer output decided to be utilized as an etalon for segmentation accuracy comparisons. Of note, the FinnBrain Neuroimaging Lab has extensive quality control and manual edit protocol for controlling the quality of FreeSurfer output, which includes but is not limited to all steps recommended by Enigma consortium (http://enigma.ini.usc.edu/protocols/imaging-protocols/).

3.3.1 Intermediary image manipulations

FSL tools have shown a particular sensitivity towards the field of view (FOV) of the image. The presence of the large amount of non-neural tissues of the neck and lower head resulted in an overestimation of brain volume at the step of brain extraction.

To assist the FSL BET algorithm, images' FOV was edited using FSL command "robustfov" to exclude the neck part of the image. Removal of the part of non-neural tissues from the image significantly improved the accuracy of the produced brain masks.

The cropped images were back-registered to full-size, and the same transformation matrix was used to bring segmentations to space with the same dimensions to assess the accuracy of the segmentations against CAT-12 and FreeSurfer output.

3.3.2 FSL VBM

VBM data preprocessing was performed on structural images using FSL VBM tool (www.fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLVBM), which includes following steps: brain extraction, tissue segmentation (into grey matter, white matter and CSF), creation of study-specific GM template, non-linear registration of GM tissue images to the template and smoothing.

Brain extraction for 5-year-olds was performed with FSL BET parameters "-b -R -f 0.3". Brain masks were visually inspected to confirm successful skull stripping.

Brain masks for toddler scans provided by FSL BET were consistently erroneous, despite the FOV and extraction parameters manipulations. Consequently, additional tools have been used in attempts to obtain better quality masks. ROBEX (Iglesias et al., 2011) and LABEL skull stripping toolkit (Shi et al., 2012) were tried, but the results were not satisfactory. After reviewing the available methods, it was decided to utilize brain masks provided by FreeSurfer software

(http://surfer.nmr.mgh.harvard.edu/) and its "recon-all" command, with flags "-autorecon1 -wsmore". Obtained brain masks were then transformed back from FreeSurfer conformed space to the subjects' native space and further used in VBM analysis.

After gross tissue segmentation, partial volume GM images were non-linearly registered to GM ICBM-152 standard space template. Resulting images were averaged and flipped along the x-axis to create a symmetric study-specific GM template. Native space GM images were then non-linearly registered to the template and underwent Gaussian smoothing with a sigma of 3.3973 mm, which is equivalent to FWHM of 8 mm.

3.3.3 FSL SIENAX

FSL SIENAX (Smith et al., 2002) was used to obtain volumetric measures. T1-weighted images underwent brain extraction and affine registration to the standard MNI template. Normalized brains were used for tissue segmentation, and estimates of total intracranial, gray, and white matter volumes, normalized for the head sizes as well as non-normalized, were reported.

As SIENAX takes as an input original T1-weighted images and does not provide an option to add custom brain-extracted images, it has not been possible to use it on toddler scans (as FSL BET fails in the case of younger children). Thus, only the analysis of 5-years old participants was performed.

3.3.4 CAT-12

Using SPM toolbox CAT-12, T1-weighted scans were corrected for bias-field inhomogeneities, segmented into grey matter, white matter, and cerebrospinal fluid and spatially normalized using diffeomorphic non-linear registration algorithm (DARTEL). Total intracranial volume, as well as GM, WM, and CSF volumes, were calculated based on segmentation in the native space to be further used as covariates in statistical analysis. GM images in the standard MNI space were then smoothed with a Gaussian kernel of 8 mm, matching the one used with FSL VBM.

3.3.5 FreeSurfer

Visually inspected and edited (Pulli et al., in preparation) FreeSurfer segmentations were registered back to subjects' native space. Grey and white matter masks with excluded cerebellum and brainstem were created for assessing the quality of FSL and CAT-12 segmentations. For each subject, cerebellum and brainstem masks have also been created to exclude those areas

out of FSL and CAT-12 images as well. This was done since the primary interest was on the cortical and subcortical nucleus segmentation.

As subcortical segmentation can be especially challenging due to non-population-specific tissue priors (Loh et al., 2016), exclusion masks containing subcortical nuclei, cerebellum, and brain-stem were created to additionally assess agreement between segmentation procedures in cortical GM.

For TIV as well as grey and white matter volumes aseg statistical data was extracted. Total white matter volume was estimated as the sum of cerebral WM, cerebellar WM, brainstem, and corpus callosum.

3.4 Statistical analysis

3.4.1 Volumetric measures

Total intracranial, grey matter and white matter volume estimations were directly compared using Pearson's correlation coefficients to evaluate the inter-method agreement. A statistical significance threshold of $p<0.05$ before Bonferroni correction for multiple comparisons was chosen.

3.4.2 VBM analysis

To investigate differences in brain structure between genders and the two age groups, smoothed grey matter images were entered into 2x2 ANOVA GLM model with age and sex as explanatory variables.

For FSL output permutation-based nonparametric testing was performed using FSL randomize method (Winkler et al., 2014) with 10000 permutations. Significant clusters were identified using threshold-free cluster enhancement (TFCE) with the significance threshold $p < 0.05$.

In the case of CAT-12 output, statistical analysis was performed using SPM. The significance threshold was set up as $p<0.05$, FWE corrected. According to CAT-12 manual recommendations, TIV was used as a covariate.

3.4.3 Dice scores

To evaluate the quality of brain segmentations, the Dice similarity coefficient (DSC) was used (Dice, 1945; Taha & Hanbury, 2015). DSC is a spatial overlap index, value of which varies

from 0, meaning no overlap between two binarized segmentations, to 1, which indicates complete overlap. The calculation is performed using the following simple formula: DSC(A, B)= 2(A∩B)/(A+B), where ∩ refers to the intersection, A and B refer to segmented tissue provided by processing pipeline under consideration and chosen golden standard accordingly. For purposes of this work, DSC computation was estimated using Python 3.8.1, NiBabel (NIfTI images processing), NumPy (mathematical functions) and os, glob, and pathlib libraries (system and shell manipulations).

Visually inspected and manually corrected FreeSurfer segmentations have been chosen as a gold standard or the "closest to ground truth" segmentations for this study.

4. Results

4.1 Brain extraction

As discussed previously, the brain extraction step presents significant difficulties for pediatric data. FSL tool for skull stripping, BET, utilizes variability in intensities of pixels representing different tissues to differentiate between tissues. In the adult population, brain tissues are surrounded by a noticeable dark band of CSF in subarachnoid space and skull, which is further encircled by hyper-intensive voxels of the scalp. The resulting changes in voxel intensities are used to determine brain outlines. However, as the distance between the cerebral cortex and the scalp is significantly smaller in younger children, changes in intensity may not be sufficient, resulting in the estimated mask expanding further. For the data used in this study, FSL BET produced erroneous brain masks for the images of children aged between 18 and 24 months. The best results have been obtained using the following parameters: "-b -R -f 0.3 -g -0.2". However, the resulting brain masks still contained a large amount of non-neural tissues.

LABEL and ROBEX algorithms faced the opposite problem, erroneously removing significant areas of the cortex. Figure 1 illustrates brain extraction results for one of the toddler subjects for FSL BET, LABEL, ROBEX, and FreeSurfer. Based on similar results obtained for the rest of the subjects, it was concluded that FreeSurfer is the most reliable tool for skull stripping for the pediatric population of the given age.

4.2 Neural tissue volumes

Total intracranial, grey, and white matter volumes were estimated using CAT-12, FSL SIENAX, and FreeSurfer. The results are detailed in Table 2. Only data on 5-year-olds is

available for FSL SIENAX, as it has not been possible to utilize it with toddler subjects,making the overall mean skewed compared to the other two software tools.

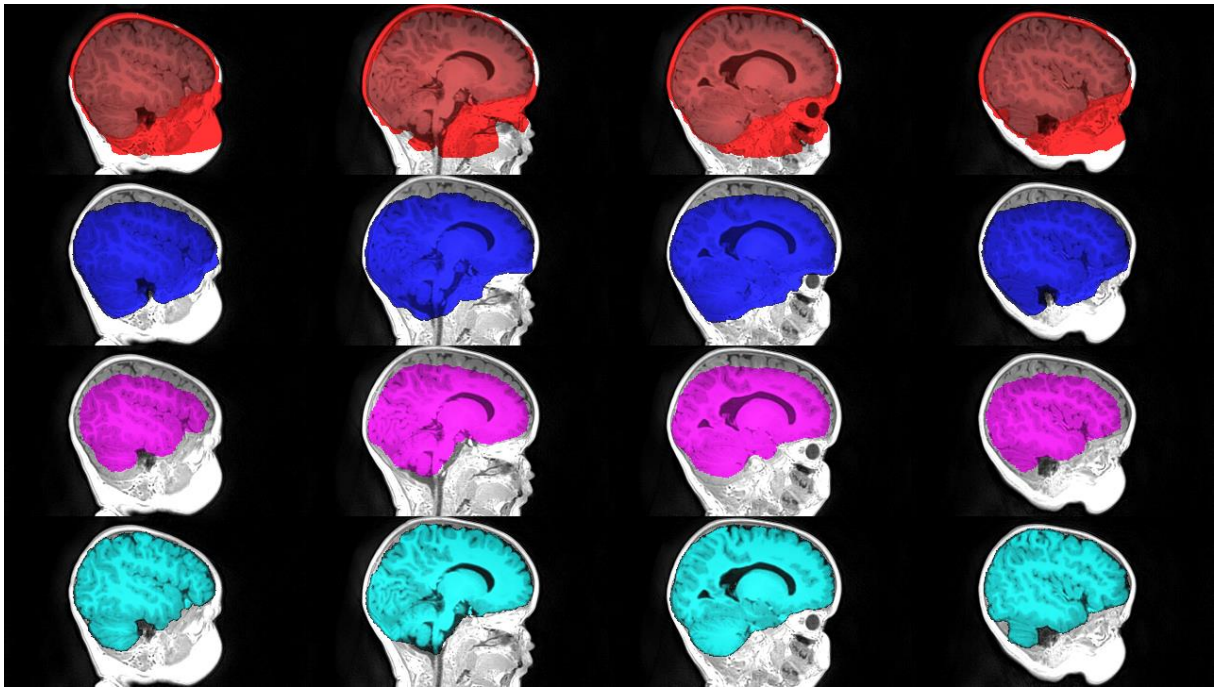

Figure 1. Example of brain extraction results from FSL BET (red), LABEL (blue), ROBEX (violet) and FreeSurfer (cyan).

	TIV, cm^3			GM volume, cm^3			WM volume, cm^3		
Gender	CAT-12	FSL SIENAX*	Free-Surfer	CAT-12	FSL SIENAX*	Free-Surfer	CAT-12	FSL SIENAX*	Free-Surfer
Female	1379(106)	1257(97)	1353(209)	765(55)	748(54)	752(62)	402(45)	509(47)	423(71)
Male	1498(106)	1369(80)	1517(220)	793(48)	799(49)	805(67)	460(55)	569(41)	425(61)
Mean diff.**	7.9%	8.2%	10.8%	3.5%	6.4%	6.6%	12.6%	10.5%	0.5%

*Table 2. Total intracranial, grey, and white matter volume estimations obtained using CAT-12, FSL SIENAX, and FreeSurfer. Data reported as Mean(SD). *: Values for SIENAX are provided only for 5-years-old subjects. **: % difference between means is calculated as* $100 * \frac{\overline{Male} - \overline{Female}}{\overline{Male}}$ *for each column.*

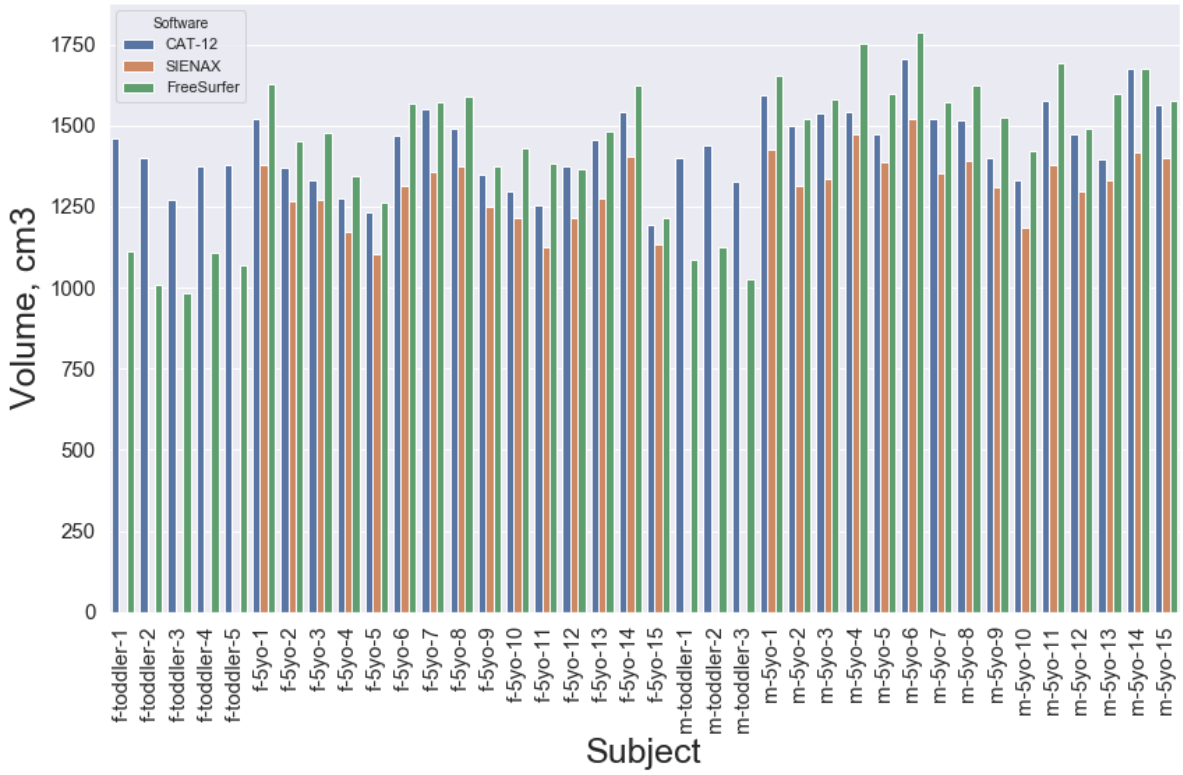

Figure 2. Total intracranial volume estimation obtained using CAT-12, FSL SIENAX, and FreeSurfer. Subjects are coded in the following way: "gender-age group-number".

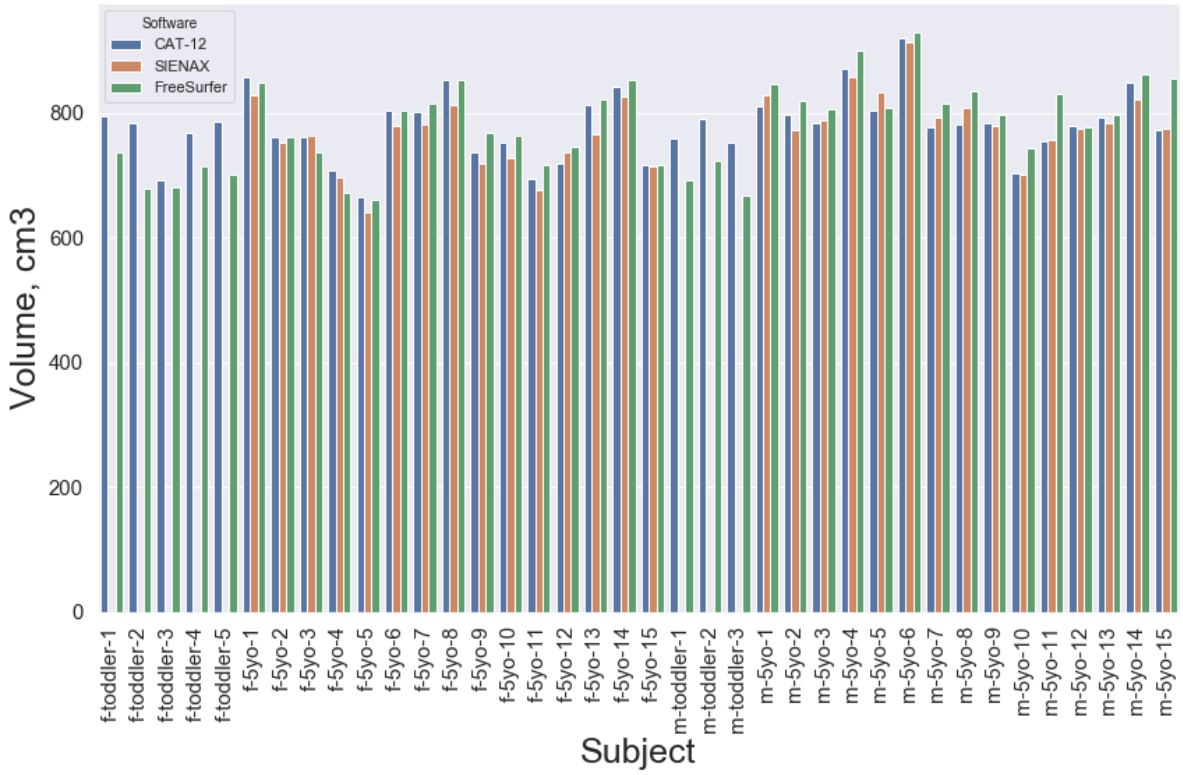

Figure 3. Grey matter volume estimations.

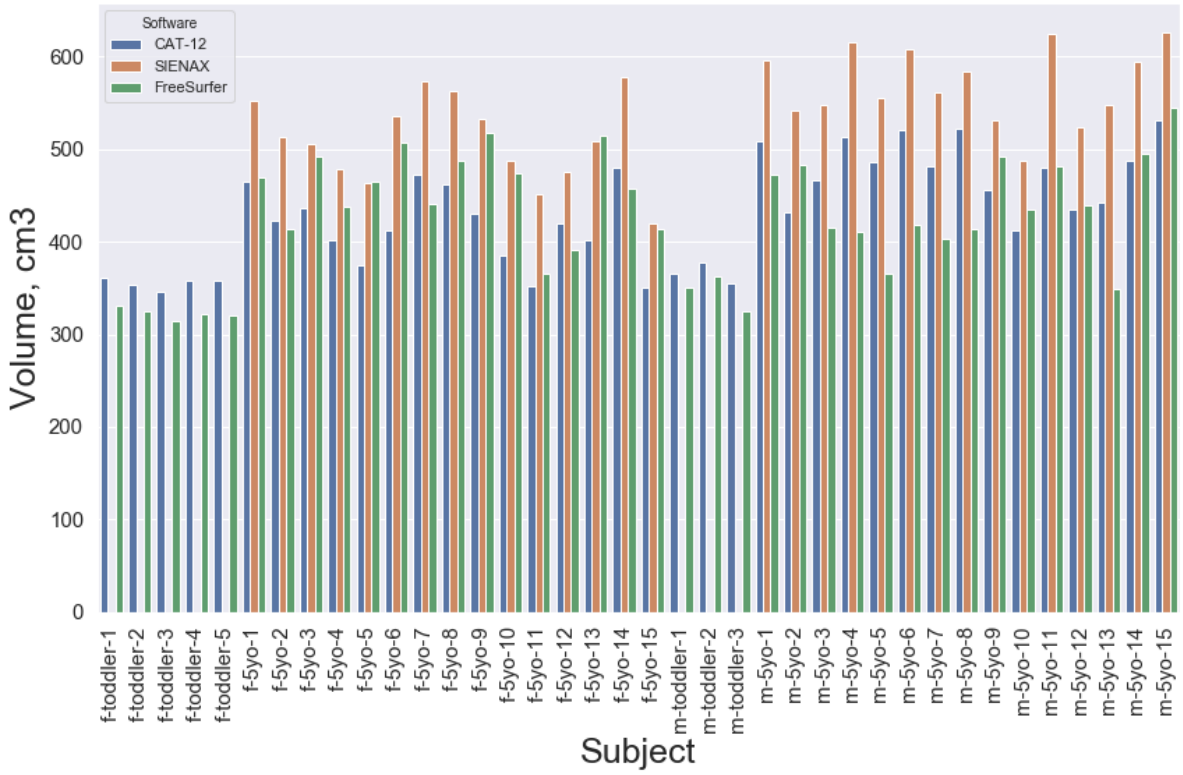

Figure 4. White matter volume estimations.

As can be seen in Figure 2, total intracranial volume estimates are not very consistent across image processing packages, with FSL SIENAX giving the smallest estimates. Grey matter volume estimates (Figure 3) are the most consistent out of the measures across software packages as well as across both age groups. In the case of white matter volume (see Figure 4), FSL SIENAX provides the highest estimates. The expected increase in white matter volume between toddlers and 5-years-olds is also clearly visible.

Assessing the differences in the obtained results, it is important to consider that the methods utilized to estimate the volumes differ from software to software. While the most straightforward way to calculate TIV from an MRI scan would be to count the number of voxels within the skull borders, such approach is rarely used due to the ambiguity of the skull-CSF border in T1-weighted images. Both FreeSurfer (Buckner et al., 2004) and FSL SIENAX (Smith et al., 2002) base their estimations on scaling factors and transformation matrices from the subject's native space to the standard MNI space. While CAT-12 documentation does not clearly state the method used for the neural tissue volume estimation, it is likely to be based on the warps as well. Consequently, while it is clear that volumetric measures are not consistent between software packages, they may still correlate with each other.

Table 3 presents the correlation coefficients for the compared pairs of software tools and corresponding p-values. In the case of 5-year-olds, both TIV and GM volumes are strongly correlated. However, the addition of the toddler data significantly lowers the correlation coefficients for all types of tissue between CAT-12 and FreeSurfer. Interestingly, FreeSurfer's white matter volume estimations do not seem to correlate with either CAT-12 or FSL SIENAX, while the latter two show a strong correlation.

		TIV	GM	WM
CAT-12 vs FreeSurfer	5-years-olds	0.911 (2.93E-16)	0.911 (2.78E-12)	0.091 (0.633)*
	whole population	0.674 (3.53E-6)	0.787 (4.56E-9)	0.528 (6E-4)
CAT-12 vs FSL SIENAX		0.932 (7.44E-14)	0.974 (2.73E-15)	0.935 (4E-14)
FSL SIENAX vs FreeSurfer		0.958 (1.05E-16)	0.912 (2.46E-12)	0.222 (0.238)*

*Table 3. Pearson's correlation coefficients and paired t-test p-values for volumetric measures, comparing CAT-12, FSL SIENAX, and FreeSurfer output. As FSL SIENAX could not provide volume estimations for toddlers, correlation with FSL SIENAX is calculated only for the 5-years-old part of the study population. For CAT-12 and FreeSurfer, correlation coefficients were calculated for both 5-year-olds subset and the whole population. * marks the results that have not reached statistical significance.*

4.3 VBM

4.3.1 FSL

FSL VBM revealed statistically significant clusters both for age and gender effect. However, no significant interaction between age and gender effects was observed.

A comparison between females and males showed a small area of increased GM volume in the left auditory cortex (Figure 5).

FSL VBM further revealed statistically significant areas in which the 5-year-olds showed higher GM volume compared to toddlers: right superior temporal gyrus, putamina, precentral and superior frontal gyri (more pronounced in the right hemisphere), precuneus cortex, frontal orbital cortex, temporal poles and cerebellum (Figure 6). The large cluster next to the brainstem is likely to be due to differences in brain extraction between FSL BET (5-year-olds) and FreeSurfer (toddlers).

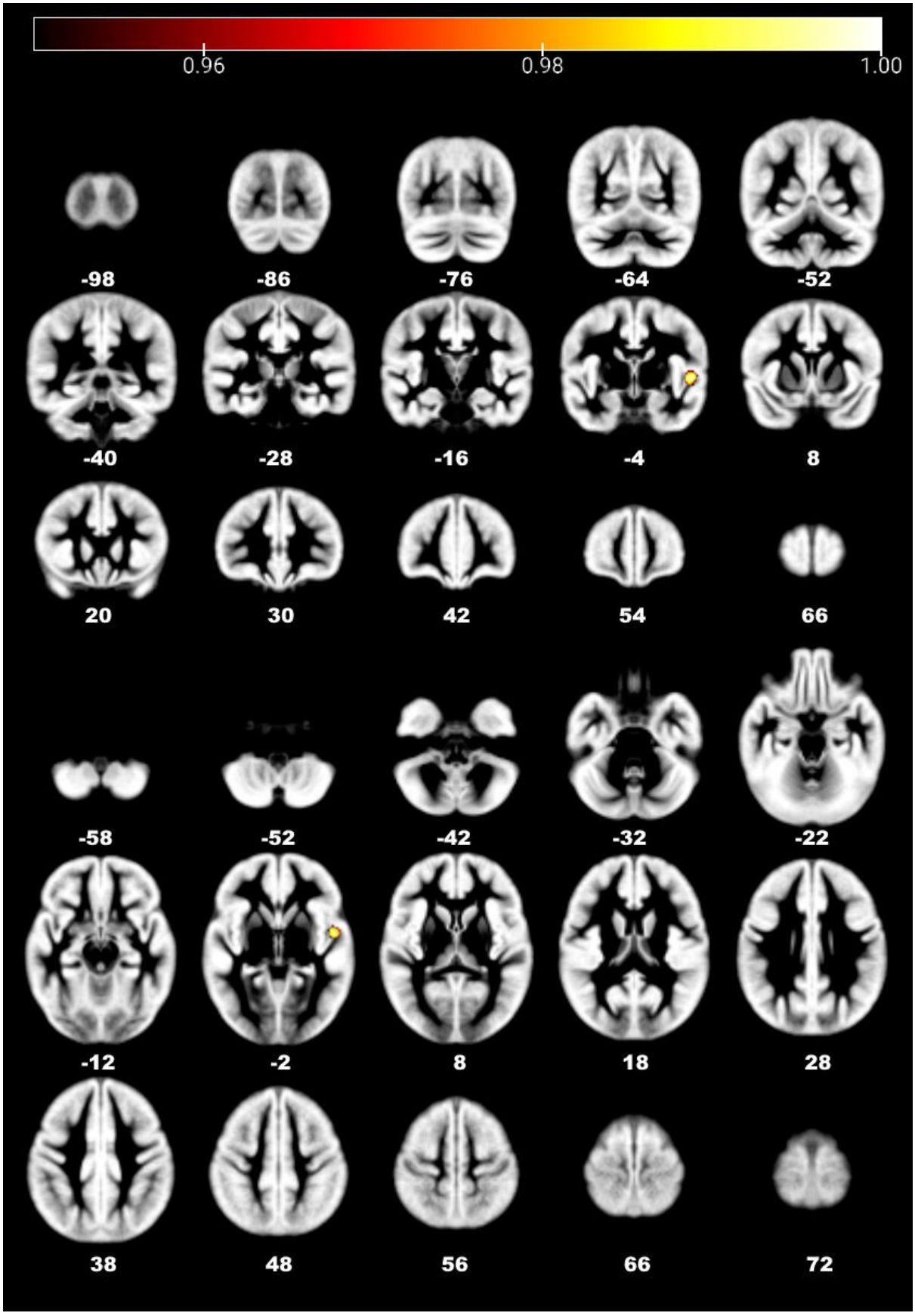

Figure 5. FSL-VBM detected increased GM volume in females compared to males in the left auditory cortex.

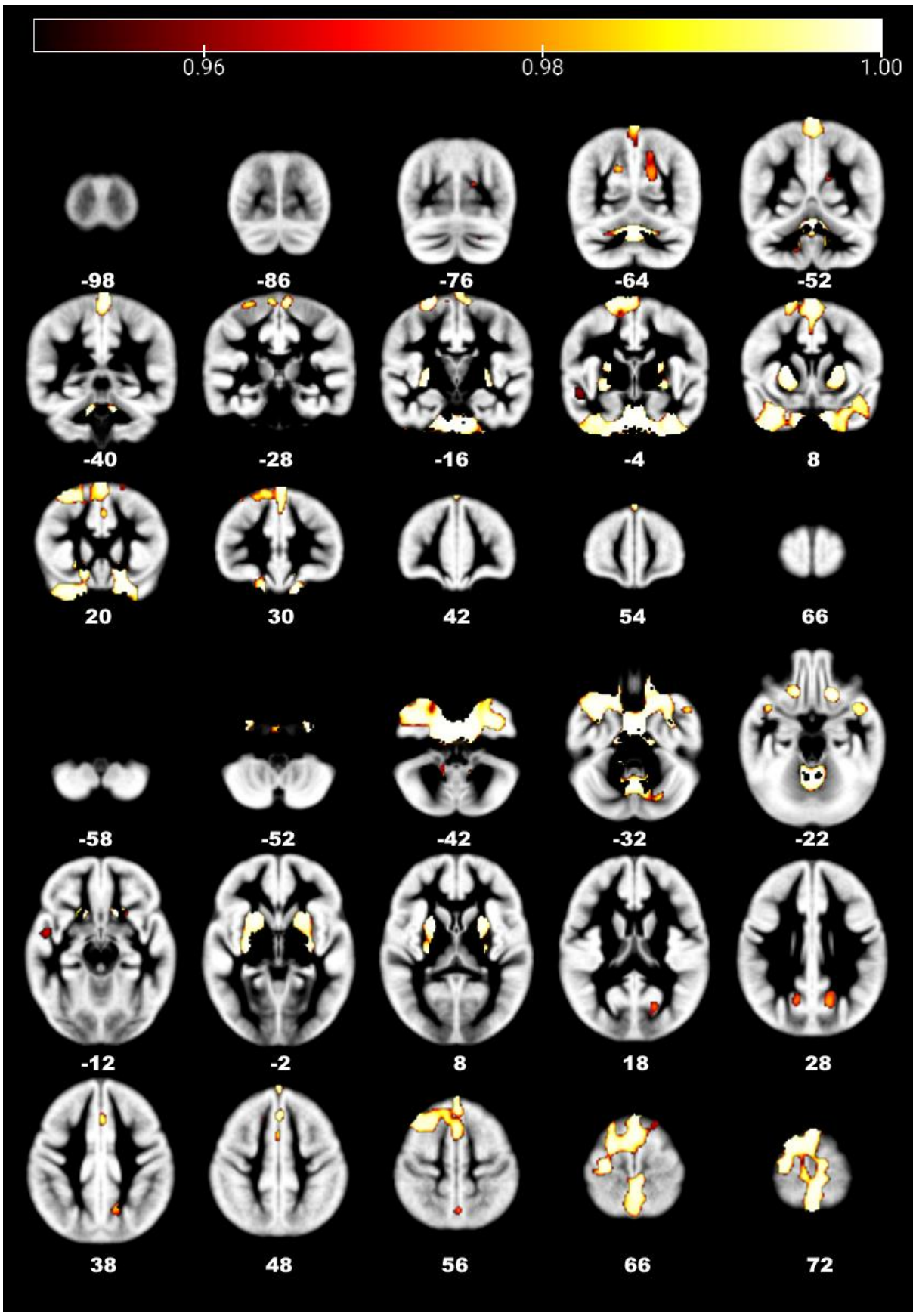

Figure 6. FSL-VBM results revealed increased GM volumes in 5-year-olds compared to the toddlers in putamina, frontal lobe, precuneus cortex, temporal poles, and cerebellum.

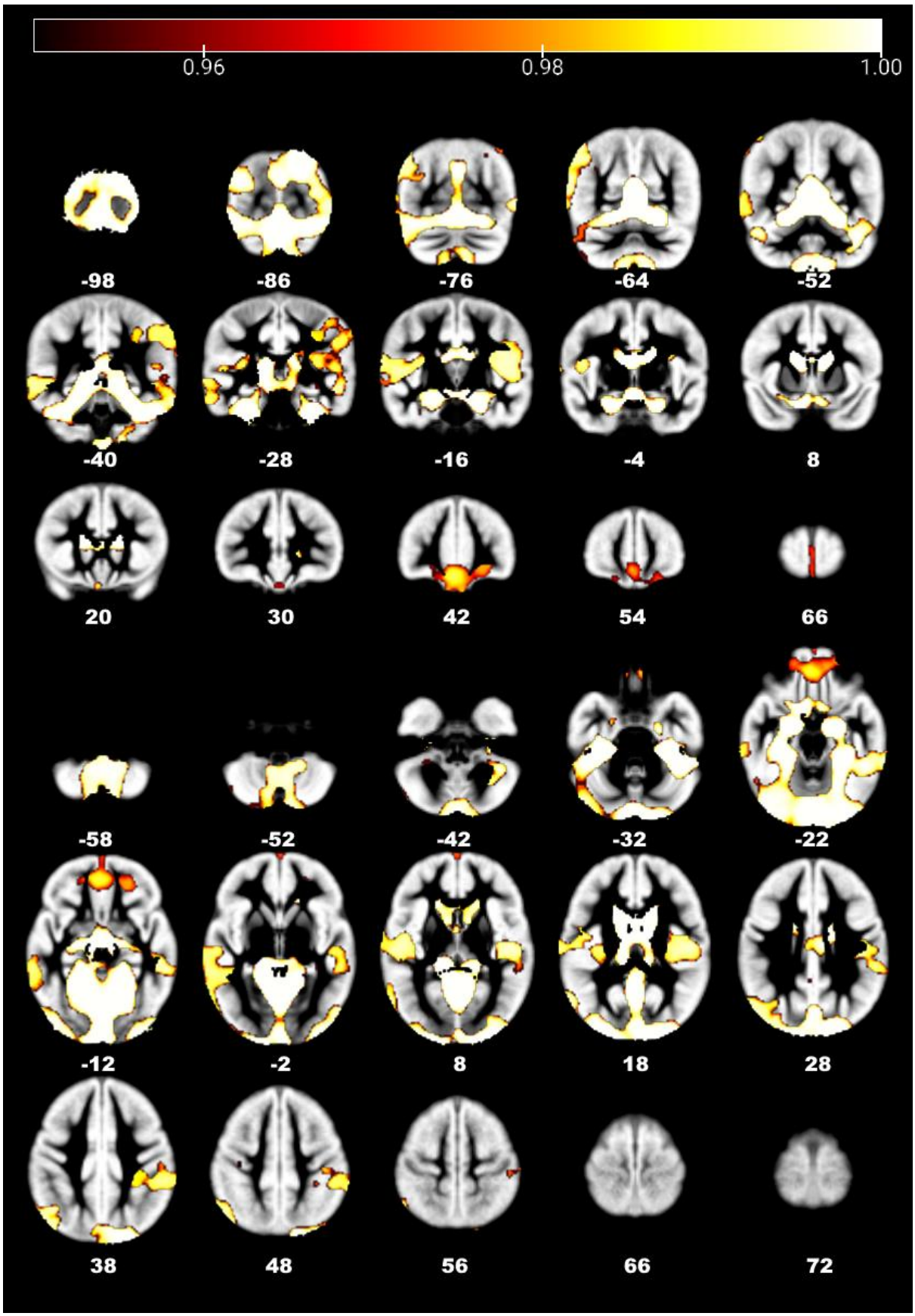

Figure 7. FSL-VBM results showed that in toddler subjects compared to 5-year-olds a higher GM volume is presented in occipital pole, lingual gyrus, middle temporal gyrus, frontal pole, and medial frontal cortex.

In the case of toddlers compared to 5-year-olds, FSL VBM revealed statistically significant clusters in occipital pole, lateral occipital cortex in both hemispheres, caudates, thalami, left supramarginal gyrus, auditory cortex (Heschl's gyrus, planum temporale) on both sides, left one showing a larger cluster, frontal pole, medial frontal cortex in both hemispheres, lingual gyri, hippocampi, occipital fusiform gyri, cuneal cortex, cerebellum (Figure 7).

4.3.2 CAT-12

In contrast with FSL output, CAT-12 VBM results for gender differences have not reached a statistically significant level. One obvious explanation for that would be the small population size. It is also important to consider that the statistical algorithms used in FSL and CAT-12 differ, and it is likely for them to exhibit different levels of sensitivity. However, investigating the reasons behind the contradictory results exceeds the purposes of this study, and remains to be explored with a larger sample.

In 5-year-olds compared to toddlers CAT-12 revealed areas of increased GM volumes in putamina and the thalami (cluster on the left side is significantly larger) (Figure 8).

In toddlers compared to the 5-year-olds, higher GM volume areas were located in medial frontal gyri, frontal pole, cuneal cortex, lateral occipital cortex in both hemispheres, occipital pole, lingual gyri and cerebellum (Figure 9).

4.4 Dice coefficients

Gross segmentations of brain tissues into the gray and white matter were compared by creating an overlap map between the outputs obtained from each tool. Dice similarity coefficient (DSC) was calculated to act as a statistical validation metric to evaluate the performance of the obtained segmentations against quality-controlled FreeSurfer segmentations, which serve as the gold standard for the accuracy of the procedures.

Figure 10 illustrates the results of spatial overlap between the segmentations. As can be seen, CAT-12 demonstrated a more considerable variance between the dice scores, especially in the case of the white matter segmentations. At the same time, the quality of the CAT-12 white matter segmentations is generally higher compared to FSL.

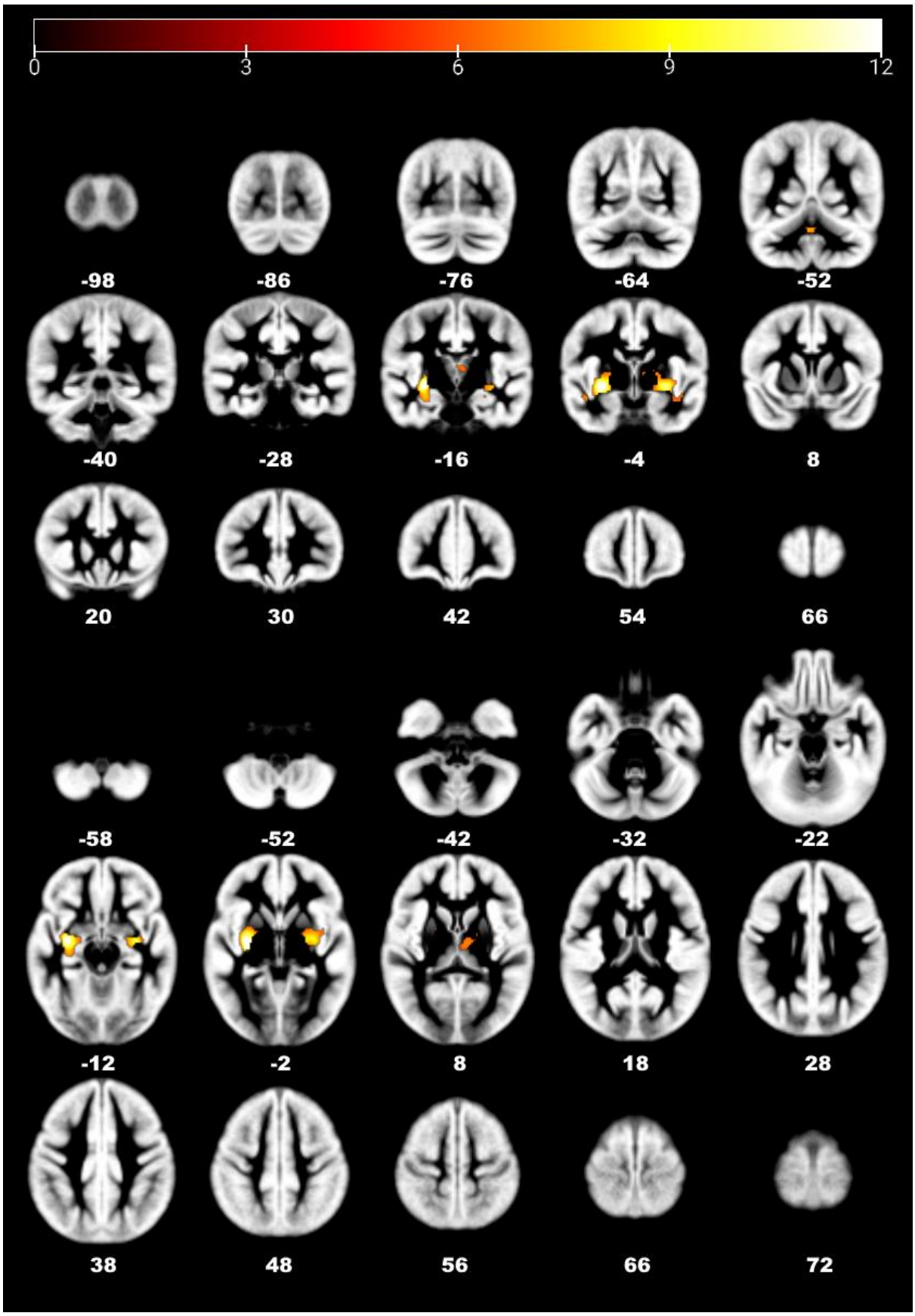

Figure 8. CAT-12 for 5-year-olds compared to toddlers showed clusters of increased GM volume in putamina and thalami.

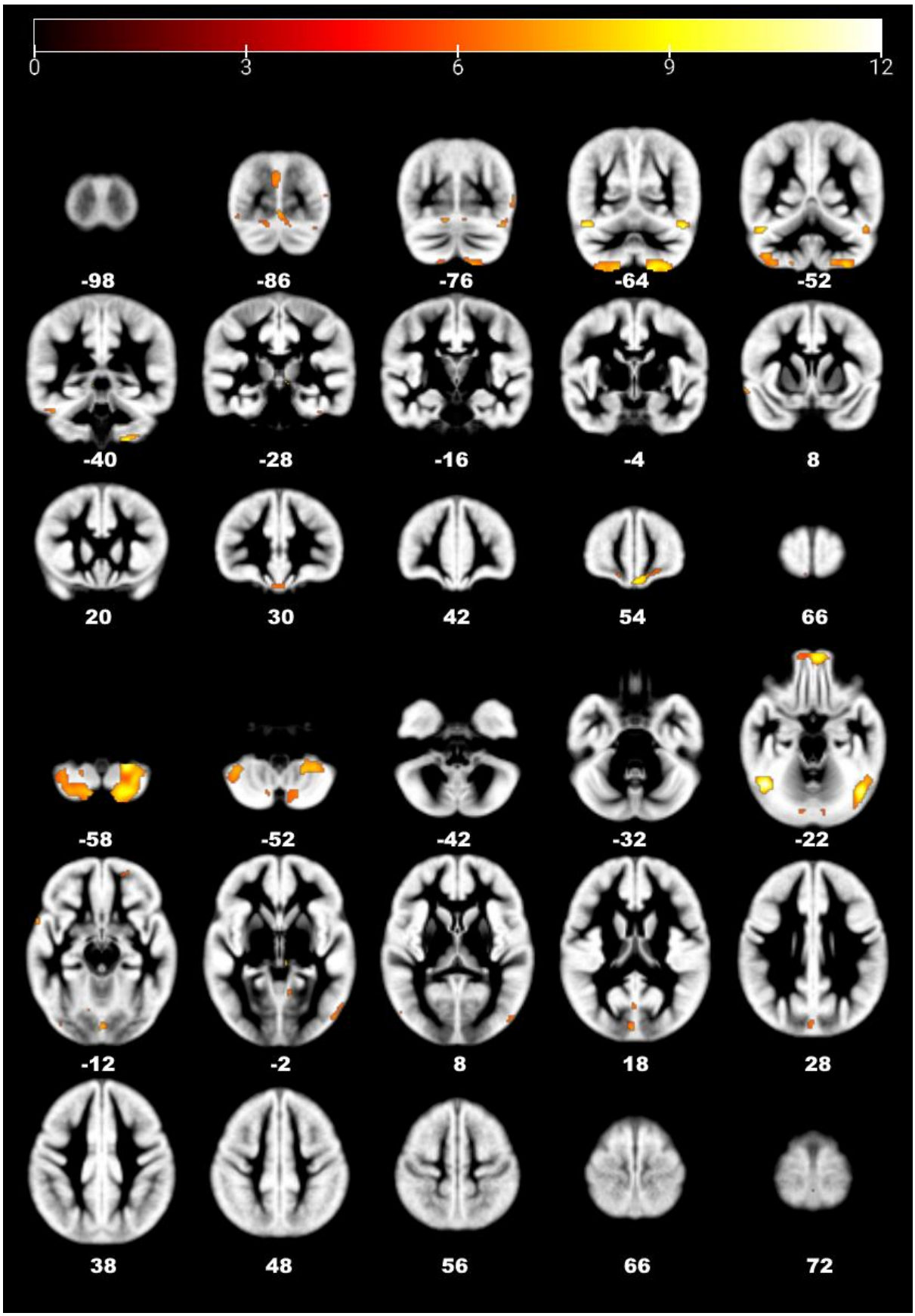

Figure 9. CAT-12 revealed areas with higher GM volume in toddlers compared to 5-year-olds in the frontal lobe, temporal occipital fusiform cortex, lingual gyri, and cerebellum.

FSL has been shown to be more robust regarding grey matter segmentations, especially in the case of younger children. However, the quality of the toddler scan segmentations is relatively low for both software packages.

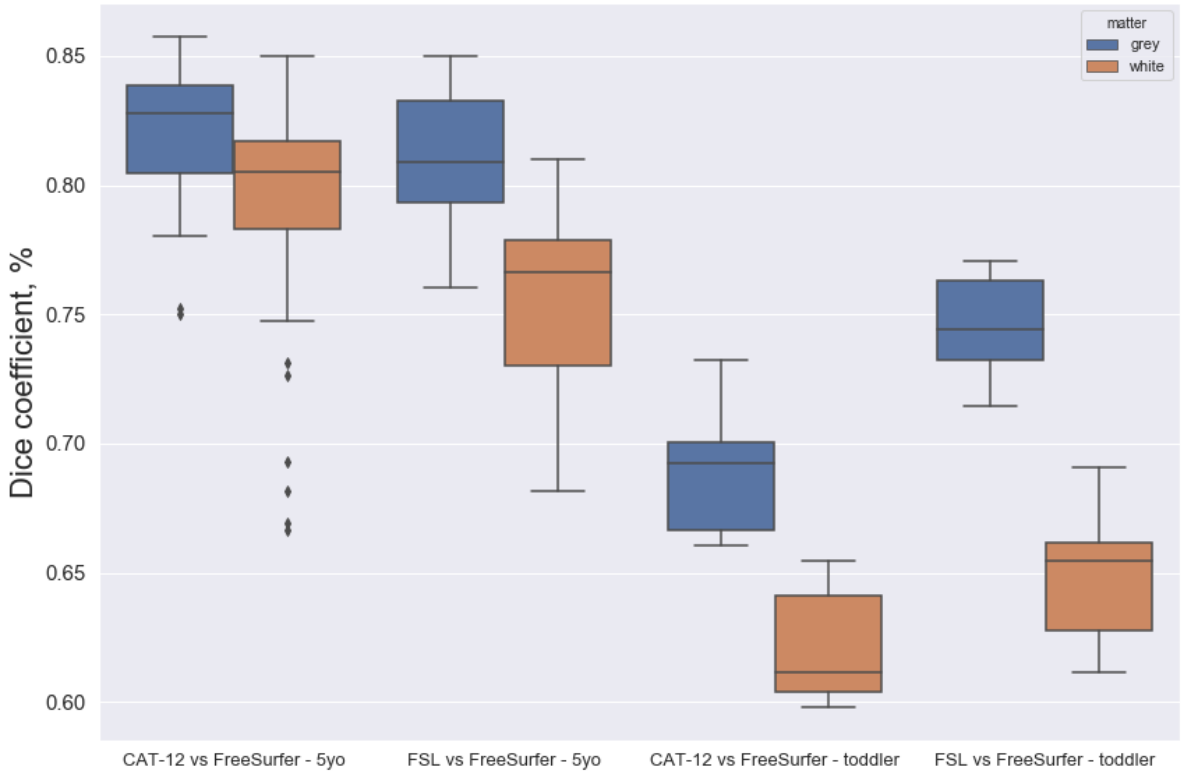

Figure 10. Dice coefficients for CAT-12 and FSL segmentations compared to FreeSurfer.

As the segmentation of subcortical structures presents even more challenges compared to cortical areas due to the lesser contrast between grey and white matter and the high dependency on tissue probability maps that do not take into account the age-related changes in the nuclei (Lorio et al., 2016). As erroneous segmentation of the deep brain structure would lead to lower Dice coefficients, the Dice scores of total grey matter segmentations were compared to the images including only the cortical part of grey matter to assess the magnitude of the effect of subcortical nuclei on the quality of segmentation. However, the results (Figure 11) revealed that the exclusion of subcortical structures either did not improve the Dice coefficients at all or did so to a minimal extent.

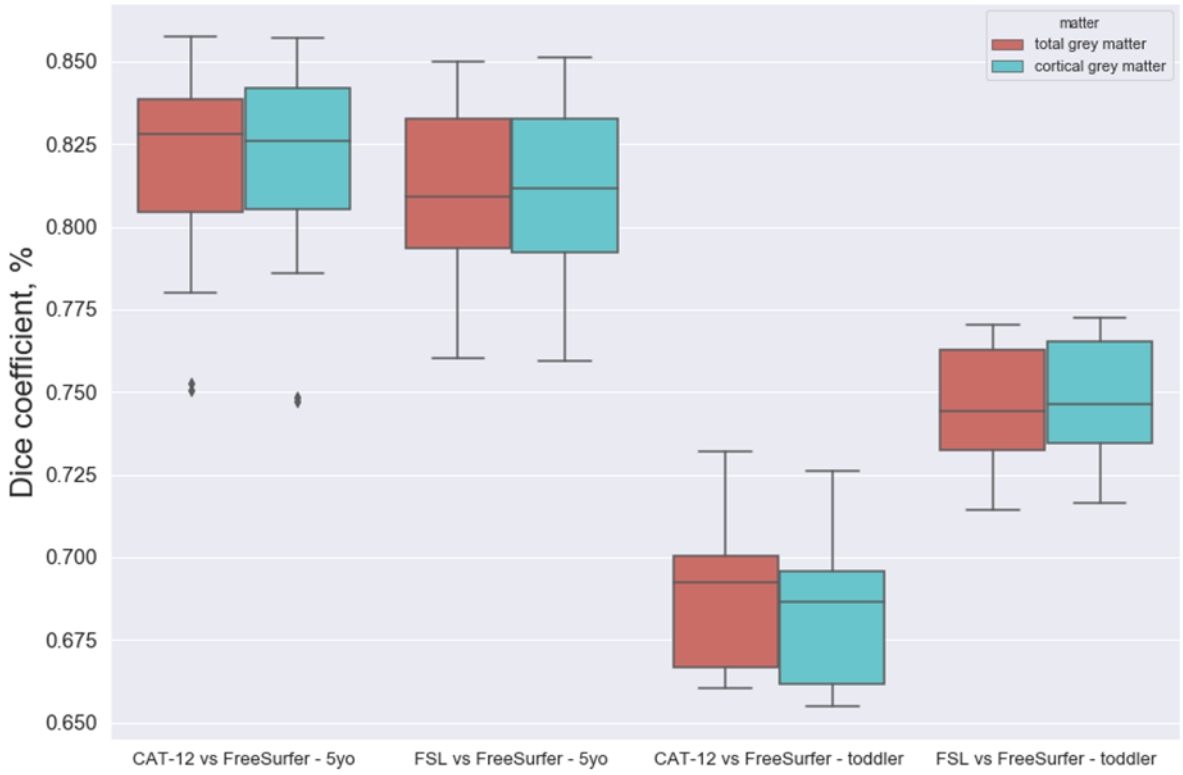

Figure 11. Dice coefficients for total and cortical grey matter segmentations.

	CAT-12 vs FreeSurfer		FSL vs FreeSurfer	
	toddlers	5-year-olds	toddlers	5-year-olds
GM	0.69 (0.03)	0.82 (0.03)	0.74 (0.02)	0.81 (0.03)
Cortical GM	0.68 (0.03)	0.82 (0.03)	0.75 (0.02)	0.81 (0.03)
WM	0.62 (0.02)	0.79 (0.05)	0.65 (0.03)	0.75 (0.03)

Table 4. Dice coefficients for GM, cortical GM, and WM. Values are reported as Mean(SD).

Due to the small sample size, it is difficult to fully assess the viability of preference of one tool over another in the case of older children. In the pediatric population of younger ages, taking into account that FSL is not sufficient to perform the skull stripping on its own, FreeSurfer segmentation would likely be the primary choice.

Finally, it was of interest to compare to what extent the CAT-12 and FSL segmentations differ from each other. While it does not necessarily give us information about the quality of the segmentation, it can help us to understand the overall agreement between tools' outputs.

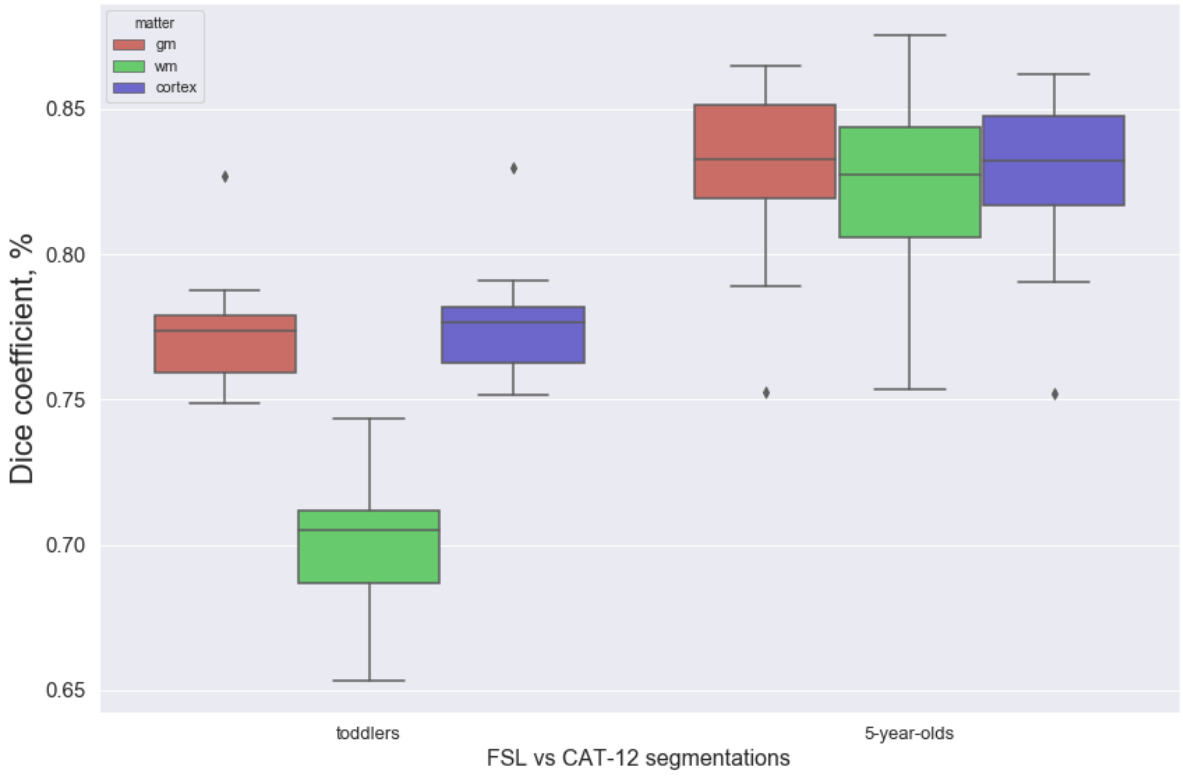

Figure 12. Dice coefficients for FSL segmentations compared against CAT-12.

As shown in Figure 12, the agreement between FSL and CAT-12 has quite similar magnitude as the agreement between two aforementioned tools and FreeSurfer, with WM segmentations being considerably less consistent across methods. Interestingly, there is noticeably better agreement between FSL and CAT-12 in the case of toddlers, than either tool has with Free-Surfer.

5. Discussion

This study investigated the age- and gender-related differences in children of preschool age and assessed the feasibility of adult processing tools in these ages. The results demonstrated that there is high inter-method variability in terms of tissue volume estimations, segmentations, and VBM results.

5.1 The challenges of image preprocessing

Extremely rapid brain development in early childhood presents not only a fascinating phenomenon to study, but also a significant challenge for the image preprocessing algorithms. While the MRI scans of 5-year-olds are close enough to adult ones to demand only minor adjustments (FOV correction, brain extraction settings calibration), in the case of younger children preprocessing can be an extremely toilsome process, involving the use of several tools and following multiple registration procedures to match the subject spaces across the software. Moreover, pediatric data demands extensive visual control of every subject after each step, which can be a problem in large-scale studies.

CAT-12 has proved itself to be much more robust when processing toddler subjects when compared to FSL. In the current study, it was decided to not create a custom age-specific template due to the relatively small sample size. However, using such templates might improve its efficiency even further.

In its turn FSL was not able to perform brain extraction procedure on toddler subjects. While it is possible to use brain extracted images produced by other tools in FSL VBM, FSL SIENAX does not provide such option. Thus the tissue volume estimations were not possible to obtain for younger children. The use of external tools may raise some problems with study repeatability, so in the light of our results, the use of FSL tools may be preferred only in the case of subjects of age 5 and older.

5.2 Volumetric measures

TIV difference between genders was observed across all software tools, being ~8% for FSL SIENAX and CAT-12 and ~11% for FreeSurfer. Despite certain disagreement between methods, the obtained numbers stay in line with the results observed in previously published literature. Larger-scale studies reported a ~6% difference for neonates (Knickmeyer et al., 2014), ~10% for older children and adolescents (Ball et al., 2012) and ~12% for adults (Ruigrok et al., 2014).

For GM volumes, the agreement was higher between CAT-12 and FreeSurfer, yielding a difference of 6.4% and 6.6%, while FSL SIENAX detected only a 3.5% difference. However, the FSL SIENAX estimations were done only for the 5-year-old part of the population, and that might have affected the result. Nonetheless, the estimated difference was significantly

lower than previously published 9 to 12% for children (Paus, 2010; Wilke et al., 2007) or 8% for newborns (Dean et al., 2018).

Differences in the WM volume estimations are the least consistent, varying from 0.5% for FreeSurfer to 10.5% for FSL SIENAX and 12.6% for CAT-12. In contrast with GM, FSL and CAT-12 WM estimations stay in line with existing literature showing 6 to 8% difference in neonates (Dean et al., 2018; Gilmore et al., 2007; Paus, 2010) and 7 to 12% in children (Paus, 2010; Wilke et al., 2007).

The results of this study clearly show that total intracranial, as well as GM and WM, volume estimations will be highly dependent on the chosen method. However, TIV and GM volume measures turned out to be highly correlated with all three software tools. WM volumes correlated well only between FSL SIENAX and CAT-12, which contradicts the previously published result (Ávila et al., 2019), where the correlation between FreeSurfer and FSL was significantly higher for WM than for TIV or GM.

While it is not possible to conclude the most accurate neural tissue volume estimation approach, it is likely that at least the TIV and GM volumes can be used for scaling purposes interchangeably between software tools. Nonetheless, it still demands to be tested in a larger study, preferably one comparing several age groups to get a better picture of the interchangeability of processing methods.

5.3 Discrepancy in VBM analysis

Before discussing VBM results, it is essential to address the significant differences in the results between FSL VBM and CAT-12. Dissimilarities in VBM results have been previously studied for FSL VBM and SPM8 in normal aging (Callaert et al., 2014) as well as in clinical population: amyotrophic lateral sclerosis (Rajagopalan et al., 2014) and multiple sclerosis (Popescu et al., 2016) patients.

While intuitively, it would seem that the results obtained from two similar procedures should replicate each other, in practice, the output is heavily influenced by the differences in preprocessing approaches, particularly in segmentation and registration.

The CAT-12 segmentation relies on several techniques; first, it utilizes SPM unified segmentation (Ashburner & Friston, 2005) to perform the spatial normalization and the skull-stripping using SPM tissue probability maps and then proceeds to the Adaptive Maximum a Posterior (AMAP) segmentation (Rajapakse et al., 1997), which does not rely on any tissue

probabilities given *a priori*. CAT-12 segmentation procedure is then finalized by an estimation of the amount of each tissue in partial volume voxels (Tohka et al., 2004). FSL FAST, in its turn, relies on the hidden Markov random field (HMRF) model and the use of the expectation-maximization algorithm (Zhang et al., 2001) to estimate the parameters for the said model. The registration procedure also differs between CAT-12 (Ashburner, 2007) and FSL (Andersson et al., 2007) in terms of deformation models, similarity measures, and regularization methods (Klein et al., 2009). These significant differences in approaches lead to a mismatch between resulting segmentations and the magnitude of this mismatch will depend on various parameters such as age (Figure 12) or image quality.

On top of the differences already pointed out, the statistical analysis approaches also vary. While FSL VBM recommends the use of non-parametric statistics with threshold-free cluster enhancement, the default statistical method for CAT-12 is parametric, so it is reasonable to expect certain variance in sensitivity between the two tools. Additionally, CAT-12 insists on the correction for TIV to be included in the model, while FSL VBM does not provide TIV estimations, making it impossible to use it as a nuisance variable.

It is difficult to assess which aspect contributes the most to the noticeable difference in the amount of statistically significant voxels in VBM results, and it is out of the scope of this study to systematically compare the preprocessing steps. However, it is important to keep in mind that certain discrepancy in the VBM results is to be expected and always to be cautious with interpretations. To detect truly reliable differences in GM volumes, large-scale studies and meta-analyses are required.

5.4 VBM: age and gender effect on GM

FSL VBM detected sex-related differences located in the left auditory cortex, namely females having higher GM volume than males. While not directly supporting the VBM output, sexual dimorphism in the auditory cortex was previously shown in terms of volumetric measures in adults (de Lima Xavier et al., 2019; Ruigrok et al., 2014) as well as in children (Lombardo et al., 2012). This potentially indicates the gender-related differences in the processing of audio cues, which have been previously hypothesized and studied using functional neuroimaging modalities (Brun et al., 2009; Burman et al., 2008; Ruytjens et al., 2007). However, there are results that both support and contradict the idea of sexual dimorphism in the auditory cortex (Kaiser et al., 2009). Conducting a VBM study with a larger sample to replicate this result would be necessary to confirm the existence of gender effect.

Age-related GM growth has been shown in the frontal orbital cortex, temporal poles, precuneus cortex, putamina, precentral and superior frontal gyri, and right superior temporal gyrus by FSL VBM, while CAT-12 revealed statistically significant clusters in putamina and thalami.

While the previously published literature for this age is scarce, some of the findings may be supported by volumetric evidence in other closely related age groups. Volume growth in frontal lobe has been previously reported to last until approximately 6 years of age (Tanaka et al., 2013). Putamen in both hemispheres has been reported to undergo significant growth from birth to 2 years of age (Nishijima et al., 2016) and then to start to reduce in relative volume starting from age 4 (Sussman et al., 2016), which would indicate a volume peak somewhere between these time points. Age-related increases in thalamus volume have been shown in children from age 4 onward (Ball et al., 2012; Muftuler et al., 2011).

Curiously, FSL VBM showed an increase in GM volume in the primary motor cortex, while it is one of the areas that would be expected to mature early (Gogtay et al., 2004).This suggests that the age-related decline in GM volume in the motor cortex may start around 5 years of age.

FSL VBM and CAT-12 agreed on the decrease in GM volume with age in the occipital lobe, cuneal cortex, lingual gyri, medial frontal gyri. Additionally, FSL VBM showed clusters in the auditory cortex, caudate nucleai, hippocampi, and thalami.

Interestingly, CAT-12 has shown age-associated growth in the anterior parts of thalami, while FSL VBM revealed GM decline in the posterior regions. While it might indicate regional differences in development trajectories, the lack of overlap between the two methods' results in that area makes it seem very inconclusive.

Previously published VBM results in older children and young adults showed the prevalence of age-dependent GM loss in parietal areas (Guo et al., 2007, Mu et al., 2017, Wilke et al., 2007). Such a tendency can be seen in this study as well, with additional involvement of the occipital lobe, which goes in line with the functional development of the visuospatial processing in young children. Interestingly, one VBM study performed on a wide age group from 7 to 23 years contradicts with presented results considering hippocampi, which were shown to have a positive correlation with age (Guo et al., 2007).

In volumetric studies in older children age-dependent GM decrease has been previously reported in visual cortex and lingual gyri (Ball et al., 2012; Muftuler et al., 2011) as well as in caudate (Giedd et al., 1996), suggesting that maturation in these areas starts in early childhood.

5.5 Segmentation quality assessment

Use of Dice coefficient to estimate volume overlap between segmentations obtained from the utilization of different methods is a quite commonly used approach (Hosseini et al., 2014; Kazemi & Noorizadeh, 2014), especially within the rapidly growing field of neural networks-based segmentation algorithms (Duong et al., 2019; Moeskops et al., 2016; Tushar et al., 2019). However, the existing literature on pediatric segmentation mostly covers agreement between methodologies on subcortical structures segmentations (Herten et al., 2019; Lee et al., 2015; Loh et al., 2016; Schoemaker et al., 2016). Accuracy of brain tissue segmentations remains scarcely investigated in children.

In this study, segmentation quality assessment was done using manually edited FreeSurfer segmentations as a gold standard. The results for both CAT-12 and FSL were fairly similar for 5-year-olds in the case of GM and cortical GM segmentations with a mean value of DSC being 0.82 for CAT-12 and 0.81 for FSL. WM segmentation quality was relatively higher, with CAT-12 – 0.79 versus 0.75 with FSL. This indicates that in older children, the choice of the processing method can be arbitrary, as long as possible segmentation errors are considered in further analysis.

The agreement between segmentations of toddler scans was generally lower than in 5-year-olds. FSL segmentations showed better result than CAT-12: 0.74 versus 0.69 for GM and 0.65 versus 0.62 for WM. However, the DSCs are relatively low, and considering the need to provide externally brain-extracted images to FSL, FreeSurfer may be the best approach in the processing of toddler data.

While 80% overlap between FreeSurfer output and CAT-12 and FSL segmentations can be considered satisfactory in some cases, segmentation-sensitive analysis such as VBM would benefit greatly from more robust methods with a better agreement.

6. Conclusions

The current study has shown that both volumetric and VBM measures may vary significantly depending on the chosen method. While in the case of TIV or GM volume estimations high correlation between results may allow us to disregard the differences in absolute volumes and rather analyze development trajectories, we can hardly ignore the partial disagreement in VBM results. This highlights the importance of practicing caution in interpretation and performing extensive quality controls while dealing with pediatric data. It should be emphasized that there is still a strong need for the development of new preprocessing tools that would be more reliable in the case of young children.

References

Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., & Raznahan, A. (2016). Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. *Human Brain Mapping*, *37*(7), 2385–2397. https://doi.org/10.1002/hbm.23180

Andersson, J. L. R., Jenkinson, M., & Smith, S. (2007). Non-linear registration aka spatial normalisation. In *FMRIB Technical Report TRO7JA2*.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2007.07.007

Ashburner, J., & Friston, K. (2003). Morphometry. In *Human Brain Function: Second Edition*. https://doi.org/10.1016/B978-012264841-0/50038-X

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry - The methods. *NeuroImage*. https://doi.org/10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2005.02.018

Ávila, H., Raulino Silva, V., & Magalhães, D. S. F. (2019). Volume estimation of the brain, white matter, and gray matter using FreeSurfer and FSL: consistency between methods. *Research on Biomedical Engineering*, *35*(3–4), 257–263. https://doi.org/10.1007/s42600-019-00027-w

Ball, W. S., Byars, A. W., Schapiro, M., Bommer, W., Carr, A., German, A., Dunn, S., Rivkin, M. J., Waber, D., Mulkern, R., Vajapeyam, S., Chiverton, A., Davis, P., Koo, J., Marmor, J., Mrakotsky, C., Robertson, R., McAnulty, G., Brandt, M. E., … O'Neill, J. (2012). Total and regional brain volumes in a population-based normative sample from 4 to 18 years: The NIH MRI study of normal brain development. *Cerebral Cortex*, *22*(1), 1–12. https://doi.org/10.1093/cercor/bhr018

Batty, M. J., Liddle, E. B., Pitiot, A., Toro, R., Groom, M. J., Scerif, G., Liotti, M., Liddle, P. F., Paus, T., & Hollis, C. (2010). Cortical Gray Matter in Attention-Deficit/Hyperactivity Disorder: A Structural Magnetic Resonance Imaging Study. *Journal of the American*

Academy of Child & Adolescent Psychiatry, *49*(3), 229–238. https://doi.org/10.1016/j.jaac.2009.11.008

Bauer, C. C. C., Moreno, B., González-Santos, L., Concha, L., Barquera, S., & Barrios, F. A. (2015). Child overweight and obesity are associated with reduced executive cognitive performance and brain alterations: A magnetic resonance imaging study in Mexican children. *Pediatric Obesity*, *10*(3), 196–204. https://doi.org/10.1111/ijpo.241

Beauchamp, M. S., Beurlot, M. R., Fava, E., Nath, A. R., Parikh, N. A., Saad, Z. S., Bortfeld, H., & Oghalai, J. S. (2011). The developmental trajectory of brain-scalp distance from birth through childhood: Implications for functional neuroimaging. *PLoS ONE*, *6*(9), 1–9. https://doi.org/10.1371/journal.pone.0024981

Branson, H. M. (2013). Normal Myelination. A Practical Pictorial Review. *Neuroimaging Clinics of North America*, *23*(2), 183–195. https://doi.org/10.1016/j.nic.2012.12.001

Brown, T. T., Kuperman, J. M., Erhart, M., White, N. S., Roddey, J. C., Shankaranarayanan, A., Han, E. T., Rettmann, D., & Dale, A. M. (2010). Prospective motion correction of high-resolution magnetic resonance imaging data in children. *NeuroImage*, *53*(1), 139–145. https://doi.org/10.1016/j.neuroimage.2010.06.017

Brun, C. C., Leporé, N., Luders, E., Chou, Y. Y., Madsen, S. K., Toga, A. W., & Thompson, P. M. (2009). Sex differences in brain structure in auditory and cingulate regions. *NeuroReport*, *20*(10), 930–935. https://doi.org/10.1097/WNR.0b013e32832c5e65

Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, *23*(2), 724–738. https://doi.org/10.1016/j.neuroimage.2004.06.018

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia*, *46*(5), 1349–1362. https://doi.org/10.1016/j.neuropsychologia.2007.12.021

Callaert, D. v., Ribbens, A., Maes, F., Swinnen, S. P., & Wenderoth, N. (2014). Assessing age-related gray matter decline with voxel-based morphometry depends significantly on

segmentation and normalization procedures. *Frontiers in Aging Neuroscience*, *6*(JUN), 1–14. https://doi.org/10.3389/fnagi.2014.00124

Carducci, F., Onorati, P., Condoluci, C., di Gennaro, G., Quarato, P. P., Pierallini, A., Sarà, M., Miano, S., Cornia, R., & Albertini, G. (2013). Whole-brain voxel-based morphometry study of children and adolescents with Down syndrome. *Functional Neurology*, *28*(1), 19–28. https://doi.org/10.11138/FNeur/2013.28.1.019

Cox J.S., R. W. ; H. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(29), 162–173. https://ac-els-cdn-com.ezp-prod1.hul.harvard.edu/S0010480996900142/1-s2.0-S0010480996900142-main.pdf?_tid=c22bae7a-b8f5-4a8b-9dac-91c1ed841d53&acdnat=1549393398_e37181b8933a2ac88c2d7dc0eab14413

de Lima Xavier, L., Hanekamp, S., & Simonyan, K. (2019). Sexual Dimorphism Within Brain Regions Controlling Speech Production. *Frontiers in Neuroscience*, *13*(July), 1–7. https://doi.org/10.3389/fnins.2019.00795

Dean, D. C., Planalp, E. M., Wooten, W., Schmidt, C. K., Kecskemeti, S. R., Frye, C., Schmidt, N. L., Goldsmith, H. H., Alexander, A. L., & Davidson, R. J. (2018). Investigation of brain structure in the 1-month infant. *Brain Structure and Function*. https://doi.org/10.1007/s00429-017-1600-2

DelBello, M. P., Zimmerman, M. E., Mills, N. P., Getz, G. E., & Strakowski, S. M. (2004). Magnetic resonance imaging analysis of amygdala and other subcortical brain regions in adolescents with bipolar disorder. *Bipolar Disorders*, *6*(1), 43–52. https://doi.org/10.1046/j.1399-5618.2003.00087.x

Despotović, I., Goossens, B., & Philips, W. (2015). MRI segmentation of the human brain: Challenges, methods, and applications. *Computational and Mathematical Methods in Medicine*, *2015*. https://doi.org/10.1155/2015/450341

Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*. https://doi.org/10.2307/1932409

Duong, M. T., Rudie, J. D., Wang, J., Xie, L., Mohan, S., Gee, J. C., & Rauschecker, A. M. (2019). Convolutional neural network for automated flair lesion segmentation on clinical

brain MR imaging. *American Journal of Neuroradiology*, *40*(8), 1282–1290. https://doi.org/10.3174/ajnr.A6138

Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, *54*(1), 313–327. https://doi.org/10.1016/j.neuroimage.2010.07.033

Gaser, C., & Dahnke, R. (2016). CAT-A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. *Human Brain Mapping*, *32*(7), 336–348.

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., Hakonarson, H., Gur, R. E., & Gur, R. C. (2017). Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *Journal of Neuroscience*, *37*(20), 5065–5073. https://doi.org/10.1523/JNEUROSCI.3550-16.2017

Ghosh, S. S., Kakunoori, S., Augustinack, J., Nieto-Castanon, A., Kovelman, I., Gaab, N., Christodoulou, J. A., Triantafyllou, C., Gabrieli, J. D. E., & Fischl, B. (2010). Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11years of age. *NeuroImage*, *53*(1), 85–93. https://doi.org/10.1016/j.neuroimage.2010.05.075

Giedd, J. N., Raznahan, A., Mills, K. L., & Lenroot, R. K. (2012). Review: Magnetic resonance imaging of male/female differences in human adolescent brain anatomy. *Biology of Sex Differences*, *3*(1), 1–9. https://doi.org/10.1186/2042-6410-3-19

Giedd, J. N., Snell, J. W., Lange, N., Rajapakse, J. C., Casey, B. J., Kozuch, P. L., Vaituzis, A. C., Vauss, Y. C., Hamburger, S. D., Kaysen, D., & Rapoport, J. L. (1996). Quantitative Magnetic Resonance Imaging of Human Brain Development: Ages 4–18. *Cerebral Cortex*, *6*(4), 551–559. https://doi.org/10.1093/cercor/6.4.551

Gilmore, J. H., Lin, W., Prastawa, M. W., Looney, C. B., Vetsa, Y. S. K., Knickmeyer, R. C., Evans, D. D., Smith, J. K., Hamer, R. M., Lieberman, J. A., & Gerig, G. (2007). Regional gray matter growth, sexual dimorphism, and cerebral asymmetry in the neonatal brain. *Journal of Neuroscience*, *27*(6), 1255–1260. https://doi.org/10.1523/JNEUROSCI.3339-06.2007

Gilmore, J. H., Shi, F., Woolson, S. L., Knickmeyer, R. C., Short, S. J., Lin, W., Zhu, H., Hamer, R. M., Styner, M., & Shen, D. (2012). Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cerebral Cortex*, *22*(11), 2478–2485. https://doi.org/10.1093/cercor/bhr327

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(21), 8174–8179. https://doi.org/10.1073/pnas.0402680101

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, *14*(1 I), 21–36. https://doi.org/10.1006/nimg.2001.0786

Guo, X., Chen, C., Chen, K., Jin, Z., Peng, D., & Yao, L. (2007). Brain development in Chinese children and adolescents: A structural MRI study. *NeuroReport*, *18*(9), 875–880. https://doi.org/10.1097/WNR.0b013e328152777e

Herten, A., Konrad, K., Krinzinger, H., Seitz, J., & von Polier, G. G. (2019). Accuracy and bias of automatic hippocampal segmentation in children and adolescents. *Brain Structure and Function*, *224*(2), 795–810. https://doi.org/10.1007/s00429-018-1802-2

Hosseini, M. P., Nazem-Zadeh, M. R., Pompili, D., & Soltanian-Zadeh, H. (2014). Statistical validation of automatic methods for hippocampus segmentation in MR images of epileptic patients. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, 4707–4710. https://doi.org/10.1109/EMBC.2014.6944675

Hu, S., Pruessner, J. C., Coupé, P., & Collins, D. L. (2013). Volumetric analysis of medial temporal lobe structures in brain development from childhood to adolescence. *NeuroImage*, *74*, 276–287. https://doi.org/10.1016/j.neuroimage.2013.02.032

Iglesias, J. E., Liu, C. Y., Thompson, P. M., & Tu, Z. (2011). Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging*, *30*(9), 1617–1634. https://doi.org/10.1109/TMI.2011.2138152

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). Review FSL. *NeuroImage*, *62*, 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jones, J. E., Jackson, D. C., Chambers, K. L., Dabbs, K., Hsu, D. A., Stafstrom, C. E., Seidenberg, M., & Hermann, B. P. (2015). Children with epilepsy and anxiety: Subcortical and cortical differences. *Epilepsia*. https://doi.org/10.1111/epi.12832

Kaiser, A., Haller, S., Schmitz, S., & Nitsch, C. (2009). On sex/gender related similarities and differences in fMRI language research. *Brain Research Reviews*, *61*(2), 49–59. https://doi.org/10.1016/j.brainresrev.2009.03.005

Karlsson, L., Tolvanen, M., Scheinin, N. M., Uusitupa, H. M., Korja, R., Ekholm, E., Tuulari, J. J., Pajulo, M., Huotilainen, M., Paunio, T., & Karlsson, H. (2018). Cohort Profile: The FinnBrain Birth Cohort Study (FinnBrain). *International Journal of Epidemiology*, *47*(1), 15–16j. https://doi.org/10.1093/ije/dyx173

Kazemi, K., & Noorizadeh, N. (2014). Quantitative Comparison of SPM, FSL, and Brainsuite for Brain MR Image Segmentation. *Journal of Biomedical Physics & Engineering*, *4*(1), 13–26. http://www.ncbi.nlm.nih.gov/pubmed/25505764%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4258855

Kennedy, J. T., Collins, P. F., & Luciana, M. (2016). Higher adolescent body mass index is associated with lower regional gray and white matter volumes and lower levels of positive emotionality. *Frontiers in Neuroscience*. https://doi.org/10.3389/fnins.2016.00413

Klein, A., , Jesper Andersson, Babak A. Ardekani, John Ashburner, B. A., Ming-Chang Chiang, Gary E. Christensen, D. L. C., Gee, J., Hellier, P., Joo Hyun Song, Mark Jenkinson, C. L., , Daniel Rueckert, P. T., & Tom Vercauteren, Roger P. Woods, J. John Mann, and R. V. P. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage*, *46*(3), 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037.Evaluation

Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., Hamer, R. M., Lin, W., Gerig, G., & Gilmore, J. H. (2008). A Structural MRI Study of Human Brain Development from Birth to 2 Years. *Journal of Neuroscience*, *28*(47), 12176–12182. https://doi.org/10.1523/JNEUROSCI.3479-08.2008

Knickmeyer, Rebecca C., Wang, J., Zhu, H., Geng, X., Woolson, S., Hamer, R. M., Konneker, T., Lin, W., Styner, M., & Gilmore, J. H. (2014). Common variants in psychiatric risk genes predict brain structure at birth. *Cerebral Cortex*, *24*(5), 1230–1246. https://doi.org/10.1093/cercor/bhs401

Knickmeyer, Rebecca C., Wang, J., Zhu, H., Geng, X., Woolson, S., Hamer, R. M., Konneker, T., Styner, M., & Gilmore, J. H. (2014). Impact of sex and gonadal steroids on neonatal brain structure. *Cerebral Cortex*, *24*(10), 2721–2731. https://doi.org/10.1093/cercor/bht125

Lee, J. K., Nordahl, C. W., Amaral, D. G., Lee, A., Solomon, M., & Ghetti, S. (2015). Assessing hippocampal development and language in early childhood: Evidence from a new application of the Automatic Segmentation Adapter Tool. *Human Brain Mapping*, *36*(11), 4483–4496. https://doi.org/10.1002/hbm.22931

Loh, W. Y., Connelly, A., Cheong, J. L. Y., Spittle, A. J., Chen, J., Adamson, C., Ahmadzai, Z. M., Fam, L. G., Rees, S., Lee, K. J., Doyle, L. W., Anderson, P. J., & Thompson, D. K. (2016). A New MRI-Based Pediatric Subcortical Segmentation Technique (PSST). *Neuroinformatics*, *14*(1), 69–81. https://doi.org/10.1007/s12021-015-9279-0

Lombardo, M. v., Ashwin, E., Auyeung, B., Chakrabarti, B., Taylor, K., Hackett, G., Bullmore, E. T., & Baron-Cohen, S. (2012). Fetal testosterone influences sexually dimorphic gray matter in the human brain. *Journal of Neuroscience*, *32*(2), 674–680. https://doi.org/10.1523/JNEUROSCI.4389-11.2012

Lorio, S., Fresard, S., Adaszewski, S., Kherif, F., Chowdhury, R., Frackowiak, R. S., Ashburner, J., Helms, G., Weiskopf, N., Lutti, A., & Draganski, B. (2016). New tissue priors for improved automated classification of subcortical brain structures on MRI. *NeuroImage*, *130*, 157–166. https://doi.org/10.1016/j.neuroimage.2016.01.062

Mayer, K. N., Latal, B., Knirsch, W., Scheer, I., von Rhein, M., Reich, B., Bauer, J., Gummel, K., Roberts, N., & Tuura, R. O. (2016). Comparison of automated brain volumetry methods with stereology in children aged 2 to 3 years. *Neuroradiology*, *58*(9), 901–910. https://doi.org/10.1007/s00234-016-1714-x

Mazaika, P. K., Weinzimer, S. A., Mauras, N., Buckingham, B., White, N. H., Tsalikian, E., Hershey, T., Cato, A., Aye, T., Fox, L., Wilson, D. M., Tansey, M. J., Tamborlane, W., Peng, D., Raman, M., Marzelli, M., Reiss, A. L., Coffey, J., Cabbage, J., … Moran, A.

(2016). Variations in brain volume and growth in young children with type 1 diabetes. *Diabetes*, *65*(2), 476–485. https://doi.org/10.2337/db15-1242

McAlonan, G. M., Cheung, V., Cheung, C., Suckling, J., Lam, G. Y., Tai, K. S., Yip, L., Murphy, D. G. M., & Chua, S. E. (2005). Mapping the brain in autism. A voxel-based MRI study of volumetric differences and intercorrelations in autism. *Brain*, *128*(2), 268–276. https://doi.org/10.1093/brain/awh332

Mechelli, A., Price, C., Friston, K., & Ashburner, J. (2005). Voxel-Based Morphometry of the Human Brain: Methods and Applications. *Current Medical Imaging Reviews*. https://doi.org/10.2174/1573405054038726

Merz, E. C., He, X., & Noble, K. G. (2018). Anxiety, depression, impulsivity, and brain structure in children and adolescents. *NeuroImage: Clinical*, *20*(April), 243–251. https://doi.org/10.1016/j.nicl.2018.07.020

Moeskops, P., Viergever, M. A., Mendrik, A. M., de Vries, L. S., Benders, M. J. N. L., & Isgum, I. (2016). Automatic Segmentation of MR Brain Images with a Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, *35*(5), 1252–1261. https://doi.org/10.1109/TMI.2016.2548501

Mu, S. H., Xu, M., Duan, J. X., Zhang, J., & Tan, L. H. (2017). Localizing Age-Related Changes in Brain Structure Using Voxel-Based Morphometry. *Neural Plasticity*, *2017*. https://doi.org/10.1155/2017/6303512

Muftuler, L. T., Davis, E. P., Buss, C., Head, K., Hasso, A. N., & Sandman, C. A. (2011). Cortical and subcortical changes in typically developing preadolescent children. *Brain Research*. https://doi.org/10.1016/j.brainres.2011.05.018

Murgasova, M., Dyet, L., Edwards, D., Rutherford, M., Hajnal, J. v., & Rueckert, D. (2006). Segmentation of brain MRI in young children. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/11866565_84

Nishijima, Daniel; K. Simel, David L; Wisner, David H; Holmes, J. F. (2016). Putamen Development in Children 12 to 21 Months Old. *Physiology & Behavior*, *176*(1), 139–148. https://doi.org/10.1016/j.physbeh.2017.03.040

Nordahl, C. W., Lange, N., Li, D. D., Barnett, L. A., Lee, A., Buonocore, M. H., Simon, T. J., Rogers, S., Ozonoff, S., & Amaral, D. G. (2011). Brain enlargement is associated with regression in preschool-age boys with autism spectrum disorders. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 20195–20200. https://doi.org/10.1073/pnas.1107560108

Okada, N., Yahata, N., Koshiyama, D., Morita, K., Sawada, K., Kanata, S., Fujikawa, S., Sugimoto, N., Toriyama, R., Masaoka, M., Koike, S., Araki, T., Kano, Y., Endo, K., Yamasaki, S., Ando, S., Nishida, A., Hiraiwa-Hasegawa, M., & Kasai, K. (2018). Abnormal asymmetries in subcortical brain volume in early adolescents with subclinical psychotic experiences. *Translational Psychiatry*, *8*(1). https://doi.org/10.1038/s41398-018-0312-6

Ou, X., Andres, A., Pivik, R. T., Cleves, M. A., Snow, J. H., Ding, Z., & Badger, T. M. (2016). Voxel-based morphometry and fMRI revealed differences in brain gray matter in breastfed and milk formula-fed children. *American Journal of Neuroradiology*, *37*(4), 713–719. https://doi.org/10.3174/ajnr.A4593

Paus, T. (2010). Sex differences in the human brain. A developmental perspective. *Progress in Brain Research*, *186*(C), 13–28. https://doi.org/10.1016/B978-0-444-53630-3.00002-6

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. https://doi.org/10.1038/nrn2513

Perlaki, G., Molnar, D., Smeets, P. A. M., Ahrens, W., Wolters, M., Eiben, G., Lissner, L., Erhard, P., van Meer, F., Herrmann, M., Janszky, J., & Orsi, G. (2018). Volumetric gray matter measures of amygdala and accumbens in childhood overweight/obesity. *PLoS ONE*, *13*(10), 1–17. https://doi.org/10.1371/journal.pone.0205331

Peterson, B. S., Anderson, A. W., Ehrenkranz, R., Staib, L. H., Tageldin, M., Colson, E., Gore, J. C., Duncan, C. C., Makuch, R., & Ment, L. R. (2003). Regional brain volumes and their later neurodevelopmental correlates in term and preterm infants. *Pediatrics*. https://doi.org/10.1542/peds.111.5.939

Phan, T. V., Smeets, D., Talcott, J. B., & Vandermosten, M. (2018). Processing of structural neuroimaging data in young children: Bridging the gap between current practice and

state-of-the-art methods. *Developmental Cognitive Neuroscience*, *33*(July 2017), 206–223. https://doi.org/10.1016/j.dcn.2017.08.009

Popescu, V., Schoonheim, M. M., Versteeg, A., Chaturvedi, N., Jonker, M., de Menezes, R. X., Garre, F. G., Uitdehaag, B. M. J., Barkhof, F., & Vrenken, H. (2016). Grey matter atrophy in multiple sclerosis: Clinical interpretation depends on choice of analysis method. *PLoS ONE*, *11*(1), 1–17. https://doi.org/10.1371/journal.pone.0143942

Pulli E.P., Tuulari J.J. (2020). Manuscript in preparation. Faculty of Medicine, University of Turku, Turku, Finland.

Rajagopalan, V., Yue, G. H., & Pioro, E. P. (2014). Do preprocessing algorithms and statistical models influence voxel-based morphometry (VBM) results in amyotrophic lateral sclerosis patients? A systematic comparison of popular VBM analytical methods. *Journal of Magnetic Resonance Imaging*, *40*(3), 662–667. https://doi.org/10.1002/jmri.24415

Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral mr images. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/42.563663

Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & Scotto di Covella, L. (2018). Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience and Biobehavioral Reviews*, *84*(November 2016), 434–452. https://doi.org/10.1016/j.neubiorev.2017.08.001

Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., & Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage*, *107*, 107–115. https://doi.org/10.1016/j.neuroimage.2014.12.006

Ruigrok, A. N. V., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. v., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience and Biobehavioral Reviews*, *39*, 34–50. https://doi.org/10.1016/j.neubiorev.2013.12.004

Ruytjens, L., Georgiadis, J. R., Holstege, G., Wit, H. P., Albers, F. W. J., & Willemsen, A. T. M. (2007). Functional sex differences in human primary auditory cortex. *European Journal of Nuclear Medicine and Molecular Imaging*, *34*(12), 2073–2081. https://doi.org/10.1007/s00259-007-0517-z

Saunders, D. E., Thompson, C., Gunny, R., Jones, R., Cox, T., & Chong, W. K. (2007). Magnetic resonance imaging protocols for paediatric neuroradiology. *Pediatric Radiology*, *37*(8), 789–797. https://doi.org/10.1007/s00247-007-0462-9

Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., Gauthier, S., & Pruessner, J. C. (2016). Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation. *NeuroImage*, *129*, 1–14. https://doi.org/10.1016/j.neuroimage.2016.01.038

Schoemaker, D., Buss, C., Head, K., Sandman, C. A., Davis, E. P., Chakravarty, M. M., Gauthier, S., & Pruessner, J. C. (2018). Corrigendum to "Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation"[NeuroImage 129 (2016) 1–14] (S1053811916000537) (10.1016/j.neuroimage.2016.01.038)). *NeuroImage*, *173*(February), 1–2. https://doi.org/10.1016/j.neuroimage.2018.02.009

Shi, F., Wang, L., Dai, Y., Gilmore, J. H., Lin, W., & Shen, D. (2012). LABEL: Pediatric brain extraction using learning-based meta-algorithm. *NeuroImage*, *62*(3), 1975–1986. https://doi.org/10.1016/j.neuroimage.2012.05.042

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P. M., Federico, A., & de Stefano, N. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, *17*(1), 479–489. https://doi.org/10.1006/nimg.2002.1040

Sussman, D., Leung, R. C., Chakravarty, M. M., Lerch, J. P., & Taylor, M. J. (2016). Developing human brain: Age-related changes in cortical, subcortical, and cerebellar anatomy. *Brain and Behavior*, *6*(4), 1–15. https://doi.org/10.1002/brb3.457

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging*, *15*(1). https://doi.org/10.1186/s12880-015-0068-x

Tanaka, C., Matsui, M., Uematsu, A., Noguchi, K., & Miyawaki, T. (2013). Developmental trajectories of the fronto-temporal lobes from infancy to early adulthood in healthy individuals. *Developmental Neuroscience*, *34*(6), 477–487. https://doi.org/10.1159/000345152

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2004.05.007

Tondelli, M., Pizza, F., Vaudano, A. E., Plazzi, G., & Meletti, S. (2018). Cortical and subcortical brain changes in children and adolescents with narcolepsy type 1. *Sleep*, *41*(2), 1–7. https://doi.org/10.1093/sleep/zsx192

Tushar, F. I., Alyafi, B., Hasan, M. K., & Dahal, L. (2019). Brain tissue segmentation using neuronet with different pre-processing techniques. *2019 Joint 8th International Conference on Informatics, Electronics and Vision, ICIEV 2019 and 3rd International Conference on Imaging, Vision and Pattern Recognition, IcIVPR 2019 with International Conference on Activity and Behavior Computing, ABC 2019*, *I*, 223–227. https://doi.org/10.1109/ICIEV.2019.8858515

Vijayakumar, N., Mills, K. L., Alexander-Bloch, A., Tamnes, C. K., & Whittle, S. (2018). Structural brain development: A review of methodological approaches and best practices. *Developmental Cognitive Neuroscience*, *33*(November 2017), 129–148. https://doi.org/10.1016/j.dcn.2017.11.008

Villagran, A., Woermann, F., & Pohlmann-Eden, B. (2013). Neuroanatomical abnormalities in juvenile myoclonic epilepsy: A voxel-based morphometry study. *Epilepsy Currents*.

Vogel, S. E., Matejko, A. A., & Ansari, D. (2016). Imaging the developing human brain using functional and structural Magnetic Resonance Imaging: Methodological and practical guidelines. *Practical Research with Children (Research Methods in Developmental Psychology: A Handbook Series)*, *September*, 334.

Wilke, M., Holland, S. K., & Krägeloh-Mann, I. (2007). Global, regional, and local development of gray and white matter volume in normal children. *Experimental Brain Research*, *178*(3), 296–307.

Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2014.01.060

Xavier Castellanos, F., Lee, P. P., Sharp, W., Jeffries, N. O., Greenstein, D. K., Clasen, L. S., Blumenthal, J. D., James, R. S., Ebens, C. L., Walter, J. M., Zijdenbos, A., Evans, A. C., Giedd, J. N., & Rapoport, J. L. (2002). Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. *Journal of the American Medical Association*, *288*(14), 1740–1748. https://doi.org/10.1001/jama.288.14.1740

Xiao, Z., Qiu, T., Ke, X., Xiao, X., Xiao, T., Liang, F., Zou, B., Huang, H., Fang, H., Chu, K., Zhang, J., & Liu, Y. (2014). Autism spectrum disorder as early neurodevelopmental disorder: Evidence from the brain imaging abnormalities in 2-3 years old toddlers. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-014-2033-x

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1109/42.906424