TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

MAPPING CORTICAL BRAIN DEVELOPMENT, EARLY LIFE EXPOSURES AND NEURAL CORRELATES OF LANGUAGE SKILLS: FINDINGS FROM THE FINNBRAIN BIRTH COHORT STUDY

Eero Silver

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1736 | MEDICA – ODONTOLOGICA | TURKU 2023

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

MAPPING CORTICAL BRAIN DEVELOPMENT, EARLY LIFE EXPOSURES AND NEURAL CORRELATES OF LANGUAGE SKILLS: FINDINGS FROM THE FINNBRAIN BIRTH COHORT STUDY

Eero Silver

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1736 | MEDICA – ODONTOLOGICA | TURKU 2023

University of Turku

Faculty of Medicine
Department of Clinical Medicine
Psychiatry
Doctoral Programme in Clinical Research

Supervised by

Associate Professor Jetro Tuulari
Department of Psychiatry
University of Turku
Turku, Finland

Professor Hasse Karlsson
Department of Psychiatry
University of Turku
Turku, Finland

Reviewed by

Associate Professor Juha Salmitaival
Department of Psychology
University of Helsinki
Helsinki, Finland

Associate Professor Tuula Hurtig
Department of Child Psychiatry
University of Oulu
Oulu, Finland

Opponent

Professor Leena Haataja
Department of Child Neurology
University of Helsinki
Helsinki, Finland

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

ISBN 978-951-29-9423-6 (PRINT)
ISBN 978-951-29-9424-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2023

To my family and friends

UNIVERSITY OF TURKU
Faculty of Medicine
Department of Clinical Medicine, Psychiatry
EERO SILVER: Mapping cortical brain development, early life exposures and neural correlates of language skills: findings from the FinnBrain Birth Cohort Study
Doctoral Dissertation, 153 pp.
Doctoral Programme in Clinical Research
August 2023

ABSTRACT

The human brain starts to develop *in utero* during the first trimester of pregnancy, and rapid structural growth continues postnatally during the first years of life. This suggests that early childhood is a critical period for various neurodevelopmental disorders, but also an ideal point in time for interventions to have a great positive effect on later neurodevelopment. Recent evidence suggests that many environmental factors and parental health profiles are associated with a child's brain development already in the prenatal period. Despite increasing knowledge about structural brain development collected from magnetic resonance imaging (MRI) studies, there are still significant gaps in knowledge about the structural mechanisms and timing of the brain's maturation during the first five years of life.

In this thesis, we assess modern segmentation tools and present the semiautomated segmentation protocol that we have applied in the FinnBrain Birth Cohort Study. We also study early cortical brain development and how it is associated with changing perinatal factors and environmental variables at five years of age. Finally, we review the latest findings of language-related structural MRI studies carried out with children aged from birth to two years.

While presenting our segmentation protocol, we discovered that with appropriate quality assessments, the FreeSurfer segmentation tool can be reliably used to analyse cortical metrics of T1-weighted images of the five-year-olds. We found that sex, obstetric variables and maternal perinatal health profile were associated with cortical development in five-year-olds. Different environmental variables modulated cortical development in five-year-olds in comparison with infant MRI studies, carried out with similar analyses. During the first two years of life, the implicated cerebral areas included many established components of the language networks; including frontal and temporal regions. These regions associated with language skills.

This thesis confirms that early exposures affect brain development through early childhood, and we have created a segmentation protocol that can be implemented with the structural MRI data. Language development is associated with school performance, and this thesis lays the foundations for future studies that will continue during children of school age also in the FinnBrain Birth Cohort Study.

KEYWORDS: brain, children, structural development, early development, language development, segmentation tool.

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Kliininen laitos, Psykiatria
EERO SILVER: Varhaiset altisteet, lapsen rakenteellinen aivojen kehitys ja kielelliset taidot: kuvantamistutkimus FinnBrain syntymäkohorttitutkimuksessa
Väitöskirja, 153 s.
Turun kliininen tohtoriohjelma
Elokuu 2023

TIIVISTELMÄ

Ihmisaivot alkavat kehittyä kohdussa jo äidin ensimmäisen raskauskolmanneksen aikana, ja aivojen rakenteellinen kehitys jatkuu vauhdikkaana syntymän jälkeen ensimmäisen kahden vuoden ajan. Varhaislapsuus on keskeinen ajanjakso neurokognitiivisten häiriöiden synnylle, ja siten myös myöhempää neurokognitiivista kehitystä tukevien hoitomuotojen tarjoamiselle. Lisäksi monet ympäristötekijät ja vanhempien hyvinvointi vaikuttavat lapsen aivojen kehitykseen jo ennen syntymää. Vaikka magneettikuvantamistutkimusten ansiosta tiedämme aivojen varhaisvaiheen rakenteellisesta kehityksestä jo suhteellisen paljon, on tietämys aivojen kehityksestä erityisesti viiden ensimmäisen ikävuoden ajalta vielä pitkälti puutteellista. Tutkimustyömme on osa FinnBrain syntymäkohorttitutkimusta. Arvioimme yleisesti aivokuvantamisessa käytettyjä segmentaatiotyökaluja ja esittelemme laadunvarmistusprotokollan, jonka mukaan olemme käsitelleet keräämämme 5-vuotiaiden magneettikuvantamisaineiston. Toiseksi tutkimme, kuinka raskausajan muuttuvat olosuhteet sekä varhaisiän ympäristötekijät vaikuttavat aivojen rakenteelliseen kortikaaliseen kehitykseen viiden vuoden iässä. Kolmannen osatyön systemaattisessa katsausartikkelissa keräämme yhteen tuoreimmat tulokset kielen kehitystä tarkastelevista kuvantamistutkimuksista ensimmäisen kahden ikävuoden ajalta.

Asianmukaisella laaduntarkkailulla FreeSurfer-segmentaatiotyökalu soveltuu 5-vuotiaiden magneettikuvantamisaineiston käsittelyyn hyvin. Toiseksi havaitsimme, että sukupuolella, äidin terveydellä sekä synnytykseen liittyvillä muuttujilla havaittiin yhteyksiä aivojen kortikaalirakenteiden kehitykseen 5-vuotiailla. Aikuisilla havaitut perinteisesti kielellisistä toiminnoista vastaavat aivoalueet, kuten ohimo- ja otsalohkon alueet, olivat havaittavissa ja yhdistettävissä kielellisiin taitoihin jo alle kahden vuoden iässä. Tämä väitöskirja vahvistaa, että varhaiset altisteet vaikuttavat aivojen kehitykseen yli varhaislapsuuden sekä sen, että olemme luoneet rakenteellisen MRI-aineistoon soveltuvat käytänteet. Kielenkehitys liittyy muun muassa koulumenestykseen ja väitöskirja luo vahvan pohjan tulevalle tutkimukselle, joka tulee jatkumaan kouluiässä FinnBrain-tutkimuksessa.

AVAINSANAT: aivot, lapset, rakenteellinen kehitys, varhainen kehitys, kielellinen kehitys, segmentaatiotyökalu

Table of Contents

Abbreviations

BMI	Body mass index
CNS	Central nervous system
CT	Cortical thickness
DTI	Diffusion tensor imaging
ENIGMA	Enhancing neuroimaging genetics through meta-analysis
GM	Grey matter
IQ	Intelligence quotient
MRI	Magnetic resonance imaging
MSEL	The Mullen scales of early learning
PMC	Prospective motion correction
ROI	Region of interest
SA	Surface area
SES	Socioeconomical status
SNRI	Serotonin and norepinephrine reuptake inhibitors
SSRI	Selective serotonin reuptake inhibitor
SSS	Superior sagittal sinus
WM	White matter

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I. Pulli, E. P., Silver, E., Kumpulainen, V., Copeland, A., Merisaari, H., Saunavaara, J., Parkkola, R., Lähdesmäki, T., Saukko, E., Nolvi, S., Kataja, E-L., Korja, R., Karlsson, L., Karlsson, H., Tuulari, J. Feasibility of FreeSurfer processing for T1-weighted brain images of 5-year- olds: semiautomated protocol of FinnBrain Neuroimaging Lab. *Frontiers in neuroscience*, 2022, vol 16.

II. Silver, E., Pulli, E. P., Kataja, E. L., Kumpulainen, V., Copeland, A., Saukko, E., Saunavaara, J., Merisaari, H., Lähdesmäki, T., Parkkola, R., Karlsson, L., Karlsson, H., & Tuulari, J. J. Prenatal and early-life environmental factors, family demographics and cortical brain anatomy in 5-year-olds: an MRI study from FinnBrain Birth Cohort. *Brain Imaging and Behavior*, 2022, 16(5), 2097–2109.

III. Silver, E.; Korja, R.; Mainela-Arnold, E.; Pulli, E.P.; Saukko, E.; Nolvi, S.; Kataja, E.L.; Karlsson, L.; Karlsson, H.; Tuulari, J.J. A systematic review of MRI studies of language development from birth to 2 years of age. *Developmental Neurobiology*, 2021, 81(1), 63–75

The original publications have been reproduced with the permission of the copyright holders.

1 Introduction

Magnetic resonance imaging (MRI) has given rise to new possibilities in the study of human brain development during its most dynamic phases of growth and development – from infancy to early adulthood. Even though the number of MRI studies carried out on children has risen remarkably during recent decades, especially prospective neuroimaging studies during infancy to early childhood, when the fast brain growth coincides with various developmental milestones, are lacking (Knickmeyer et al., 2008). The youngest ages pose significant challenges for performing the MRI measurements that necessitate good compliance during the scans (Copeland et al., 2021). It then follows that there is a relatively small amount of data available on the essential periods of brain development, and although many large-scale efforts to fill this gap are ongoing, there is a clear need for foundational studies in developmental cognitive neuroscience of early childhood.

Paediatric neuroimaging is a quickly developing field of study, which potentially leads to methodological issues. For instance, the commonly used automated segmentation tools, which are essential for reasonable and time-efficient research work with MRI data, are often based on templates of adults and older children (Ghosh et al., 2010; Thompson et al., 2020). As a result, this can cause errors when analysing paediatric imaging data. Another key challenge is the variation of atlases and automated segmentation tools and, further, the lack of "gold standard" quality control protocols, which complicates comparisons between studies.

The human brain develops dynamically during infancy and throughout early childhood, a time when a child is sensitive to both genetic programming and extrinsic exposures. Current research suggests that prenatal and early life environmental factors and family demographics could predict the cortical brain morphology in infants (Cattarinussi et al., 2021; Knickmeyer et al., 2017). Associations have been found in later ages too, but often they differ from infant studies (Piccolo et al., 2016). In addition, there is evidence that exposure to many prenatal factors alter infant brain development even before birth via the mother (Choe et al., 2013; Pulli et al., 2018; Qiu et al., 2013).

Relatively little is known about how brain maturation is associated with the remarkable learning that happens during the first years of life. Learning to produce and comprehend spoken language provides an interesting example of general cognitive development, and an opportunity to capture the aspects of brain growth and acquired skills that are fundamental and unique to humans (Gaudet et al., 2020; Perani et al., 2011). Neurocognitive functions related to language development start to develop long before the first words are produced (Shultz et al., 2014). After the first two years of life, cognitive development is based mostly on maturation and re-organization of white matter (WM) trajectories connecting the specific brain regions (Deoni et al., 2015; O'Muircheartaigh et al., 2014). By which mechanisms these developmental trajectories are associated with a child's cognitive performance and according to what timeframe it will be expressed are still largely unknown.

This thesis focuses on early life structural brain development with an interest to optimize methods, assess perinatal and early life factors that influence cortical brain development, and pave the way to MRI studies of early language development by conducting a systematic review.

2 Review of the Literature

2.1 Structural development of the brain

Postnatal human brain development begins with a rapid increase of the total brain volume, and by the age of two years, the human brain has reached nearly 90 percent of its final adult volume (Knickmeyer et al., 2008). This demonstrates that the first years of life are a critical period for brain development, when many neurogenic disorders are likely to originate. On the other hand, it is also a period when early diagnostics and further, possible therapeutic interventions could have significant effects on later brain development. As there are a lot of quality MRI studies carried out with preterm or otherwise developmentally impaired children (Li et al., 2016; Shaw et al., 2012; Tusor et al., 2014; Xiao et al., 2014; Yang et al., 2016) they are not within scope of this work and, hence, pathologies will not be discussed further in this thesis. Instead, understanding the normal structural development of the human brain is essential for studying the effects of psychiatric disorders, neurocognitive impairments or other early life adverse outcomes on brain maturation. The structural growth curves of the brain are presented in Figure 1 (Bethlehem et al., 2022).

2.1.1 Grey matter development

Rapid early life cortical growth is reflected in all typically used metrics, such as cortical thickness (CT), surface area (SA), gyrification and cortical volumes (Remer et al., 2017). These metrics reflect distinct patterns of brain maturation and are frequently discussed separately in brain imaging studies (Kuhl et al., 2020; Lyall et al., 2015; Yang et al., 2016).

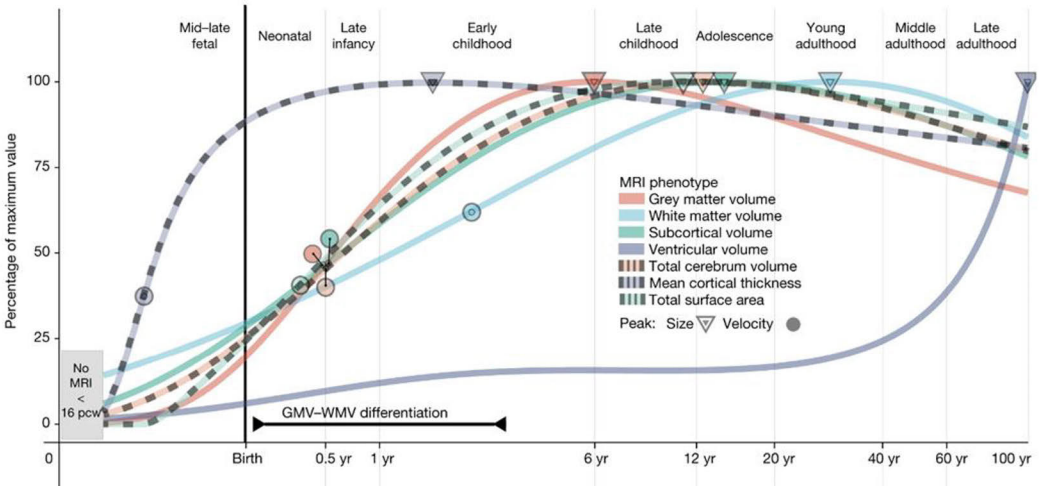

Figure 1. Brain charts of the human lifespan. The figure was originally published in the article by Bethlehem et al., (2022), which is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

The development of CT features both thickening and thinning during early life. Despite numerous child MRI studies, there is still discussion on whether the trend is cortical thickening or thinning during early childhood (Natu et al., 2019; Walhovd et al., 2016). Further, there has been no consensus on the age of peak thickness of the cortex. Recent evidence suggests that the whole brain average CT increases dynamically during the first year, and peaks at about 14 months of age (Wang et al., 2019). The study by Lyall et al. (2015) reported that CT reaches about 97% of the adult values at the age of two years. After that, the apparent trend seems to be linear cortical thinning during toddlerhood and preschool period (Gilmore et al., 2020; Levman et al., 2017). Remer et al. (2017) have reported general cortical thinning of 10–20%, with the exception of cortical thickening in posterior regions of the brain between the ages one and six years.

While CT growth is reported to peak after approximately two years of age, SA expansion continues throughout the preschool and school period (Sussman et al., 2016). This suggests that CT is relatively more established than SA at birth and, on the other hand, that SA is a major determinant in cortical volumetric development after the first two years of life. Lyall et al. (2015) reported regionally heterogenous SA expansion of 114.56% (range 64.55–149.57%) on average in the first two years of life. The heterogenous and non-linear expansion extends through the preschool period until SA peaks at around nine years of age (Wierenga et al., 2014).

As cortical volumetric growth is composed of SA and CT, its developmental paths mimic those of cortical SA. The rapid volumetric growth during the first year of life is characteristic of early development, with more a stable and gradual increase thereafter (Gilmore et al., 2012). Recently published brain growth charts reported the following "growth peaks": 1.7 years for CT, 11–12 years for SA and 5.9 years for cortical volumes (Betlehem et al., 2022).

2.1.2 White matter development

White matter (WM) consists of glial cells and axons (mostly covered with myelin), and it is usually studied as a structure connecting the pathways between functional brain regions. After four years of age, WM growth has been reported to increase faster than grey matter (GM) (Gilmore et al., 2020). While GM development takes place during the first years of life, and partly in the preschool period, WM development extends through young adulthood (Bethlehem et al., 2022).

The organization of what will eventually be the adult-like neural network between brain regions begins during the third trimester of gestation (Innocenti & Price, 2005). Simultaneously and subsequently to the network establishment, the WM fiber pathways become more mature and efficient in information transfer through myelination. Myelination is described as a process where the myelin sheaths wrap around the axons, which enables faster conduction of nerve impulses (Dubois et al., 2014). Even though myelination is a complex process with high interindividual variation as well, several themes are found to be characteristics of it. It occurs first in proximal and sensory (somatosensory, vision and audition) pathways, and later in distal and motor pathways. Consequently, myelination has been found to develop caudal-cranially, starting with the occipital lobe and then spreading to the parietal, temporal and frontal lobes (Lebel & Deoni, 2018). The majority of myelination takes place during the first years of life, but the process continues throughout adulthood (Yakovlev and Lecours, 1967), which highlights the role of WM as a critical marker in later brain development. Indeed, recently published volumetric brain growth charts place WM peak volume at 28.7 years (Betlehem 2022).

Hemispheric structural asymmetry, typically with non-linear developmental trajectories, is a phenomenon commonly reported in MRI studies (Brain Development Cooperative Group, 2012). Cortical asymmetry is first detected already in newborns (Lehtola et al., 2019; Li et al., 2014). Interestingly, the direction of lateralization seems to change with age: newborn studies have reported leftward volumetric asymmetry (Gilmore et al., 2007), whereas children and adults exhibit mostly rightward asymmetry (Toga & Thompson, 2003). The age-related trends of

cortical asymmetry form a potentially interesting phenotype for, for example, language-related studies.

Dimorphism between the sexes has been reported in the adult brain, with males presenting larger cortical and subcortical volumes than females (Duerden et al., 2020; Goldstein et al., 2001) and similar trends have been observed in studies with newborns and children (Bethlehem et al., 2022). Infant males had approximately 10% larger cortical GM than females, and the difference stays relatively the same in adults (Allen et al., 2003; Giedd et al., 1999). Cortical WM was also larger in infant males; however, the sex difference was smaller in infancy and early childhood (approximately 7%) compared to the adult brain (approximately 15%) (Allen et al., 2003; Gilmore et al., 2007). These findings support the view that WM maturation, compared to GM, is more prolonged. Dimorphism between the sexes in GM and WM are presented in Figure 2 (Bethlehem et al., 2022).

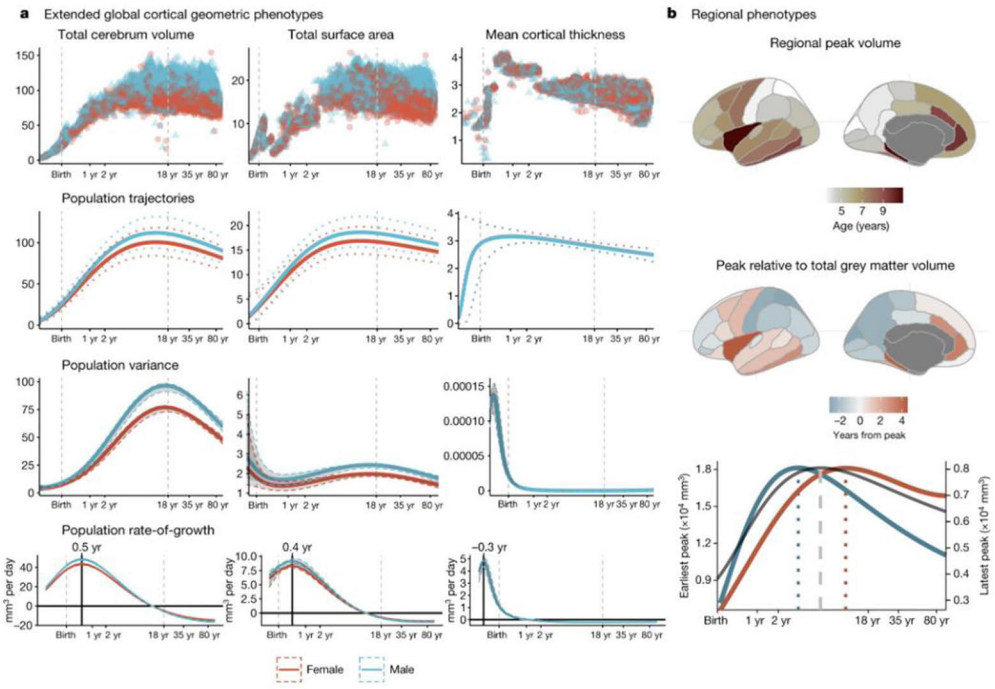

Figure 2. Extended global and regional cortical morphometric phenotypes. The figure was originally published in an article by Bethlehem et al., (2022), which is licensed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

2.2 Typical analyses and issues in the processing of structural magnetic resonance images

Paediatric neuroimaging studies encounter various methodological challenges that can affect the quality of the collected data. Structural MRI sequences are conducted fairly slowly (a time scale of minutes), and the scan protocols are very sensitive to movement. As children are not allowed to be sedated for scientific purposes due to the risk of later adverse effects, they are required to lie still while awake during the whole sequence, which poses a challenge (Poldrack et al., 2002; Theys et al., 2014). Of note, rather surprisingly, this challenge seems to be less prominent in newborns, which is reflected in relatively large sample sizes and the number of studies in this population. Movement during the scan results in increased motion artefact of the images and thus, repetitive measurements of the same sequences. This causes prolonged periods of scanning (Blumenthal et al., 2002). Sometimes a subtle motion causes challenges in the segmentation to different tissued compartments such as GM and WM even if the visible artefact was not seen (Alexander-Bloch et al., 2016). To ensure high-quality data, paediatric neuroimaging studies are typically required to schedule more time for preparations and scanning, which makes the study visits longer and increases the expenses of carrying out paediatric studies. This is often reflected in smaller sample sizes, too.

A visual quality check of the data and processed segmentations is beneficial and relatively fast (Blumenthal et al., 2002; Phan et al., 2018). However, the benefits of manual edits to improve the segmentation outcomes are not always clear (Beelen et al., 2020; Waters et al., 2019). While the errors in the output are often obvious and easy to recognize, the edits are often small, the process takes time and does not necessarily affect the cortical thickness, surface area or volumetric values significantly (Ross et al., 2021). In addition, the possible inter- and intra-rater bias during manual editing can decrease the reliability and repeatability of the data.

The final quality validation of the segmented and edited images can be challenging. A common type of quality control is a dichotomous pass or fail scale. The exclusion is often based on excessive motion artefact (Boutzoukas et al., 2020; Vanderauwera et al., 2018), pathology issues (al Harrach et al., 2019) or outlier cases (Nwosu et al., 2018). In inclusion, the data is reported to have a sufficient quality, without any detailed criteria or explanation (Barnes-Davis et al., 2020). Another approach is to rate the image quality on the Likert scale based on a number of categories from excellent to unusable (Blumenthal et al., 2002). The challenge is that even if the categories are presented, it is difficult to accurately define the cut off criteria between them (Blumenthal et al., 2002; Shaw et al., 2007). This issue can

reduce intra- and inter-rater reliability within a study and especially in comparisons between different studies. In addition, the use of many categories often leads to a situation where a researcher must draw a line of exclusion between two chosen categories, which fundamentally leads to a dichotomous scale: images with acceptable and unacceptable image quality. Instead of quality classification based on the amount of visible artefact, it might be more reasonable to evaluate brain regions of interest separately. This is justified by the fact that early life brain development is characterized by non-linear growth patterns, which is not necessarily taken into account in adult-based templates in segmentation tools (Phan et al., 2018; Wilke et al., 2003). The quality check based on regions of interest (ROI) would detect the issues of this phenomenon better compared to evaluation-based visible motion artefact severity.

For inclusion and exclusion criteria of images, two major approaches are proposed: a micro- and macro scale approach. In the micro scale assessment, errors can be found in the segmented image and score the image based on their number and size. This requires a segmentation tool that enables the visual check of the output. However, this approach has multiple challenges. Firstly, there are many errors that do not warrant exclusion of the ROI in question, for example, small errors in the border between WM and GM. In some cases, these types of errors can be abundant, despite rarely meeting the exclusion criteria. Thus, it is challenging to determine how to scale these errors and consider them in comparison with larger errors. Secondly, in many cases it is not obvious whether there is an error in the slice or not (one typical example is an image with poor WM–GM contrast). Counting errors by the number of slices with a certain type of error could lead to differences between raters and further, significant differences between these cases. Thirdly, quality control protocols are often described on a general level in scientific studies (Barnes-Davis et al., 2020; Boutzoukas et al., 2020; El Marroun et al., 2016; Kamson et al., 2016). As a result, there are many ways to assess and scale the errors in the automated segmentation. In contrast, in the macro scale assessment, one can quickly look at the brain image and assess the amount of motion artefact and the clarity of the WM–GM border. In borderline cases, the image can be segmented and then assessed for major segmentation errors, such as errors in the WM–GM border or large unsegmented areas. One key challenge also with this approach is the lack of objectiveness, as these errors are very difficult to quantify or to describe in articles or instructions. On the other hand, an experienced rater can make the same visual check for all images and is capable of quickly learning to exclude images of poor quality, and therefore a high internal reliability should be attainable. Furthermore, as this type of assessment can be done quickly, unclear cases can be verified by other raters with little additional time commitment.

Another critical challenge is the variation is segmentation tools and techniques. As there are multiple automated and semiautomated segmentation tools, most of their templates are based on adult imaging data and, consequently, there is no gold standard for paediatric neuroimaging tool or segmentation protocol. These deficiencies in paediatric studies can make the comparison between studies unreliable (Schoemaker et al., 2016).

FreeSurfer (http://surfer.nmr.mgh.harvard.edu/) is open-source software designed to process brain MRI images, and it is commonly used in paediatric neuroimaging with healthy population but also when studying pathologies (al Harrach et al., 2019; Boutzoukas et al., 2020; el Marroun et al., 2016; Wedderburn et al., 2020). The automated segmentation protocol is based on surface-based parcellation of cortical regions, which is based on patterns of cortical folding and previous knowledge of the structural anatomy (Dale et al., 1999). After automated segmentation with FreeSurfer, the output images can be visually checked and manually edited (Figure 3). Critical locations in manual editing are skull-stripping and errors in the border between WM and GM and in the outer border of GM. The FreeSurfer instructions suggest that the visual check and manual edits should last approximately 30 minutes. In our experience, this is an underestimate because careful manual editing alone usually takes over 60 minutes. As the automated segmentation tools are designed to speed up the process of reliably analysing neuroimaging data, the time requirement for manual editing has a critical role and is therefore evaluated in Study I.

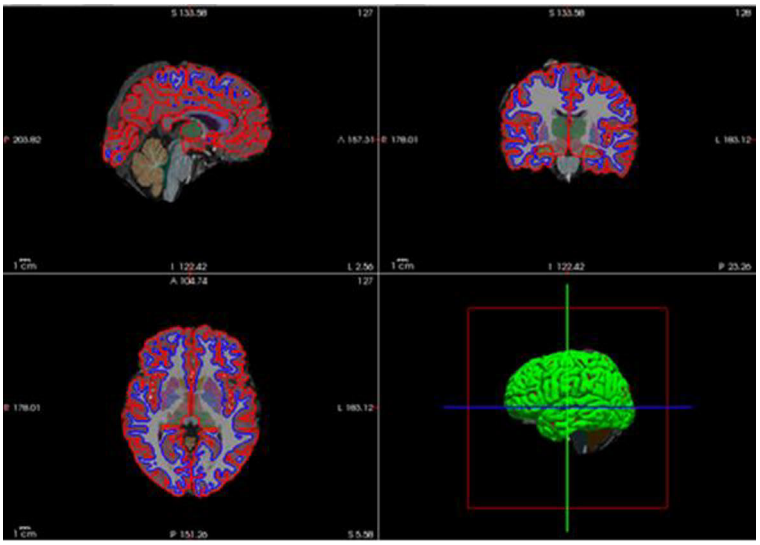

Figure 3. An example of segmented FreeSurfer output in Freeview from Study I (Pulli et al., 2022

2.3 Environmental factors affecting brain development

Cortical development is sensitive to extrinsic factors starting from the perinatal period, and the early life exposures may have long-lasting and adverse effects on neurocognitive development among individuals later in life (Duan et al., 2019; Lehtola et al., 2020). Many psychiatric and neurocognitive disorders have been postulated to result from alterations in cortical and subcortical maturation in early childhood (Hazlett et al., 2012; Li et al., 2016), and to capture such signals is of importance for both clinical and scientific purposes. For example, the ability to select an appropriate set of covariates for the data analyses is dependent on knowledge of important and potentially confounding factors.

The central nervous system (CNS) starts to develop early in the prenatal period during the first trimester of pregnancy (Stiles & Jernigan, 2010). Thus, the fetus is extremely vulnerable to environmental changes, chemical exposures and particularly to the mother's pre- and perinatal health profile. It is known that a mother's substance abuse (e.g., cocaine, alcohol, or drugs) during pregnancy is significantly harmful to fetal development. Alcohol is a chemical exposure known to have wide effects on a child's brain development: it is documented that it results in smaller GM volumes, for example, in frontal and parietal lobes and subcortical regions (Donald et al., 2016). Smoking during pregnancy disrupts normal brain growth and also leads to smaller total brain and GM volumes, probably due to nicotine-induced hypoxia (Lambers & Clark, 1996; Salminen et al., 2019). Other chemical exposures from the mother, including the use of a wide variety of illicit drugs and medication for psychiatric disorders, is suggested to influence brain development (Pulli et al., 2018).

Maternal psychiatric disorders, such as depression and anxiety, during pregnancy are a common problem and are considered a notable health risk for the child (Malm et al., 2015; Scheinost et al., 2017). Maternal anxiety and depression have been found to increase the risk for the child to experience physical, emotional and cognitive impairment later in life (Pechtel & Pizzagalli, 2011). It is estimated that children whose mothers have had depression in the perinatal and postpartum period are more likely to develop depression later in life compared to children whose parents have not experienced corresponding symptoms (Capron et al., 2015; Koutra et al., 2013; Susser et al., 2016). Prenatal depression and psychological distress symptoms have been reported to affect cortical gyrification and are associated with hippocampal and

amygdala volumes, both negatively and positively (Duan et al., 2019; Lehtola et al., 2020; Wu et al., 2020). Interestingly, not all children with a maternal anxiety background develop any kind of psychiatric or cognitive deficits later in life. Thus, it is still unknown which neurogenic mechanisms activate these developmental variations (Lautarescu et al., 2020; Moog et al., 2022).

Recent evidence suggests that the early organization of cortical structure is affected by both maternal and paternal factors. The role of paternal features in early cortical maturation can be aroused by different mechanisms. As paternal education has been found to associate with cortical thinning in the child, it could reflect the father's ability to psychosocially support the mother during pregnancy, which is reflected in healthy maternal outcomes (Shapiro et al., 2017) and further in the cortical maturation of the child. Another theory suggests that children with a higher intelligence quotient (IQ) have thinner cortices due to a more prolonged period of maturation of the higher order networks (Shaw et al., 2006). This is hypothesized to be beneficial for later cognitive outcomes. Of note for the interpretation of these findings, detected thinner cortices could also reflect increased myelination of the underlying WM (Sowell et al., 2004), as paternal education has been reported to correlate positively with overall WM volumes in newborns (Knickmeyer et al., 2017). Overall, the changes in paternal features associated neonatal brain maturation profiles could be driven by genetic influences.

Families' socioeconomical status (SES) is found to predict better cognitive performance and is associated with structural changes in cortical maturation. SES is commonly defined via one or a combination of the following factors: household income, parental education level and occupational status. While there are few infant studies carried out, it is suggested that lower SES is associated with smaller GM volumes globally in this age group (Betancourt et al., 2016). After infancy, there are an increasing number of studies in this field of interest, which support the trend of SES correlating positively with GM (Hanson et al., 2013) and WM (Luby et al., 2013) volumes during childhood. Cortical thickness is found to follow different maturational patterns in different SES classes from early childhood to adulthood (Leonard et al., 2019; Piccolo et al., 2016). However, there are also major studies that have not found associations between SES and structural development (Brain Development Cooperative Group, 2012). The variance in the results between studies demonstrates the complexity of SES as a variable (Tooley et al., 2021).

As described earlier, a developing brain encounters multiple variables from the perinatal period to late childhood that can affect the brain maturation directly or indirectly via maternal profile. The growing evidence demonstrates the significant role of perinatal surroundings, family demographics and obstetric variables for cortical brain anatomy (Jha et al., 2019; Knickmeyer et al., 2017; Sheridan et al.,

2012; Wu et al., 2020). As these cortical changes in later childhood possibly highlight the role of postnatal environment as an important predictor, the found associations between different variables and a newborn's brain could demonstrate the genetic influence in brain maturation. The extent to which these environmentally associated structural features persist into later ages during childhood is still unclear.

2.4 Associations between language skills and structural brain development

Children develop at a rapid pace during the first two years of life, with significant increases in cognitive and language skills, and this process coincides with the rapid structural growth of the brain. Cognitive development during the early years can be divided into a number of main dimensions, of which the most important are attention, perception, memory, executive functions and language. This thesis focuses primarily on language learning, which is closely linked with the development of other cognitive domains.

Learning to understand and produce spoken language is a complex process that requires the co-operation of many cognitive abilities, including memory, motor function and audiovisual stimulation. Consequently, coherent language production requires the synchronized function of numerous brain regions. There is great variation among individuals in the timing of acquired language skills that is still considered normal development (Dosman et al., 2012).

The critical foundations for language development are laid in the first 24 months of life. A newborn's auditory-perceptual abilities allow encoding of spoken language. Newborns have memory for and the ability to react to their mother's voice (DeCasper & Fifer, 1980) and prefer to listen to human speech as opposed to other sounds (Vouloumanos & Werker, 2007). By the age of 12 months, infants' perceptual abilities are well tuned to the speech sounds of the ambient language being learned (Kuhl et al., 2006). The ability to learn probabilities among sequences of sounds and syllables is already present in infants (Saffran et al., 1996), which is thought to be a foundation for learning the structures of words and sentences.

Vocal development progresses from newborn crying and vegetative sounds to cooing and laughing at six weeks of age, babbling at 16 weeks, reduplicated babbling at six months, and variegated babbling at 10 months. The first words are typically produced around 12 months of age. After that, the vocabulary is added to during the second year of life with great variation among individuals. Children also begin to form simple grammatical structures as early as 12 months, and most children

generatively combine words into utterances of two or a few words by the age of 24 months. Children's ability to maintain speech sounds in working memory (Gathercole et al., 2006) and executive function, the ability to allocate mental resources to relevant information, are also closely aligned with language abilities.

Broca's and Wernicke's areas are the brain areas traditionally related to language skills. They are located unilaterally in the frontotemporal lobe in the left hemisphere for most people (Figure 4, Encyclopædia Britannica, 2022). Broca's area is in the inferior frontal gyrus and is linked to expressive speech processing, whereas Wernicke's area lies posteriorly to Broca's area, in the superior temporal lobe, and is responsible for receptive language skills. These regions are connected to each other by the WM fiber tract called the arcuate fasciculus, and together these structures form the perisylvian language network that is considered important for speech production and processing (Hickok & Poeppel, 2007). As earlier language-related studies have possibly highlighted the role of specific traditional brain regions in language development, the current research emphasizes that many brain regions contribute to language skills and together form a whole-brain language network. In addition, even the classical areas entail multiple cell types (http://www.talairach.org) and can have differences in gene expression profiles (https://portal.brain-map.org/). In addition, Wernicke's and Broca's areas contain multiple brain regions in many of the contemporary atlases (https://www.lead-dbs.org/helpsupport/knowledge-base/atlasesresources/cortical-atlas-parcellations-mni-space/).

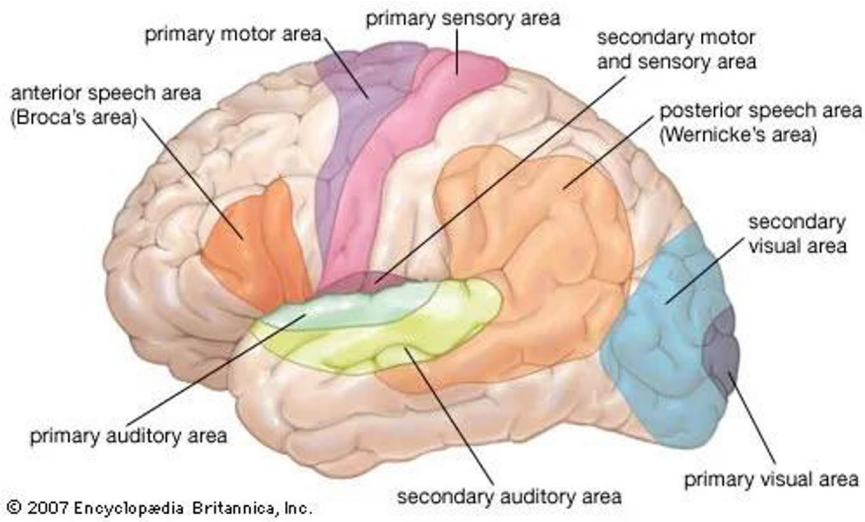

Figure 4. Functional areas of the human brain. © Encyclopædia Britannica, Inc

The cerebellum is the largest structure of the hindbrain that lies at the back of the skull, below the temporal and occipital lobes. Its feedforward tracts are connected to other brain regions via cerebellar peduncles, which connect the cerebellum to pons (Heck & Sultan, 2002). Respectively, the feedback tracts are projected to cerebral areas via the thalamus (Salmi et al., 2009). As its role is traditionally related to motor coordination and fine motor movement, recent studies have found that the cerebellum is significantly associated with cognitive and emotional functions (Catani et al., 2008; Koziol et al., 2014). Strong evidence suggests that the cerebellum is an essential part of language processing (Highnam & Bleile, 2011; Mariën & Borgatti, 2018), which is also detected in the very early phases of life (Deniz Can et al., 2013). Impaired language abilities are reflected in structural cerebellar changes also with preterm born children and in neurocognitive disorders (Dellatolas & Câmara-Costa, 2020; Stipdonk et al., 2021).

In addition to the perisylvian language network and arcuate fasciculus, a few other WM pathways are commonly linked to language abilities among other functional tasks (Figure 5, Swanson et al., 2015). The uncinate fasciculus connects limbic system structures to the frontal cortex, whereas the inferior longitudinal fasciculus binds the temporal lobe to the occipital lobe and connections to the uncinate fasciculus (Tufo et al., 2019; Olson et al., 2015). WM trajectories between temporal and frontal regions are often divided into the dorsal and ventral pathways (Brauer et al., 2013). The dorsal pathway includes the arcuate fasciculus and parts of the superior longitudinal fasciculus, whereas the ventral pathway includes the uncinate fasciculus and inferior longitudinal fasciculus with its lateral branches. As they are responsible for phonological and semantic processing, respectively, damage in these networks is associated with impaired language skills (Yang et al., 2017; Zhang et al., 2021). From the developmental perspective, other WM tracts, such as the corpus callosum and thalamocortical pathway, have been recognized as affecting early language processing (Alcauter et al., 2014; Swanson et al., 2015).

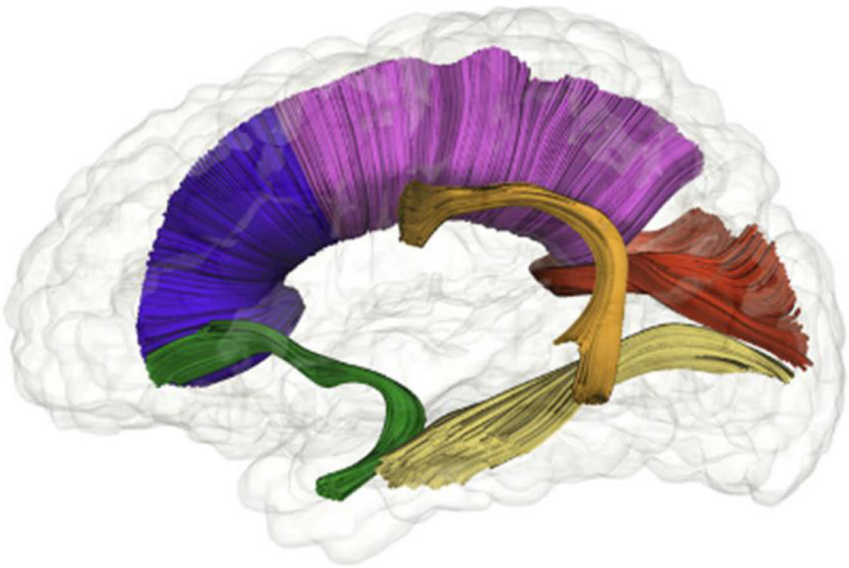

Figure 5. Language-related WM fiber pathways: blue, anterior portion of the corpus callosum; magenta, corpus callosum body; red, splenium; green, uncinate fasciculus; yellow, inferior longitudinal fasciculus; orange, arcuate fasciculus (Swanson et al., 2015). Reprinted with permission.

Although the language network in the brain is relatively well established in adults, much less is known about structural and functional characteristics that underlie language development in young children. Hemispheric asymmetry is characteristic of the language-related brain structures (Toga & Thompson, 2003). For instance, the frontal and temporal regions connected to language skills show mostly left-sided dominance (Friederici & Alter, 2004; Binder et al., 2000), albeit with marked inter-individual variability (Rosen et al., 2011; Schonwiesner et al., 2005). It is suggested that newborns do not exhibit such hemispheric asymmetries (Perani et al., 2011), but the left lateralized and localized language network would be achieved during the first two years of life (Emerson et al., 2016). However, there is currently a marked paucity of imaging data between the neonatal period and two years of age, when brain growth is the fastest and coincides with the most dynamic phases of language development.

3 Aims

This thesis is foundational work for developmental cognitive neuroscience that uses measures derived from structural MRI. We critically assessed the feasibility of adult segmentation software for five-year-olds, carefully mapped important demographics that explained cortical anatomy, and finally performed a systematic review of neuroimaging and early life language development as an example of the key findings and limitations in the current state of the art.

I. To describe in the semiautomated segmentation and quality control protocol of structural brain images that was used in the FinnBrain Birth Cohort Study, which relies on the FreeSurfer v6.0 and ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis) consortium tools. To compare the images of the FinnBrain quality control protocol and unedited MR images that are automatically segmented with FreeSurfer. In addition, to compare these results to the ones that are automatically segmented with computational anatomy toolbox (CAT12).

II. To quantify the associations between environmental and obstetric variables, family demographics and gross anatomy of cortical SA and volumes in healthy, typically developing five-year-olds. To compare these results to the ones carried out with infants to evaluate whether these factors could have programming effects and, on the other hand, study how environmental factors affect cortical development.

III. To carry out a systematic review of studies that have combined magnetic resonance imaging and language-related measures in healthy normally developing infants from birth to two years of age. The aim was to describe the current knowledge on the topic and assess the quality of how the background information on the subjects had been reported.

4 Materials and Methods

This thesis consists of two prospective MRI studies (I and II) using original data from the FinnBrain Birth Cohort Study, and one retrospective systematic review (Study III) of language-related MRI studies carried out with children aged from birth to two years. The basis of the thesis is to study the normal brain development of a typically developing population.

The participants in studies I and II are part of a larger FinnBrain Birth Cohort Study population (www.finnbrain.fi) (Karlsson et al., 2018). Initially, mothers and their spouses were recruited at three maternal welfare clinics in Southwest Finland and the Åland Islands during the first trimester ultrasound visits. Of the current sample of participating mothers, 96.6% were of Finnish ethnicity. Parental background information was gathered pre- and postnatally by questionnaires and included monthly income, educational level, diagnosed medical conditions, medications affecting CNS, and substance use during pregnancy. For CNS affecting medications, only serotonin reuptake inhibitors (SSRI) and norepinephrine reuptake inhibitors (SNRI) or benzodiazepines were reported. No illicit drug use substance was reported in the current sample. Parental psychiatric history variables were determined on the basis of questionnaires for mothers, which screened the parents for symptoms of depression, anxiety disorder and psychotic symptoms. Obstetric data was retrieved from the Finnish Medical Birth Register of the National Institute for Health and Welfare (www.thl.fi).

At five years of age, subjects were invited to research visits assessing the child's somatic health, neuropsychological and language development as well as neuroimaging. The inclusion criteria for the neuroimaging measurements were a visit to a FinnBrain Child Development and Parental Functioning Lab, including neuropsychological measurements. However, five participants were included without neuropsychological visits: three participants had an exposure to maternal prenatal synthetic glucocorticoid treatment (recruited separately for a nested case–control sub-study). In addition, two participants were initially enrolled for pilot scans. The exclusion criteria for studies I and II were: 1) born before gestational week 35 (before gestational week 32 for those with exposure to maternal prenatal

synthetic glucocorticoid treatment); 2) developmental anomaly or abnormalities in senses or communication (e.g., blindness, deafness, congenital heart disease); 3) known long-term medical diagnosis (e.g., epilepsy, autism); 4) ongoing medical examinations or clinical follow up in a hospital (meaning there has been a referral from primary care setting to special health care); 5) child's use of continuous, daily medication (including per oral medications, topical creams and inhalants, except for desmopressin medication, which was allowed); 6) history of head trauma (defined as concussion necessitating clinical follow-up in a health care setting or worse); and 7) metallic ear tubes (to assure good-quality scans), and routine MRI contraindications (certain types of cardiac pacemakers, cochlear implants or other implants held by magnet, metallic clips).

4.1 MRI data acquisition

All MRI scans were performed for research purposes by the research staff that consisted of a research nurse, four PhD students and two MR technologists / radiographers. Before a visit, each family was contacted and recruited via telephone calls by a research staff member. If the family was willing to participate, a research member met families personally to share information about the study visit and to bring practice materials for home rehearsal.

During the study visit, a flexible timetable was arranged for the subjects to get familiar with the research staff. Imaging was practised using a mock scanner that consisted of a toy tunnel and homemade wooden head coil. The aim of this was to introduce the scanning procedure to the child and simultaneously demonstrate the importance of lying still and practising communication during scanning. These preparations before the official scanning were highly variable as we did our best to accommodate each child's characteristics in cooperation with the family. Finally, a light meal was served before the scan.

The subjects were scanned during natural sleep or awake while they were watching a movie or television show of their choice. Hearing protection included wax earplugs and headphones. Foam padding was used to help the head stay still and ensure a comfortable position. A member of the research staff stayed in the scanner room with one parent throughout the scan. Subjects were given a "signal ball" to throw in case they needed or wanted to stop the scan at any point. Many of the

methods used to reduce anxiety and motion during the scan have been described in earlier studies (Greene et al., 2016).

4.1.1 Image acquisition

Subjects were scanned with a Siemens Magnetom Skyra fit 3T (Siemens Medical Solutions, Erlangen, Germany) using a 20-element head/neck coil. The Generalized Autocalibrating Partially Parallel Acquisition (GRAPPA) technique was used to accelerate image acquisition (Parallel acquisition techniques factor of two was used). The MRI data was acquired as part of max. 60-minute scan protocol. For the purposes of the current study, we acquired high resolution 3D T1 images using a magnetization prepared rapid gradient echo (MPRAGE) sequence with the following sequence parameters: TR = 1900 ms, TE = 3.26 ms, TI = 900 ms, flip angle = 9 degrees, voxel size = 1.0 × 1.0 × 1.0 mm^3, FOV 256 × 256 mm^2. In addition, the MRI protocol consisted of a T2 turbo spin echo (TSE), a 96-direction (b = 1000 s/mm^2, 9 b0s, TR/TE = 9300/87.0 ms) diffusion tensor imaging (DTI) sequences (Merisaari et al., 2019) and a 7-minute functional MRI acquisition. All images were viewed by one neuroradiologist, who then consulted a paediatric neurologist, if necessary, which is described in detail in the incidental findings article (for details, see Kumpulainen et al., 2020).

4.2 MR image analysis

4.2.1 Data Processing in Study I

In Study I, we used a subsample of our MRI study population of five-year-olds, including the participants who were scanned before a temporary stop on visits due to the restrictions caused by the COVID-19 pandemic. The scans were performed between October 29, 2017, and March 1, 2020. In total, 146 (40% of the reached families) participants attended imaging visits. Due to scanning failures or excess motion artefacts in the T1 images, 12 subjects were excluded. Thereafter, 134 T1 images (mean age 5.34 years, SD 0.06 years, 72 boys, 62 girls) entered the processing pipeline. After the first segmentation, 13 T1 images were excluded and, finally, 121 images were selected for the ROI exclusion-based ENIGMA quality control protocol. A flowchart depicting the different exclusion steps is presented in Figure 6. The demographics of the subjects are presented in Appendix A, and detailed information about the quality control protocol with large supplementary materials is presented in the original article I (Pulli et al., 2022).

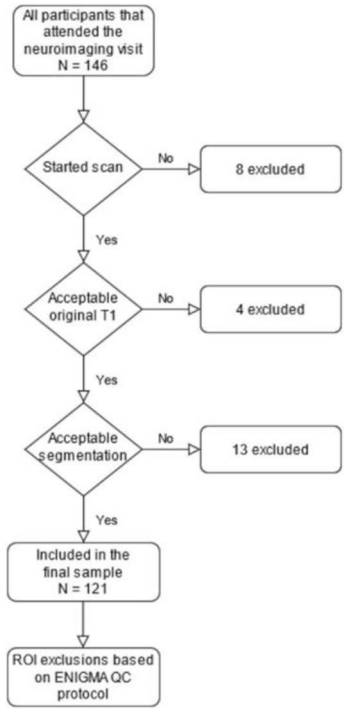

Figure 6. A flowchart of the steps leading to the final sample size of 121 subjects in Study I.

Data Processing with FreeSurfer

FreeSurfer recon-all automated segmentation

Cortical reconstruction and volumetric segmentation for 134 images were performed with the FreeSurfer software suite, version 6.0.0.7. If there were multiple T1 images, we selected the one with the least motion artifact and ran the "recon-all" processing stream with default parameters. The process begins with transformation to the Talairach space, intensity inhomogeneity correction, bias field correction (Sled et al., 1998), and skull-stripping (Ségonne et al., 2004). Thereafter, WM is separated from GM and other tissues and the volume within the created WM–GM boundary is filled. Then the surface is tessellated and smoothed. After these preprocessing steps, the surface is inflated (Fischl et al., 1999) and registered to a spherical atlas. This method adapts to the folding pattern of each individual brain, utilizing consistent folding patterns, such as the central sulcus and the sylvian fissure as landmarks, allowing for high localization accuracy (Fischl et al., 1999). FreeSurfer uses a probabilistic approach based on Markov random fields for automated labeling of brain regions. Cortical thickness is calculated as the average distance between the WM–GM boundary and the pial surface on the tessellated surface (Fischl and Dale, 2000). The cortical thickness measurement technique has been validated against post-mortem histological and manual measurements (Kuperberg et al., 2003; Salat et al., 2004).

FreeSurfer Manual Edits and the Freeview Quality Control Protocol

Freeview was used to view and edit the images using the standard command recommended by the FreeSurfer instructions with the addition of the Desikan–Killiany atlas that made it possible to correctly identify the ROIs where errors were found. Images with excess motion artifact or large unsegmented regions were excluded. There were 13 participants who were excluded due to erroneous segmentation. The images that passed the initial quality check were manually edited. All images were examined in three directions and one hemisphere at a time, and the edits were made for every slice regardless of the ROI in question. Subsequently, we ran the automated segmentation process again according to the FreeSurfer instructions. The images were then inspected again for errors, and the ROIs with errors that affect WM–GM or pial borders were excluded in the Freeview quality control protocol. The Freeview protocol presented in this study was adapted locally

for the FinnBrain Neuroimaging Lab as a method to assess errors in a slice-by-slice view from the official quality control procedure provided in the FreeSurfer instructions.

Errors in Borders

The automatically segmented images generated by FreeSurfer were visually inspected, and the errors were either manually corrected or the ROI with the error was excluded depending on the type of error. Excess parts of the skull were removed where the pial border was affected by them (Figures 7A, B). A common error was that the superior sagittal sinus (SSS) was in some parts included within the pial border. We stopped editing the SSS after an interim assessment as it was time consuming and had little effect on the final outcome.

Arteries were removed to avoid segmentation errors between them and WM (especially relevant for anterior temporal areas and the insulae). This was done by setting the eraser to only delete voxels with an intensity between 130 and 190 in the brainmask volume. The arteries were removed throughout the image regardless of whether they caused issues in the segmentation on that specific slice or not, as can be seen in Figure 7C. In cases where an error appeared in a junction between ROIs, all adjoining ROIs were excluded.

Some errors could not be fixed easily. Occasionally, the pial border cut through the cortex (Figure 7D). In these cases, the remaining GM mask is too small, and this error cannot be easily fixed in Freeview. Manual segmentation of a T1 image is arduous and unreliable with a 1 mm^3 resolution even when the edits would cover small areas. Moreover, the FreeSurfer instructions do not recommend this approach. Therefore, we simply excluded these ROIs.

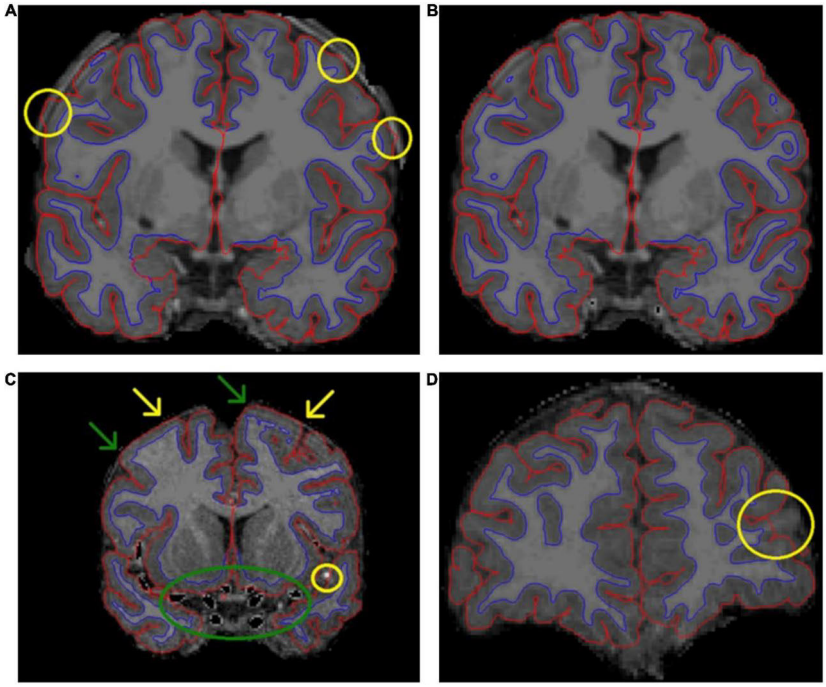

Figure 7. Common errors and fixes in the pial border and non-brain tissues. (A) Skull fragments cause errors in the pial border (yellow circles). (B) The same subject with skull fragments removed. In panel (C), arteries were removed (green circle). Voxels were removed with an intensity between 130 and 190, and therefore some parts of arteries were not removed (yellow circle). (C) Also demonstrates the challenges with artifact, meninges, and the pial border. In some areas, the pial border may extend into the meninges (yellow arrows). Meanwhile, at the other end of the same gyrus, the border may seem correct (green arrows). It is difficult to fix these errors manually. Additionally, the visible motion artifact adds further challenges to manual edits of the pial border. In panel (D), the pial border cuts through a gyrus.

Small errors in the WM–GM border were prevalent throughout the brain. The corrections were made by erasing excess WM mask (Figure 8). WM–GM border was inspected after the manual edits. A continuous error of 10 slices or more in the coronal view led to the exclusion of all the ROIs directly impacted by the error. Furthermore, ubiquitous errors in the WM–GM border, as markers of motion artifact, led to exclusion of the whole brain (as shown in Figure 9).

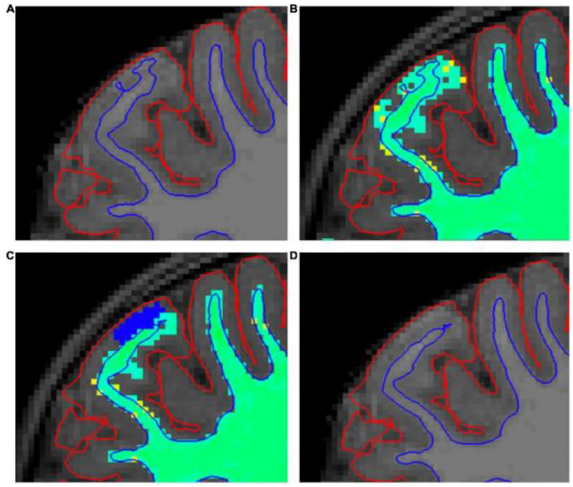

Figure 8. A demonstration of the white matter (WM) mask editing protocol. (A) Shows a typical error in the border between WM and GM, where WM extends too close to the pial border, searched for in the "brainmask" volume (A, D). (B) Shows the same error in WM volume with "Jet" colormap (B, C). (C) Shows how the errors are fixed by erasing the erroneous WM mask (blue voxels). (D) The final result after manual edits and the second recon-all.

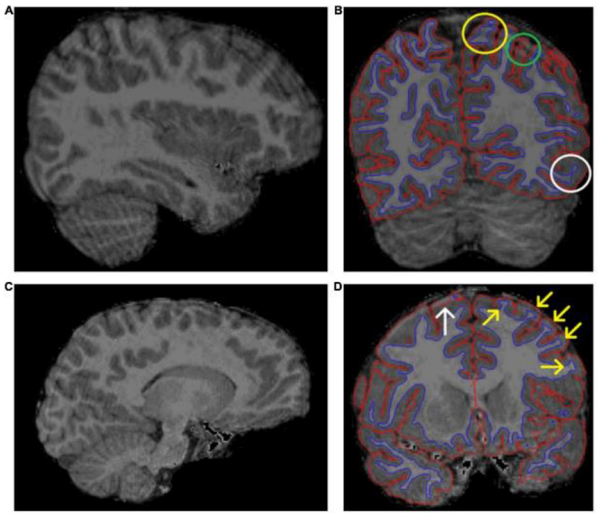

Figure 9. Examples of excluded brain images. (A) "Waves" in the image, marking motion artifact. (B) The same subject as in panel (A) in a coronal view and borders visible, which shows motion artifact-related errors in the border between WM and GM (yellow circle). Additionally, there is a potentially unsegmented area (green circle) and poor contrast between WM and GM (white circle). (C, D) Another excluded image. The motion artifact in panel (C) is not as pronounced as in panel (A). However, (D) still shows some typical errors for images with a lot of artifact. There is a pial error (white arrow). The yellow arrows show typical cases where the "ringing" causes the WM mask to "widen" where the actual WM meets the ringing motion artifact.

There were some error types that cannot be easily fixed but also do not warrant exclusion, such as the pial border errors that often extend into the cerebrospinal fluid or meninges around the brain. These errors are detected systematically in almost every T1 image analysed by the FreeSurfer recon-all segmentation protocol. The challenge is that sometimes the real border between GM and the surrounding meninges cannot be denoted visually and therefore the error cannot be reliably fixed. This problem is further complicated by the fact that the motion artifact may mimic the border between GM and meninges, making the visual quality control challenging (Figure 7C).

Errors in Cortical Labeling

A common challenge was the presence of WM hypointensities in the segmented images, which could erroneously appear in the cortex. These errors were typically small and did not cause errors in pial or WM–GM borders, and in those cases did not require exclusion. The hypointensities themselves were rarely successfully fixed by editing the WM mask and therefore were left unedited unless they caused errors in the GM–WM border. In those cases, removing the WM mask often fixed the error in the border, even though the incorrect hypointensity label often remained in the WM segmentation. If the correction of the errors in the WM-GM border was unsuccessful, the ROI was excluded (Figures 10A, B). Of note, these errors can only be seen with the anatomical labels as overlays, unless they affect the WM–GM border.

One typical error occurred at the posterior end of the lateral ventricles, where it causes segmentation errors in the adjacent cortical regions, typically the precuneus and the lingual gyrus. These regions were excluded from analyses if there was a distortion in the GM–WM border (Figures 10C, D). Unfortunately, hypointensities often appeared in ROI junctions, leading to exclusion of multiple regions due to one error.

Of note, we also briefly analysed the errors in subcortical labeling considering the putamen, but the edits were largely unsuccessful. The subcortical segmentation is not further discussed here. However, a separate study has been carried out to validate subcortical segmentation procedures for our data, and interested readers are referred to the original publications (Lidauer et al., 2022; Pulli et al., 2022).

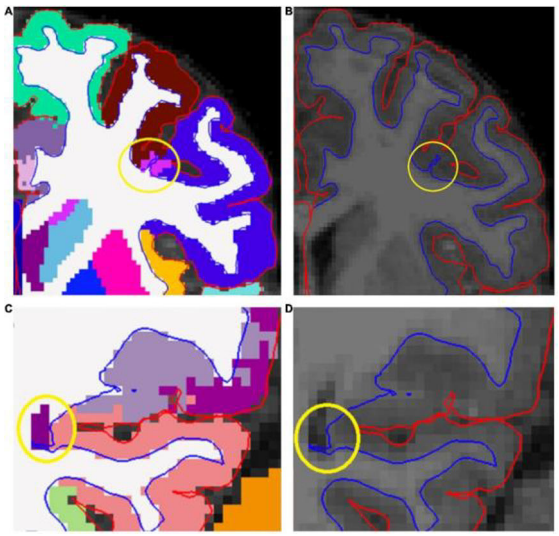

Figure 10. (A, B) WM hypointensity affecting the border between WM and GM, denoted by a yellow circle. (C, D) The posterior part of the lateral ventricle causes distortion to the WM–GM border (yellow circle). If the error was not successfully fixed, all regions adjoining the error were excluded.

ENIGMA Quality Control Protocol

After the quality control including manual edits, we conducted a quality check with the ENIGMA Cortical Quality Control Protocol 2.0 (April 2017). There, the FreeSurfer cortical surface measures were extracted and screened for statistical outliers using R10 and visualized via Matlab (Mathworks) and bash scripts. Visual representations of the external 3D surface and internal 2D slices were generated and visually inspected according to the instructions provided by ENIGMA in https://drive.google.com/file/d/0Bw8Acd03pdRSU1pNR05kdEVWeXM/view. The ENIGMA Cortical quality check instructions point out how certain areas have a great anatomical variation and therefore they note the possibility of being more or less stringent in their quality control. Considering this and the fact that the example images provided in the ENIGMA instructions are limited in number, we deemed it necessary to describe how we implemented these instructions in our sample.

The External View

We started by viewing the external image. The pre- and postcentral gyri were assessed for meninge overestimations, which can manifest as “spikes” or flat areas. These errors were rare in our sample. These cases were excluded as instructed.

The supramarginal gyrus has a lot of anatomical variability and when quality checking it, we decided to be flexible as suggested by the ENIGMA instructions. The images were excluded only if the border between supramarginal and inferior parietal regions cuts through a gyrus, leading to discontinuous segments in one of the regions (Figure 11A). This type of error led to an exclusion also in other regions. Similarly, in cases with supramarginal gyrus overestimation into the superior temporal gyrus, we only excluded clear errors. Another common error is overestimation of the insula extending to the midline (Figure 11B). In these cases, we excluded insula and the regions adjacent to it in the midline (e.g., the medial orbitofrontal region in the case of Figure 11B).

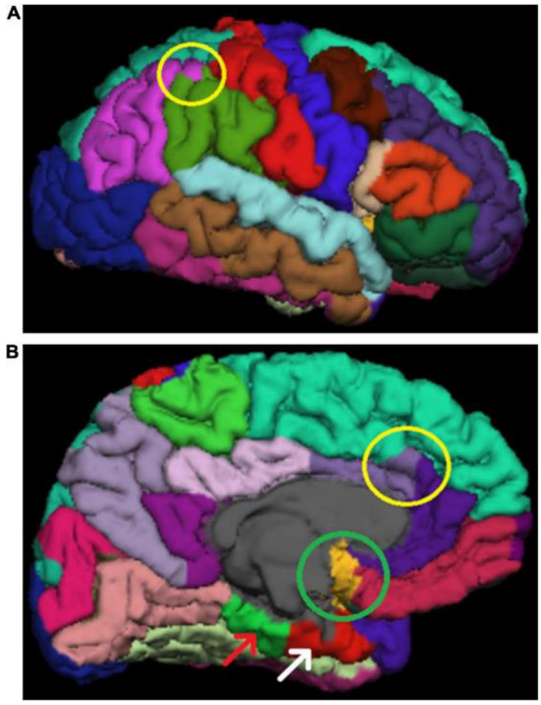

Figure 11. (A) Shows an error (yellow circle) where the inferior parietal region (purple) cuts through a whole gyrus in the supramarginal region (green). (B) Insula overestimation in the midline (green circle). Additionally, there is an error in the border between the superior frontal and caudal anterior cingulate. This border should follow the sulcal line. The rostral anterior cingulate was not considered erroneous in these cases.

The border between the superior frontal region and the cingulate cortex (Figure 11B) is typical place for errors. A prominent paracingulate sulcus may cause underestimation of the cingulate cortex and, consequently, overestimation of the superior frontal region. This was most often seen on the left caudal anterior cingulate (Figure 11B), where we excluded the cases where the border did not follow sulcal lines anteriorly, as demonstrated in the image examples in the instructions.

The size of the pericalcarine region was overestimated in some cases. According to the instructions, cases where the segmentation is confined to the calcarine sulcus should be accepted. Therefore, we excluded cases where the pericalcarine region extended over a whole gyrus into the lingual gyrus or the cuneus. Cases of superior parietal overestimation were excluded as instructed. These errors were rare in our sample. Similarly, errors in detecting the banks of the superior temporal sulcus were excluded as instructed.

The border between the middle and inferior temporal gyrus was not assessed, as the instructions suggested that most irregularities seen there are normal variants or relate to the viewing angle. Similarly, we did not quality check the entorhinal/parahippocampal regions in the external view, due to a lot of variation in the area. Furthermore, this region looks poor in practically all images (e.g., in Figure 11B) as do all the regions adjacent to the base of the skull and therefore, the quality assessment in those regions requires additional procedures, which are beyond the scope of the current study, to confirm their usability in statistical analyses.

The Internal View

In the internal view, regions with unsegmented GM were excluded. These errors often reflect WM hypointensities seen in Freeview. Interestingly, even relatively large hypointensities do not necessarily equate to errors in the borders set by FreeSurfer and therefore do not necessarily result in false CT calculations.

Temporal pole underestimations were sometimes seen. However, the cases were rarely as clear as presented in the instructions. Therefore, we had to use both coronal and axial views to assess the situation and make exclusions when both views supported an error in segmentation.

One commonly detected error in our sample was the false pial surface delineation in the lateral parts of the brain. This was particularly prevalent in the middle temporal gyri. Notably, it is possible to attempt to fix these types of topological errors, such as by using control points or brain mask edits. Some previous studies (e.g., Ross et al., 2021) have done this. However, the editing time was 9.5 hours on average and

the edits did not affect the conclusions. Therefore, this type of edit was considered too time-consuming and challenging compared to the expected effect on results. Consequently, the ROIs affected by these errors were excluded from the analyses. This error was assessed from 2D slices, in which what seems to be an error may be caused by partial volume effects. Consequently, we only considered an exclusion when clear errors were seen in two adjacent slices. A particularly clear example of this can be seen in Figure 12, where the WM extends outside the segmentation. The error is also visible in the external view, where these regions do not appear as smooth as normally; however, the decisions to exclude a ROI were always made on the basis of the internal view. This kind of error was harder to recognize in Freeview and represents the most striking difference in results between the ENIGMA and Freeview quality control protocols.

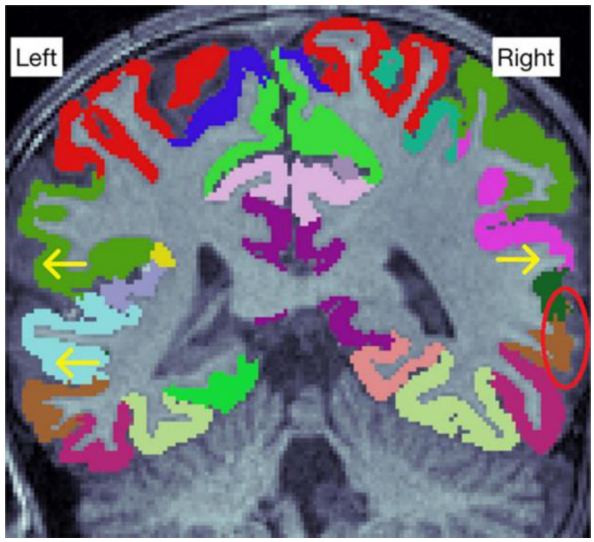

Figure 12. Visible errors in the lateral parts of the image (arrows). An especially clear error is denoted by the red circle, where some WM is seen outside the cortical segmentation.

Statistical Outliers

After the systematic viewing of the problem regions, the statistical outliers were inspected. This rarely led to new exclusions, as many of the statistical outliers were among the excluded subjects or the outliers were ROIs where the instructions did not give any tools to assess whether they were correct. Therefore, we had to simply double check the internal view to rule out segmentation errors.

ENIGMA Exclusion Differences Between Edited and Unedited Images

We performed the full ENIGMA quality control protocol for all edited images that were included in the ROI-based analyses. To assess how manual edits affect the number of excluded regions, we performed the ENIGMA quality control protocol also on a half sample (n = 61) of unedited images. In borderline cases (usually located in the borders between the supramarginal and superior temporal gyri and/or between the caudal anterior cingulate and superior frontal gyri), we consulted the ENIGMA quality control protocol of the edited images to make the same ruling if the error was similar. Likewise, in the cases where the edited image passed the internal or external view with no ROI exclusions, but did not pass in the unedited version, the images were directly compared to each other to ensure the reason for not passing is an objective difference, as opposed to a human error or a different ruling in a borderline case.

Exclusions

We used a dichotomous rating scale: pass or fail. The amount of motion artifact (marked by "concentric rings" or "waves") and the quality of the WM–GM border were assessed from the original T1 image. In borderline cases, we ran the standard recon-all and made a new assessment based on the segmented image output. Considerable segmentation errors, such as large missing areas or ubiquitous errors in WM–GM border, were reasons for exclusion. Additionally, ENIGMA exclusion criteria were implemented as instructed. In some borderline cases, another expert rater assessed the image quality. Some images that were considered for inclusion but excluded after the first recon-all can be seen in Figure 9. These images had significantly more artifact than other images in our sample, although arguably they could have been included since the amount of artifact could be described as "moderate." However, we decided to implement strict exclusion criteria to ensure the high quality of data.

In addition to the reported FreeSurfer and ENIGMA quality control protocols, we compared FreeSurfer default recon-all to recon-all with the "-mprage" and "-schwartzya3t-atlas" optional flags. We also carried out quality assessments with alternate processing tools: CAT12 (Seiger et al., 2018) and Qoala-T (Klapwijk et al., 2019). These analyses are reported in detail in the original article (Pulli et al., 2022).

4.2.2 Data processing in Study II

In Study II, we used the full sample of our MRI study population of five-year-olds. The MRI scans were performed between October 29, 2017, and March 1, 2021. Altogether 541 families were contacted and 478 (88%) of them reached. In total, 203 (42% of the reached families) participants attended imaging visits. After excluding 30 subjects due to the failure or excess motion in T1-weighted images (in addition to three subjects who were born before 35 weeks of gestation), a total of 170 subjects qualified for the analyses. The descriptive statistics of the included subjects are presented in the Tables 1 and 2.

Cortical reconstruction and volumetric segmentation for all images that were included in the processing pipeline were performed with the FreeSurfer software suite, version 6.0.0 (http://surfer.nmr.mgh.harvard.edu/). The recon-all processing stream was run with default parameters. The editing was carried out according to the FinnBrain quality control protocol, which is presented in detail in Study I and in the original article I (Pulli et al., 2022).

The FreeSurfer pipeline divides the cortex to 33 regions bilaterally. However, instead of analysing each region specifically, we were interested in the gross cortical anatomy and the ROIs were determined for volumes and SAs as follows: total, left and right hemispheres, and lobe division bilaterally to four main lobes (frontal, temporal, parietal and occipital). We included the MRI data that passed the FinnBrain quality control protocol (n=170). Specifically, after analyzing the T1-images through the quality control protocol, only a small percentage of the subjects passed the protocol without any ROI exclusion (as reported in the Study I). Consequently, we did not exclude participants if they had minor errors in a single or few ROIs within the lobar measures, as volumes and SAs in specific excluded regions were relatively small and are not likely to affect the measurements of the main lobes and hemispheres. We also noticed that typical errors in borders and cortical labeling were rather located inside main lobes than areas between lobes or between hemispheres. The ROI selection was justified by the notion that prior work (Jha et al., 2019) reported associations that reflected gross anatomy in the main results. In addition, expanding the approach to smaller ROIs would have aggravated the multiple comparison issues, and we did not pursue such analyses in this work.

Table 1. Descriptive statistics for continuous demographic and obstetric history variables

Continuous variables	**Missing**	***N***	**Mean**	**SD**	**Min**	**Max**
Birth weight (g)	0	170	3550.55	477.07	2450.00	4980.00
Gestational age at birth (weeks)	0	170	39.80	1.50	35.14	42.29
Postnatal age at MRI (years)	0	170	5.40	0.13	5.09	5.79
Ponderal index at MRI	0	170	14.10	1.20	11.21	17.63
Maternal age at child's birth (years)	0	170	30.55	4.74	19.00	41.00
Paternal age at child's birth (Years)	51	119	31.88	5.14	20.00	45.00
Maternal BMI before gestation	1	169	24.27	4.33	17.48	41.95
5min Apgar score	0	170	9.10	0.69	4	10

BMI= Body mass index, MRI= Magnetic resonance imaging

Table 2. Descriptive statistics for categorical demographic and obstetric history variables

Categorical variables		***N***	**%**
Sex		170	
	Male	93	54,7
	Female	77	45,4
NICU admission		169	
	Yes	23	13,6
	No	146	86,4
	Missing	1	
Mode of delivery		169	
	Vaginal	138	81,7
	C-section	31	18,3
	Missing	1	
Maternal education*		164	
	Low & mid	83	50,6
	High	81	49,4
	Missing	6	
Paternal education*		111	
	Low & mid	74	66,7

	High	37	33,3
	Missing	59	
Maternal smoking during gestation**		170	
	Yes	11	6,5
	No	159	93,5
Diagnosis of gestation diabetes		169	
	Yes	20	11,8
	No	149	88,2
	Missing	1	
SSRI/SNRI medication during gestation		155	
	Yes	6	3,9
	No	149	96,1
	Missing	15	

NICU = Neonatal intensive care unit; SSRI/SNRI = Serotonin and norepinephrine reuptake inhibitors. *Educational levels: Low & Mid = Upper secondary school, vocational school or lower & University of Applied Sciences, High = University, **Maternal smoking during gestation=a combination variable of maternal smoking during early and late pregnancy

4.3 Statistical analysis

4.3.1 Statistical analyses in Study I

Statistical analyses were conducted with IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY, United States). The ROI data was confirmed to be normally distributed using JMP Pro 15 (SAS Institute Inc., Cary, NC, United States) based on visual assessment and the similarity of mean and median values.

To compare the differences between the included (the participants that were included in ROI based analyses, n = 121) and excluded (all participants that lacked usable T1 data, n = 25) groups, we performed independent samples t-tests for age from birth at scan, gestational age at scan, gestational age at birth, birthweight, maternal age at term, and maternal body mass index (BMI) before pregnancy. In addition, we conducted Chi-Square tests for child gender, maternal education level, maternal monthly income estimates after taxes, maternal alcohol use during pregnancy, maternal tobacco smoking during pregnancy, maternal medical history, and maternal medication use at gestational week 14 (non-steroidal anti-inflammatory drugs, thyroxin, SSRI or SNRI, and corticosteroids), or at gestational week 34

(thyroxin, SSRI or SNRI, corticosteroids, and blood pressure medications). To limit the chance of false positives, the categories in history of disease and medication during pregnancy were only included in statistical analyses when there were at least four participants who had a history of the disease or used the medication.

To compare the exclusion rates between Freeview and ENIGMA quality control protocols, as well as between ENIGMA quality control protocols of edited and unedited images, we conducted Chi-Square tests (among all datapoints, single ROIs, and internal/external view passes in ENIGMA).

The inclusion criterion for the ROI based comparisons was passing the ENIGMA quality control protocol. To compare edited FreeSurfer to unedited FreeSurfer, we conducted a paired samples t-test. We calculated the absolute values of the change in CT between unedited and edited images for each ROI separately using the following formula: (CD/CU) * 100%, where CD is the absolute value of the difference in mean CT between edited and unedited images and CU is the mean CT in the unedited images. Furthermore, we conducted a paired samples t-test with the mean CT values from all ROIs to measure the change between edited and unedited images. The same analyses were performed for WM SA and GM volume.

To assess the effects of manual editing and quality control on group comparison and brain structural asymmetry results, we conducted independent samples t-tests for sex differences in CT, SA, and volume measurements between a sample without quality control (n = 121 for every ROI) and the quality-controlled sample (maximum n = 121, where number of included ROIs varies). Using these same samples, we also conducted paired samples t-tests for the 34 ROIs in both hemispheres to examine structural asymmetry.

All significances were calculated 2-tailed ($\alpha = 0.05$). To adjust for multiple comparisons in ROI-based analyses, we conducted the Bonferroni correction by setting the p value threshold to 0.05 divided by the number of comparisons (= the number of ROIs = 68), resulting in $p = 0.000735$. We note that the p value cut off for the current study is somewhat arbitrary and thus we also report the raw p values in the tables.

4.3.2 Statistical analyses in Study II

Brain variables included cortical volumes and surface area for the following regions: total cortex, both cortical hemispheres, and the aforementioned four main lobes bilaterally. The refence article by Jha et al. (2019) was set as a basis as we included 16 demographic variables into the analyses (Tables 1 and 2). Categorial variables included: sex, NICU admission, mode of delivery, maternal and paternal education

level, maternal smoking during pregnancy, diagnosis of gestational diabetes during gestation and the use of SSRI/SNRI medication during pregnancy. Continuous variables included: birth weight, gestational age at birth, postnatal age at MRI, ponderal index at MRI, maternal age at child's birth, paternal age at child's birth, maternal BMI before gestation and 5min Apgar score (a scoring system that provides a standardized health assessment for infants 1, 5 and 10 minutes after delivery). To aid interpretation of the results, we have divided the variables into three groups: 1) fixed family factors and child features at scanning (child sex, maternal and paternal education level, maternal age at child's birth, paternal age at child's birth, postnatal age at MRI, and ponderal index at MRI), 2) incidental pregnancy and delivery related factors (mode of delivery, child birth weight, gestational age at birth, NICU admission, and 5 min Apgar score), 3) maternal prenatal health features (maternal pre-pregnancy BMI, maternal smoking during pregnancy, diagnosis of gestational diabetes and the use of SSRI/SNRI medication during pregnancy). The following variables were considered too unreliable or otherwise suboptimal in our questionnaire data, and were excluded: maternal and paternal psychiatric history, household income, gestation number, and number of siblings.

Statistical analyses were conducted using IBM SPSS Statistics 27.0. (IBM Corp., Armonk, NY, USA). The MRI data was confirmed to be normally distributed using JASP Statistics 0.14.1. (https://jasp-stats.org/), based on visual assessment, skewness, kurtosis and Shapiro-wilk p values. Correlation matrices with Pearson correlation were created for cortical volumes and surface areas with JASP Statistics. Lateralization calculations for hemispheres and lobes were carried out with JASP statistics for descriptive purposes. Lateralization indices were also confirmed to be normally distributed based on the same criteria as the initial MRI data.

Linear regression models were carried out using each brain variable separately as a dependent variable and the group of 16 demographics as independent variables. Stepwise linear regression models were applied. This study was exploratory in nature and therefore we decided (a priori) not to perform formal corrections for multiple comparisons and report raw p values throughout the manuscript and this thesis. The Bonferroni corrected p value at alpha level 0.05 over the 22 comparisons over the 11 regression models × 2 brain measures (SA, volumes) was $p < 0.002$.

4.4 Ethical approval

The Study was conducted according to the Declaration of Helsinki and was reviewed and approved by the Ethics Committee of the Hospital District of Southwest Finland

(ETMK:31/180/2011). A written informed consent was acquired from both parents at the beginning of every MRI visit and child assent was acquired prior to the visit and confirmed from the parents.

4.5 Study III literature search

In Study III, which was a systematic review, we were interested in language-related MRI studies of typically developing infants from birth to two years of age. The aim of the search was to cover comprehensively the MRI studies carried out with infants studying brain development with the hypothesis, that the search included the language related studies of our ROI, too. Consequently, no search term specified the language development. We conducted a PubMed search with a search term 1: ("Magnetic Resonance Imaging"[Mesh] OR MR imaging* OR MRI OR NMRI OR fMRI OR DTI OR "diffusion tensor imaging") AND ("Brain/growth and development"[Mesh] OR brain growth* OR brain developm*) AND ("Infant"[Mesh] OR infant* OR toddler*). The filters used were the publication date, ranging from January 1, 2012, to August 6, 2018, publication language English, and species humans. After reviewing the results of this search, we decided that a further systematic search focused on language development was warranted. Considering the aim of this study, the following exclusion criteria were listed:

1. The article was published in a language other than English.
2. The study was carried out with animals.
3. Prematurely born or low birth weight infants (<37 weeks of gestational age or body weight <2500g) were studied.
4. Participants had reported pathologies known to affect neurodevelopment.
5. No MRI scans were obtained in the study.
6. The article was a review.
7. The study did not include participants imaged under two years of age.

The search term 1 yielded 724 articles. The first screening for the search term 1 was done on August 6, 2018, following the listed exclusion criteria. In the first screening, articles were excluded based on the criteria already observed on titles or the abstract. We included 173 out of 724 articles for more detailed analysis. After that, the 173 articles that passed the first screening were assessed for eligibility. The

purpose of the assessment was to find the articles that met the inclusion criteria and, in addition, dealt with language development as confirmed by the whole text of the articles. Finally, the search term 1 was refreshed on February 19, 2020, and 949 articles were found. The new articles were screened based on titles and abstracts to see if there were articles suitable for the final review. We identified nine articles that met the inclusion criteria and dealt with language development. The study selection protocol for the search term 1 is described in Figure 13.

An additional and more targeted search was done to find new articles for the review, but also to get a more comprehensive view of the field of language studies and language development in general. The search allowed studies carried out with older, preschool age children, to present a full search of the performed studies. We conducted a PubMed search with a search term 2: ("Language Development"[Mesh] OR (“Language”[All Fields] AND “Development”[All Fields]) OR "programming languages"[MeSH Terms] OR ("programming"[All Fields] AND "languages"[All Fields]) OR "programming languages"[All Fields] OR "language"[All Fields] OR "language"[MeSH Terms]) AND ("brain"[MeSH Terms] OR "brain"[All Fields]) AND ("magnetic resonance imaging"[MeSH Terms] OR ("magnetic"[All Fields] AND "resonance"[All Fields] AND "imaging"[All Fields]) OR "magnetic resonance imaging"[All Fields] OR "mri"[All Fields]) AND ("growth and development"[Subheading] OR ("growth"[All Fields] AND "development"[All Fields]) OR "growth and development"[All Fields] OR "development"[All Fields]) AND ("humans"[MeSH Terms] AND ("infant"[MeSH Terms] OR "child, preschool"[MeSH Terms] OR "Infant, Newborn"[Mesh] OR "Child, Preschool"[Mesh])). Filters used were publication date ranging from January 1, 2012, to February 19, 2020. The search yielded 333 articles. The previously listed exclusion criteria were applied also to this search. The screening for the search term 2 aimed to find articles that met the inclusion criteria and dealt with language development. The articles were screened based on titles and abstracts, and one additional article was selected for the review (Feng et al., 2019).

In summary, 10 articles that met the criteria and studied infants’ cognitive development and language skills were selected for the review (as mentioned above, nine articles were identified through the search term 1 and one article was identified through the search term 2).

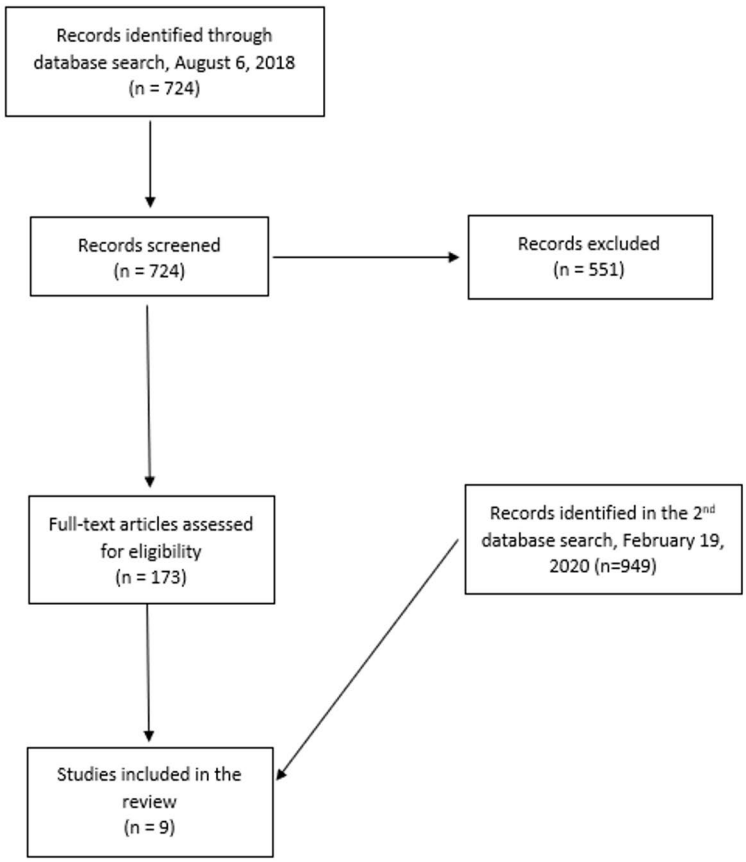

Figure 13. A flow diagram explaining the study selection protocol considering the search term 1.

5 Results

5.1 Feasibility of FreeSurfer processing for T1-weighted images of five-year-olds: FinnBrain quality control protocol (Study I)

5.1.1 Demographics

A few distinguishing features between the included and excluded subjects' groups were detected. There was a significant difference between groups in maternal BMI before pregnancy ($p = 0.03$). In the included group, mean maternal BMI was 23.9 ($n = 121$) vs. 26.0 in the excluded group ($n = 24$, information from one participant missing). Two types of medication were more common in the excluded group: SSRI or SNRI medication at 14 gestational weeks ($p = 0.03$; included group 109 no, 3 yes; excluded group 20 no, 3 yes) and blood pressure medication at 34 gestational weeks ($p = 0.03$; included group 113 no, 3 yes; excluded group 21 no, 3 yes). In addition, there was a marginally significant difference in SSRI/SNRI use at 34 gestational weeks ($p = 0.06$; included group 112 no, 4 yes; excluded group 21 no, 3 yes). Of note, here these differences are presented mainly for descriptive purposes rather than analysing if the listed early exposures predict image quality outcome comparison between unedited and manually edited FreeSurfer segmentations.

5.1.2 Comparison between unedited and manually edited FreeSurfer segmentations

Cortical Thickness

The difference in CT was not significant after Bonferroni correction in 57/68 (83.8%) regions. Unedited images had significantly larger CT values in two regions: the right rostral anterior cingulate and right superior temporal regions. Edited images had significantly larger CT values in 9 regions: the left and right caudal middle frontal, left and right inferior temporal, left and right superior parietal, right precentral, right superior frontal, and right supramarginal regions. The smallest (both

absolute and relative) change was detected in the left rostral middle frontal (0.0003 mm, 0.011%) and the largest (both absolute and relative) in the right caudal middle frontal (0.0526 mm, 1.857%) region.

The mean change in absolute CT values between the unedited and edited images was 0.0129 mm (0.441%). After including the direction of the change in the analysis, edited images had higher CT values (mean 0.00264 mm, 0.0901%), but the difference was not statistically significant ($p = 0.217$). Pearson correlations between edited and unedited images were calculated by ROI. The correlations were positive and ranged from 0.725 in the left insula to 0.984 in the left banks of the superior temporal sulcus region. All correlations remained statistically significant after Bonferroni correction.

Surface Area

The difference in SA was not significant after Bonferroni correction in 57/68 (83.8%) regions. Unedited images had significantly larger SA in 11 regions: the left and right postcentral, left and right precentral, left and right superior parietal, left and right insula, left caudal middle frontal, left superior frontal, and right inferior temporal regions. The edited images did not have significantly larger SA values in any of the areas. The smallest absolute change was observed in the right pars orbitalis (0.26 $mm^{2,}$ 0.028%) and the smallest relative change was seen in the right middle temporal gyrus (0.53 mm^2, 0.015%). The largest absolute change was found in the right superior parietal region (161.05 mm^2, 2.55%) and the largest relative change was found in the right insula (66.41 mm^2, 2.81%).

The mean change in absolute SA values between the unedited and edited images was 21.21 mm^2 (0.778%). When we include the direction of the change in the analysis, edited images had lower SA values than unedited images (mean 17.52 mm^2, 0.643%) ($p = 0.000044$). Pearson correlations between edited and unedited images were calculated by ROI, they all were positive and ranged from 0.669 in the left frontal pole to 0.995 in the left supramarginal region, and all remained statistically significant after Bonferroni correction.

Volume

The difference in volume was not significant after Bonferroni correction in 66/68 (97.1%) regions. Unedited images had significantly larger volumes in two regions: the left and right insulae. No areas where edited images had significantly larger volume values were detected. The smallest absolute change was observed in the left

precuneus (0.83 mm^3, 0.020%) and the smallest relative change was seen in the right superior parietal region (3.58 mm^3, 0.019%). The largest (both absolute and relative) change was observed in the left insula (189.56 mm^3, 2.400%).

The mean change in absolute volume values between the unedited and edited images was 31.53 mm^3 (0.345%). After including the direction of the change in the analysis, edited images had lower volume values than unedited images (mean 7.98 mm^3, 0.087%), but the difference was not statistically significant ($p = 0.175$). Pearson correlations between edited and unedited images were calculated by ROI, they all were positive and ranged from 0.744 in the right frontal pole to 0.995 in the left supramarginal region. All remained statistically significant after Bonferroni correction.

5.1.3 The ENIGMA and Freeview quality control protocols

Generally, the Freeview quality control protocol was more permissive than the ENIGMA protocol. The Freeview resulted in 7,824 accepted datapoints compared to ENIGMA's 7,208, out of possible 8,228 ($p < 0.0001$). The largest differences in both directions between these quality control protocols were detected in the left middle temporal gyrus (Freeview 119; ENIGMA 77; difference 42, $p < 0.0001$) and the left precuneus (Freeview 91; ENIGMA 110; difference 19, $p = 0.0011$). The worst quality areas (measured by total datapoints across both protocols) were the right postcentral gyrus and the right middle temporal gyrus with 187 and 188 (out of possible 242) valid datapoints, respectively. A minority of the subjects passed the protocols with no ROI exclusions: three for the Freeview volumetric protocol, 22 for the Freeview CT protocol, and three for the ENIGMA protocol (15 passes for the external and 25 passes for the internal view. Of note, the internal was rated as pass if it did not result in additional exclusions when viewed after the external view, and therefore the number of passes is overestimated).

ENIGMA exclusion differences between edited and unedited images

The sample size for the analysis was 61 participants, in total 4148 ROIs per hemisphere. In the left hemisphere, 238 edited and 318 unedited ROIs were excluded ($p = 0.0003$). In the right hemisphere, 215 edited and 319 unedited ROIs were excluded ($p < 0.0001$). In total, 453 edited and 637 unedited ROIs were excluded ($p < 0.0001$).

In the edited images, there were 10 that passed the external view without any ROI exclusions (unedited 5, $p = 0.17$), and 13 that passed the internal view (unedited 3, $p = 0.0073$). Some frequently observed differences between edited and unedited images in the ENIGMA internal view are presented in Figure 14.

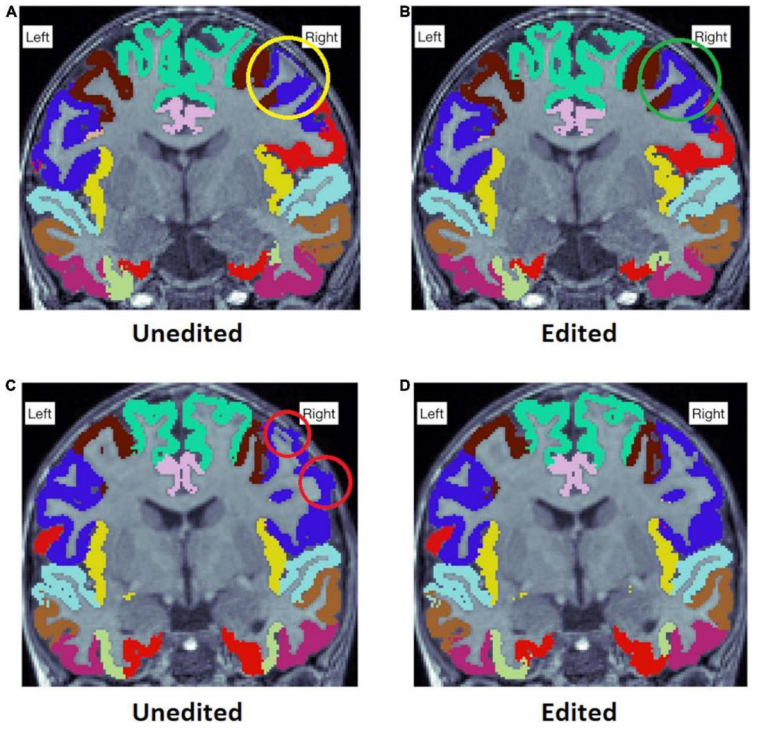

Figure 14. The right precentral gyrus was a region where differences between edited and unedited images were frequently observed both in cortical thickness and surface area (Study I). (A) An error in the right precentral gyrus, where the cortex is too thin (yellow circle). (B) is the edited image of the same participant, where the error is no longer visible in the region (green circle). (C) The precentral gyrus extending into the skull. (D) Shows the edited image of the same participant, where this error is no longer present.

5.1.4 Differences between FreeSurfer and CAT12 results

FreeSurfer and CAT12 had relatively low agreement in the sample. The difference in CT was not significantly different after Bonferroni correction ($p < 0.0007$) in 15/68 (22.1%) regions. FreeSurfer produced significantly higher values in 14/68 (20.6%) regions, and CAT12 produced significantly higher values in 39/68 (57.4%) regions.

5.2 Early-life environmental factors and family demographics influencing cortical anatomy in five-year-olds (Study II)

5.2.1 Surface area

Maternal smoking during pregnancy was associated with smaller SA in the right hemisphere ($p = 0.040$**),** specifically in the right temporal ($p = 0.043$), right parietal ($p = 0.005$) and also in the left parietal ($p = 0.039$) lobes. NICU admission predicted larger SA in total ($p = 0.046$), in both hemispheres (left $p = 0.049$, right $p = 0.020$) and in the left parietal lobe ($p = 0.010$). Two predictors were found for the right parietal lobe: maternal BMI before gestation associated with larger SA ($p = 0.071$) and higher 5 min Apgar scores associated with smaller SA ($p = 0.010$). The reported predictors survived the threshold of p-value < 0.05. However, the only predictor that survived the Bonferroni correction at $p < 0.002$ was sex (except for both occipital lobes).

5.2.2 Volumes

Higher maternal age correlated with larger volumes in the right hemisphere ($p = 0.018$) and in the right parietal lobe ($p = 0.006$). Higher maternal educational level was associated with larger volumes in total ($p = 0.021$), in the left hemisphere ($p = 0.016$), left frontal ($p = 0.029$), left temporal ($p = 0.020$), left parietal ($p = 0.002$), right temporal ($p = 0.014$) and in the right parietal lobes ($p = 0.019$). These associations survived the threshold of p-value under < 0.05, but the only predictor that survived the Bonferroni correction at $p < 0.002$ was sex (except for both temporal lobes). The regression models for total cortical SA and volume are presented in Appendix B.

5.3 Sex differences in Studies I and II

5.3.1 Cortical thickness and surface area

In Study I, girls had higher CT values than boys in all the regions with significant differences (in both samples with and without quality control). In the quality-controlled sample there were 16/68 ROIs with significant differences ($p < 0.05$) between girls and boys, and 28/68 ROIs in the sample with no quality control.

Boys had higher SA values than girls in all the regions with significant differences (in both samples with and without quality control). In the quality-controlled sample, there were 57/68 ROIs with significant differences ($p < 0.05$) between girls and boys, and 61/68 ROIs in the sample with no quality control.

In Study II, Sex was a significant predictor for SA throughout the cortex and in both hemispheres ($p < 0.001$, except for both occipital lobes ($p = 0.010$)). Males exhibited 8.5% larger total SA on average ($M = 187\,665$, $SD = 13\,881$) than females ($M = 172\,409$, $SD = 13\,065$).

5.3.2 Volumes

In Study I, boys had higher volumes than girls in all the regions with significant differences, (in both samples with and without quality control). In the quality-controlled sample, there were 42/68 ROIs with significant differences ($p < 0.05$) between girls and boys, and 39/68 ROIs in the sample with no quality control.

In Study II, Sex was a significant predictor for volumes throughout the cortex and in both hemispheres ($p < 0.001$, except in both temporal lobes (left $p = 0.048$, right $p = 0.009$)). Males exhibited 6.3% larger volumes on average ($M = 623\,594$, $SD = 41\,419$) than females ($M = 585\,748$, $SD = 43\,212$).

5.4 Cortical asymmetries in Studies I and II

5.4.1 Cortical thickness and surface area

In Study I, there were 18/34 ROIs with significant differences ($p < 0.05$) in CT between the two hemispheres (left thicker in 8, right thicker in 10) in the quality-controlled sample. In the sample with no quality control, there were 19/34 ROIs with significant CT differences between the two hemispheres (left thicker in 8, right thicker in 11).

There were 28/34 ROIs with significant differences ($p < 0.05$) in SA between the two hemispheres (left larger in 14, right larger in 14) in the quality-controlled sample. In the sample with no quality control, there were 30/34 ROIs with significant SA differences between the two hemispheres (left larger in 15, right larger in 15).

In Study II, the total SA correlated positively with hemispheric lobular SAs (rs range = 0.649-0.940) (Figure 15A). However, we found that the SA correlations were weaker between occipital lobes and other ROIs (rs range for the left occipital lobe was 0.44-0.658 and for the right occipital lobe was 0.451-0.659).

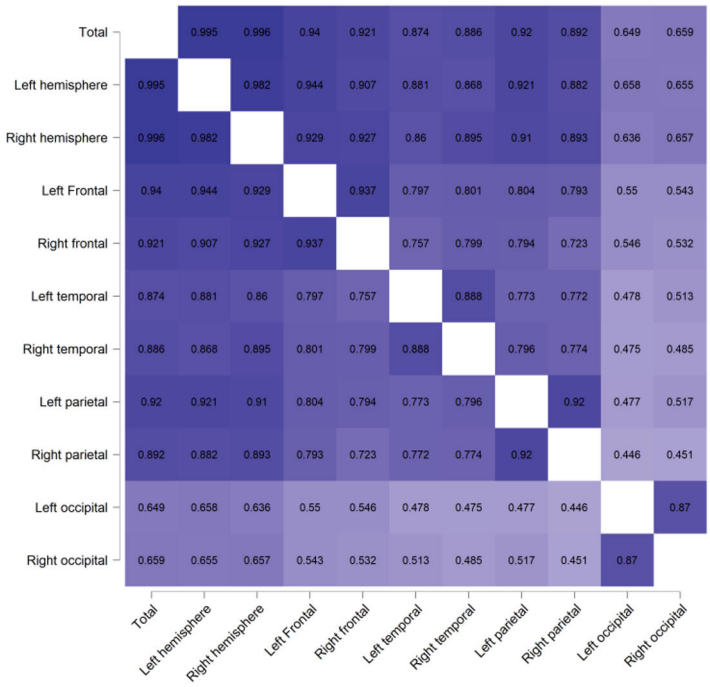

Figure 15A. Colour-scaled correlation matrix of the cortical surface areas (darker purple indicates larger correlation coefficient / effect size) (Study II).

5.4.2 Volumes

In the quality-controlled sample in Study I, there were 30/34 ROIs with significant volumetric differences ($p < 0.05$) between the two hemispheres (left larger in 15, right larger in 15). In the sample with no quality control, there were 32/34 ROIs with significant volumetric differences between the two hemispheres (left larger in 17, right larger in 15).

In Study II, the total volume correlated positively with hemispheric lobular volumes (rs range = 0.610-0.932), but the occipital lobes and other ROIs correlated weaker (rs range for the left occipital lobe = 0.398-0.611 and for the right occipital lobe = 0.432-0.636) (Figure 15B). Thus, the same phenomenon of weaker correlations between occipital lobes and other ROIs were observed both in the cortical SA and volumes. Remarkable interindividual variation in cortical metrics (both volumes and SA) in five-year-olds was observed.

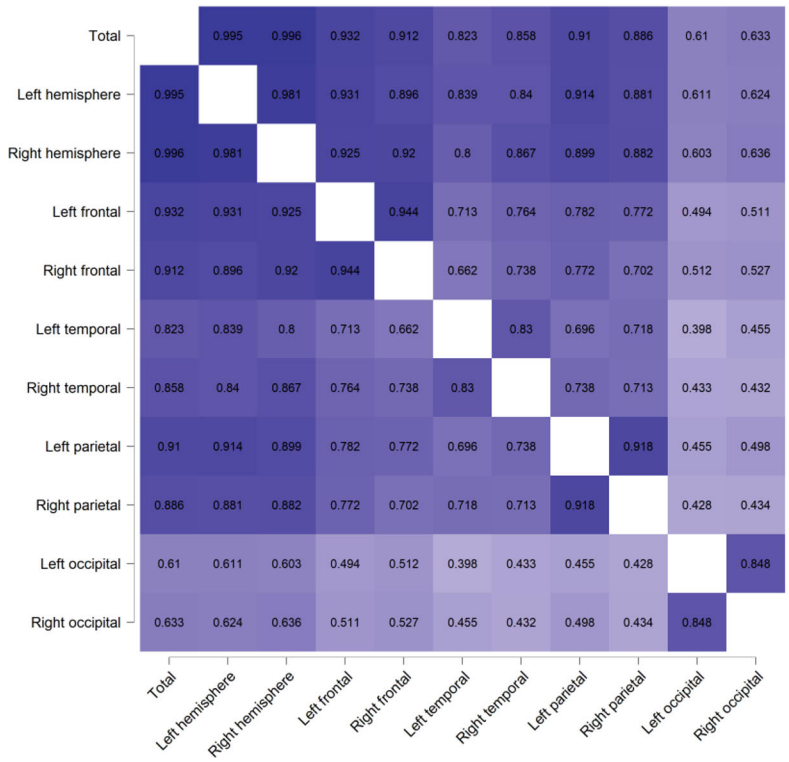

Figure 15B. Colour-scaled correlation matrix of the cortical volumes (darker purple indicates larger correlation coefficient / effect size) (Study II).

5.4.2 Cortical lateralization in Study II

The cortical SAs and volumes were mostly right lateralized. However, there was an exception in the temporal lobes, where larger volume and SA was detected on the left side (Figures 16A and 16B). However, the degree of lateralization (the difference between hemispheres or lobes of interest divided by the total volume or SA) was relatively small in our data. In addition, there was significant variation between individuals based on standard deviations.

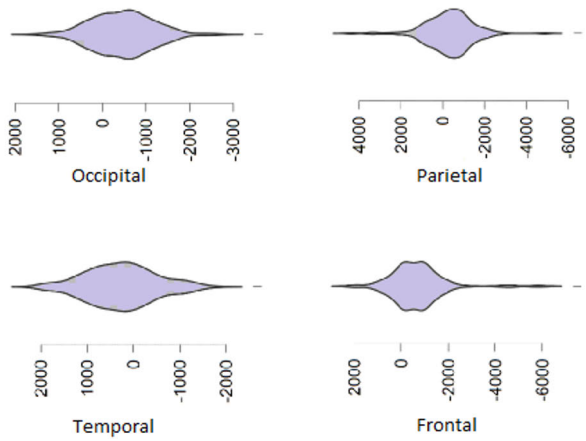

Figure 16A. Violin plots of the lateralization of the cortical surface areas (Study II). The figure represents lateralization indexes (mm^2), which is the difference between left and right cortical region.

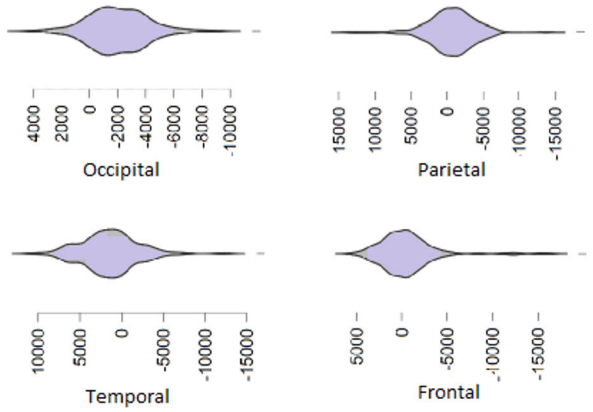

Figure 16B. Violin plots of the lateralization of the cortical volumes (Study II). The figure represents lateralization indexes (mm^3), which is the difference between left and right cortical region.

5.5 Language development from birth to two years of age (Study III)

The review included 10 articles, of which five studied structural GM and WM development and two studies dealt with functional development. Three articles studied WM tracts with DTI. The summary tables of the articles are presented in Appendix C. Four studies had longitudinal design and six were cross-sectional in terms of MR imaging. The number of participants varied greatly between studies, from 19 infants (Deniz Can et al., 2013) to 183 participants, of which 114 were infants aged 0–2 years (O'Muircheartaigh et al., 2014).

Different cognitive tests were applied in the studies. The Mullen Scales of Early Learning (MSEL) was the most commonly used method for assessing cognitive and language abilities, with eight of the articles implementing it. In addition, one study combined MacArthur-Bates Communicative Development Inventory and the Vineland Adaptive Behaviour Scales-II motor subscale score tests to MSEL results (Swanson et al., 2015). Two studies used the Bayley Scales of Infant and Toddler Development III as a cognitive test (Feng et al., 2019; Spann et al., 2014).

While most studies carried out MRI scanning and cognitive assessment on different days, some studies executed these two measurements during the same visit (Langer et al., 2015; Short et al., 2013; Spann et al., 2014; Swanson et al., 2015).

The reviewed studies utilised a range of different imaging techniques. T1 and T2-weighted sequences were applied in the cortical and subcortical studies. Two

studies used T2-weighted functional MRI scans (Alcauter et al., 2014; Emerson et al., 2016). The myelin water fraction (MWF) technique was used in three studies (Deoni et al., 2015; O'Muircheartaigh et al., 2013, 2014). The DTI studies used fractional anisotropy (FA) as the main predictor. As expected, the different methods revealed different aspects of the associations between brain development and language skills.

5.5.1 Findings from structural and functional MRI studies

The seven structural and functional MRI studies that were reviewed described various brain regions of interest: whole brain analyses (Deniz Can et al., 2013; Deoni et al., 2015; O'Muircheartaigh et al., 2013; O'Muircheartaigh et al., 2014; Spann et al., 2014), temporal regions (Emerson et al., 2016) and cortical functional networks arising from the thalamus (Alcauter et al., 2014). The asymmetry of the language-related regions was reported in most articles.

The earliest correlations between receptive and expressive communication skills and volumetric cortical brain maturation were observed during the first six weeks of life. The volumes of bilateral cingulate regions and bilateral posterior parietal regions imaged during the first six weeks of life correlated positively with expressive and receptive communication skills measured first at six-month time point and later negatively, measured at 12–24 months of age (Spann et al., 2014.). The leftward asymmetry of the temporal language-associated WM pathways has been frequently reported in adults. During infancy, these language regions showed a transition to asymmetry that began after approximately 11 months of age and continued into the second year of life. Early transition toward asymmetry correlated positively to language outcome scores in the inferior frontal gyrus, but not in the superior temporal gyrus (Emerson et al, 2016).

The MWF values and the MSEL scores in receptive language and visual reception subscales correlated positively with the left anterior thalamus and caudate at the age of one to six years (mean age approximately three years) (O'Muircheartaigh et al., 2013). These associations between the language scores and asymmetries seemed to change with age. Rightward asymmetry in the putamen correlated negatively with the expressive and receptive language scores in newborns but positively after three years of age. The leftward asymmetry in the frontal cortex correlated negatively with expressive language scores in newborns but positively after three years of age. The leftward asymmetry in the anterior thalamus and caudate correlated positively with the receptive language scores, peaking at around 20–23 months, and after that negatively after three years of age. The trends of changing asymmetry seemed to stabilise at around four years of age (O'Muircheartaigh et al.,

2013). The MSEL expressive and the receptive language scores correlated positively with the MWF values in a subset of temporal and frontal regions from three months to four years of age. These language-related regions were mostly left-lateralised or bilateral (O'Muircheartaigh et al., 2014). Better performance in cognitive tests correlated positively with the MWF values throughout the brain regions from three months to five years of age. The results suggested that children with better cognitive performance had slower but prolonged period of WM maturation during the first two years of life (Deoni et al., 2015).

Thalamic connections to the medial-visual, lateral-visual, occipital and default mode areas emerged after the age of one year. The thalamus-sensorimotor and thalamus-salience connectivity networks were the only ones present shortly after birth. The thalamus-salience network connectivity in one-year olds predicted better combined MSEL cognitive test results measured at the age of two years (Alcauter et al., 2014). Associations between cerebellar maturation and language skills were also detected, with language scores correlating positively with GM and WM intensities in right posterior cerebellar region (Deniz Can et al., 2013).

The results demonstrated that the volumetric development of language relevant brain regions is associated with developing language abilities (Alcauter et al., 2014; Deniz Can et al., 2013; Spann et al., 2014; Travis et al., 2013). Further, there are indications that lateralization of the regions responsible for language skills begins at a very early age and the change in the lateralization is associated with language development in a region-specific manner (Deniz Can et al., 2013; Spann et al., 2014).

5.5.2 Findings from DTI studies

The reviewed DTI articles studied different aspects of brain development: the maturation of early networks associated with language production (Feng et al., 2019; Swanson et al., 2015), and the structural differences between infants with or without genetic risk for developmental dyslexia (Langer et al., 2015).

In the adult literature, two major fiber connections between temporal and inferior language-related areas have been of special interest: ventral and dorsal pathways. Similar patterns have been found in the microstructure and asymmetry of these linguistic pathways between infants and adults. The dorsal pathways (arcuate fasciculus and superior longitudinal fasciculus) were less mature than the ventral pathways (middle longitudinal fasciculus and inferior longitudinal fasciculus) during infancy, but it is reported that after the first two weeks of life the gap closed over (Dubois et al., 2016), which highlights the role of peripartum period for dynamic brain changes.

The FA values in the arcuate fasciculus, uncinate fasciculus and left inferior longitudinal fasciculus did not correlate to language production during the first 24 months of life (Swanson et al., 2015). This suggests that language abilities start to develop before the maturation of connective WM trajectories take place, since the associations between language skills and temporal WM development have been widely reported later in life. On the other hand, positive correlations were found between language production and FA values in other pathways outside the temporal region (corpus callosum and right inferior longitudinal fasciculus) at 12- and 24-month time points (Swanson et al., 2015). Further, the FA values in the left arcuate fasciculus were lower among the children at risk for developmental dyslexia compared to the children at no risk at the age of one year (Langer et al., 2015). In addition, the FA values in the corpus callosum in newborns correlated positively with language scores measured at two years of age (Feng et al., 2019).

6 Discussion

6.1 Feasibility of FreeSurfer segmentation tool for T1-weighted brain images of five-year-olds and semiautomated protocol of FinnBrain Neuroimaging Lab

We described our semiautomated segmentation protocol used for image processing. While the study was based on existing guidelines by FreeSurfer and ENIGMA, we are confident that it will help investigators in the field of paediatric neuroimaging. We have added to the existing literature by assessing the effects of our manual edits on CT values and providing a detailed protocol for others to use.

The manual edits had minor effects on the CT values, less than 2% in all regions, and the differences were rarely significant. On the other hand, the larger effects (such as 0.05 mm) are bigger than the yearly change in CT in children (as estimated from figures in Walhovd et al. (2016) and Botdorf & Riggins, (2018)). Importantly, consistent bias in the absolute values does not necessarily affect longitudinal data, as the values can be scaled to only account for the relative value compared to group average. However, as this change represents measurement error due to artifact in the image rather than real difference in CT, reducing the variability should bring the results closer to the real values, whether scaled or absolute. Edited images had larger CT values in most cases where significant differences were seen. This is not surprising considering the nature of manual edits: most of the editing time is spent correcting small errors in the WM–GM border, which often results in a thicker cortex. Errors where pia extends too far into dura or cerebrospinal fluid were detected in the areas next to those structures, typically in the superior parts of frontal and parietal lobes, in regions such as the rostral and caudal middle frontal, superior frontal, superior parietal, precentral, and postcentral gyri. Therefore, they occur repeatedly in the same regions, which isolates these errors to certain areas and, therefore, cancels out some of the overall bias caused by them. These are the same

regions where WM-GM border errors were frequently detected. The pial errors were quite reliably approximately one to two voxels in thickness, while WM–GM border errors can be greater and occur anywhere in the brain. Furthermore, it is crucial to note that the errors in pial surface mainly affect the CT, whereas SA is often measured from WM–GM border and is therefore unaffected by the pial errors (Winkler et al., 2012). In addition, volumetric segmentation needs to be assessed separately from surfaces. A thinner cortex in the edited images were detected in only two regions: the right rostral anterior cingulate and superior temporal region. As there were arteries adjacent to the rostral anterior cingulate and erasing them may have resulted in a thinner cortex, the reason for the apparent cortical thinning of the superior temporal region is unclear.

Similarly, the effects of SA value edits were relatively minor, less than 3% across all areas and less than 1% on average, which is in line with previous study (McCarthy et al., 2015). However, there were a few exceptions, including some of the areas with more motion artifact and subsequently more edits (the pre- and postcentral gyri as well as the superior parietal region). Edited images had smaller SA values in the regions of detected differences, as expected considering the nature of our edits. The effects for volume were minor.

After evaluating the pros and cons of the micro- and macro approaches (discussed earlier in the Introduction), we chose macro scale assessment for the exclusion of whole images. Furthermore, we decided to apply it on a dichotomous pass or fail scale and skip further quality classification. One possible downside is the loss of subcategories in the accepted sample, since image quality can be included in regression analyses (Shaw et al., 2007). However, we performed a rigorous quality control protocol that evaluates quality on every ROI, resulting in the high quality of all datapoints in the final sample. Consequently, we believe further categorization based on overall image quality would not create any extra value for the study.

We applied the widely used ENIGMA quality control protocol (Thompson et al., 2020) to support decisions on inclusion and exclusion of ROIs. The internal view of the ENIGMA protocol gives 16 slices with colour-coded segmented ROIs, which gives a good overall view of the brain. However, it does not allow exploration of unclear cases, which leads to missing errors that are not located in the slices presented by ENIGMA. To explore this issue, we presented our own FinnBrain quality control protocol, and the use of Freeview for slice-by-slice assessment of the image was expectedly more sensitive to certain types of errors compared to the ENIGMA protocol. However, this protocol was not implemented for the final analyses due to the challenges such as large number of minor errors and the lack of consensus on how to treat them (discussed detailly in the methods section). Overall, the FinnBrain protocol was more permissive than the ENIGMA protocol, mostly

because the FinnBrain protocol lacks the external view (unlike ENIGMA) and hence, errors in borders between ROIs are not assessed. Therefore, if the FinnBrain quality control protocol were selected, it would have to be implemented together with the ENIGMA protocol.

FreeSurfer and CAT12 showed a relatively poor agreement in our sample, which was contradictory to the earlier reports of good agreement between these segmentation tools in adult population (Masouleh et al., 2020; Seiger et al., 2018). The disagreement was not systematic between these two, with CAT12 mostly overestimating CT compared to FreeSurfer, but also underestimating in some regions. As the early brain development undergoes non-linear regional growth curves (Phan et al., 2018; Wilke et al., 2003), adult templates are considered suboptimal for developing brains (Muzik et al., 2000; Phan, Smeets, et al., 2018; Yoon et al., 2009) and custom paediatric templates should be considered. Our results suggest that the choice of automated segmentation tool can be very influential.

There is no one "gold standard" processing method for paediatric images, which leads to methodological variation between different studies. Paediatric images are inherently more susceptible for segmentation errors than adult images due to increased motion during scans and the potential mismatch with adult templates and segmentation tools. This highlights the need for rigorous quality control to ensure high-quality data. We believe that detailed method descriptions are crucial for transparency that helps comparisons between studies. Here we described in the semiautomated segmentation protocol used in the FinnBrain Neuroimaging Lab, including manual edits and the implementation of the ENIGMA quality control protocol. We decided to use the standard recon-all without optional registration flags, as they did not provide additional benefits. Furthermore, we observed a surprisingly poor agreement between FreeSurfer and CAT12 output. Our semiautomated segmentation protocol provides high-quality paediatric neuroimaging data and could help investigators working with similar data sets.

6.2 Associations between environmental factors and cortical development

We quantified how the large pool of variables of family demographics and pre-/perinatal maternal factors are associated with cortical SAs and volumes at the age of five years. The variables were grouped into three groups for discussion: 1) Fixed family factors and child characteristics at scanning, 2) incidental pregnancy and

delivery related factors, and 3) maternal prenatal health features. In addition to reporting the results of our study population, we compared the results to those with newborn data (Jha et al., 2019) to evaluate whether these specific factors could have possible programming effects and, on the other hand, quantify the effects of environmental factors on early cortical development. In the reference work by Jha et al. (2019), parental factors were not found to predict cortical SA in newborns, but the delivery related factors associated with the newborn SA. Consequently, the results revealed that apart from child's sex, predictors that modulated SA in newborns were different compared to predictors that modulated SA in five-year-olds.

Sex was a significant predictor for both absolute cortical volumes and SAs, with males having larger cortex than females. These results of cortical metrics were significant globally, with the notable exception for bilateral occipital lobes. Larger absolute cortical volumes and SAs in males are also reported in previous newborn studies, with minor regional variation (Jha et al., 2019; Knickmeyer et al., 2017; Lehtola et al., 2020). Greater cortical volumes and SAs in males have been reported also in toddlers, and the similar trend lasts throughout the childhood and early adulthood (Remer et al., 2017; Wilke et al., 2007). Of note, CT is reported to show weaker dimorphism between the sexes than volumes and SA (Wierenga et al., 2014). In addition to child sex, two parental factors associated with larger cortical volumes (maternal age and maternal educational level), although the associations did not survive Bonferroni correction for significance.

Higher maternal education level during gestation correlated positively with total cortical volume, in the left hemisphere and partially in specific lobes bilaterally. In comparison with the findings from the reference study by Jha et al. (2019), parental factors were not found to predict cortical SA newborns. Instead, maternal ethnicity and paternal educational level associated with CT in newborns. Further, maternal education has been reported to correlate positively with cortical volumes in another newborn study (Knickmeyer et al., 2017).

We found that higher maternal age at delivery correlated positively with cortical volumes in the right hemisphere. In the newborn study by Jha et al. (2019), this correlation was not detected. Maternal age is a constant factor that is likely to contribute to the child's development throughout childhood. Previous clinical research has shown that higher maternal age is associated with reduced fertility and increased risk of offspring health problems (Jolly et al., 2000; Nassar & Usta, 2009). These findings of negative outcomes were based on physiological mechanisms and emphasized especially after 35 years of maternal age (Liu et al., 2011). On the other hand, epidemiologic studies have proposed that advanced maternal age (ranging from 25 to 35 years) could have a positive outcome on a child's health in terms of self-rated health status, height, obesity, diagnosed conditions and mortality

(Myrskylä & Fenelon, 2012). These results are likely to be partly explained by maternal SES (or other correlate of maternal age and education), which is also strongly associated with a positive outcome on a child's health profile (Erola et al., 2016). Other environmental factors may naturally co-influence the child's health.

NICU admission and maternal BMI before gestation correlated positively with cortical SA, whereas 5 min Apgar score and maternal smoking during pregnancy correlated negatively with cortical SA at the age of five years (these findings did not survive the Bonferroni correction for significance, though). In comparison with these findings to the reference work (Jha et al., 2019), maternal health profile during gestation and incidental delivery related features did not associate with cortical SA in the newborns.

The newborn study by Jha et al. (2019) found that child's birth weight, gestational age at birth and postnatal age at MRI associated with SA in newborns, which were not detected in our data at five years of age. However, it is suggested that birth weight is a prenatal factor with long-lasting influences in the cortical development (Gilmore et al., 2020; Walhovd et al., 2016). In our study, maternal smoking during pregnancy predicted smaller SA values comprehensively in the right hemisphere, except for the right frontal lobe showing no significant associations. Interestingly, maternal smoking during pregnancy did not predict brain morphology in newborns (Jha et al., 2019). On the other hand, only 8% of the women reported smoking during pregnancy, which could cause unreliability for analyses. Our results support the findings from previous studies, where pre/perinatal smoking exposure predicts smaller SA, CT and brain volumes (El Marroun et al., 2016; Liu et al., 2013; Salminen et al., 2019).

The lateralization of the cortical anatomy was right sided, with only the temporal lobes showing greater values on the left hemisphere. In general, the degree of lateralization was modest. Interestingly, the lateralization differed remarkably compared to what has been reported among newborns (Lehtola et al., 2019), where rightward asymmetry was detected in temporal lobe and leftward asymmetry in occipital and parietal lobes. While previous research has commonly reported rightward asymmetry of the brain from the childhood to early adulthood (Dean et al., 2018; Tanaka et al., 2013), leftward structural asymmetry has also been detected in infants (Gilmore et al., 2007).

6.3 Language-related structural brain development during the first two years of life

We found some evidence that the language circuit described in later ages is also associated to language skills during the first two years of life. During the selected age range, the human brain grows dynamically with variable speed at different ages and different regions (Knickmeyer et al., 2008). Since later in the development a major trend is apparent cortical thinning, it is critical to point out that for the described age range, thicker cortex likely relates to a more developed brain (Sowell et al., 2004). For the WM, the corresponding major developmental shift is the myelination, from mostly premyelinated after birth to nearly fully myelinated at the age of two years (Liu et al., 2018). A higher amount of myelin likely relates to more established pathways and, subsequently, faster communication between neurons.

In summary, for most used brain metrics "more is more developed". It may thus be perplexing to explain some of the reviewed findings where negative associations between cortical volumes and language skills have also been described (Spann et al., 2014). However, it is important to note that the cortex and the subcortical areas have differing patterns of development (Jha et al., 2019; Li et al., 2014), and so does the myelinating WM. It may be possible that cortical mass decreases as a result of myelination, as the WM progresses to and within the cortex (Croteau-Chonka et al., 2016; Deoni et al., 2015).

Poorly characterized background information of the participants poses some challenges to MRI studies in general (Pulli et al., 2018). In the reviewed studies, basic demographics, such as subjects' age at scan, gender distribution, recruitment process and exclusion criteria were well reported. However, detailed but very relevant information of, for example, ethnicity, parental health profile, educational history or family SES were often poorly reported. In the reviewed studies, no significant correlations were found between the SES scores and language skills or brain morphometrics in any of the five review articles that reported SES scores. However, it is reported earlier that the aspects of early environment are suggested to influence an infant's brain development especially during the pre- and postnatal periods, when a child is exposed to environmental factors mostly via maternal health and family changes (Betancourt et al., 2016; Hanson et al., 2013). In the studies that reported parental SES or educational level, the majority of the families were classified as middle class or higher with also a higher educational level (Deniz Can et al., 2013; Short et al., 2013; Spann et al., 2014; Swanson et al., 2015). It is common in MRI studies in general, that families with a higher educational level and SES are

more active in participating in studies. Considering these observations, we believe it is important for future studies to report adequate child and family demographics.

Language learning is one of the most striking developmental processes during the early years of life, and problems with language development may lead to social and cognitive difficulties later in life (Peterson & Pennington, 2012). It is thus both an interesting process in terms of mapping neural correlates of emerging cognition and an important avenue for translational studies (Langer et al., 2015; Weiss-Croft & Baldeweg, 2015). We identified a fair number of good quality studies that found numerous cerebral correlates of language development. However, across the studies there are no systematic neural correlates of language skills. This may be partly explained by a modest number of studies, scattered representation of ages in measurements and the variety of the methods used.

6.4 Strengths, limitations, and future directions

6.4.1 FinnBrain quality control protocol for semiautomated segmentation

The FinnBrain quality control protocol offers an instruction choice for manual editing that is easy to adopt, relatively time-efficient and aims to minimize interindividual bias. Errors caused by skull fragments are clear and easy to fix, and so are the errors caused by arteries by erasing the voxels between certain intensity values. The greatest effort and experience are required in the edits of WM-GM border. However, as the edits are followed by another automated segmentation that considers the human input in calculations, the editor cannot decide the delineation between WM and GM, and therefore cannot make errors that would mandate editing the image again from scratch. Such errors were possible while editing the borders between GM and SSS, which led to the exclusion of these edits after an interim analysis. While it could be argued that the effect on CT values is not worth the time that manual edits require, we believe that the systematic manual edits protocol has

an additional benefit: It increases the chance of finding and fixing regional errors that would have led to an exclusion, therefore increasing the number of valid datapoints in the final sample.

A key limitation in our protocol is the reliance on visual assessment in the quality control, which is now more based on individual experience rather than systematic evaluation. An automated, objective estimate of the severity of the motion might allow us to set universal standards on the different categories of motion severity. However, this kind of classification arouses fundamental questions: 1) How much the different levels of motion affect different aspects of brain morphology? and 2) Are the effects similar throughout the brain or are there significant regional differences? Considering these challenges, more research is needed before the effects of motion artefact can be analysed automatically. Another approach would be to lessen motion artefact by adding prospective motion correction (PMC) to the T1-weighted imaging sequence (Ai et al., 2021). The benefit is clearest in images with a lot of motion artefact, while the cost is poorer performance in some quality control measures, such as signal to noise ratio compared to a MPRAGE sequence without PMC (Ai et al., 2021). While implementation of PMC could improve the quality of our data, it would not solve the initial challenge, a need for systematic and objective assessment for motion artefacts. Therefore, the existence of this alternative imaging sequence does not impact our main findings. Although we opted for a quality control protocol that performs visual quality control on a level of individual ROIs, investigators may additionally benefit from using custom software to detect potentially low-quality data (Klapwijk et al., 2019).

One possible way of increasing the feasibility of anatomic segmentation in children is the creation and use of custom templates, which could produce more accurate results compared to the adult ones. To this day, some paediatric atlases have already been created (Phan et al., 2018), some including 5-year-olds (Fonov et al., 2011; Wilke et al., 2003). Certain typical errors can appear in unedited paediatric FreeSurfer images when applying adult templates, as presented in our work and in the literature (Phan et al., 2018). However, paediatric atlases have challenges that are characteristics for them. As the early brain development is non-linear and the developmental time intervals can be relatively short, the required specificity regarding the age range is unclear. Age-specific atlases have been created for neonates and infants (Kuklisova-Murgasova et al., 2011; Shi et al., 2011), with promising results of agreement compared to the "gold standard" manual segmentation (Oishi et al., 2011; Serag et al., 2012). The age ranges in these atlases may be very specific, covering also the preterm neonates starting from 29 gestational weeks of age (Kuklisova-Murgasova et al., 2011), for example. In older paediatric populations, the age range may be more permissive, from a few years to even more

than 10 years (Fonov et al., 2011; Wilke et al., 2003). Consequently, various paediatric atlases offer the multitude of different options, which may complicate comparisons between studies, the issue we have discussed earlier. Therefore, in our work, we decided to use the standard adult atlas with appropriate quality control measures to counter the challenges this approach has, and the results of cortical segmentation were satisfactory in general. However, for future aspects, it remains an important venue to develop and validate implemented in mainstream software such as FreeSurfer (de Macedo Rodrigues et al., 2015; Zöllei et al., 2020).

6.4.2 Environmental factors affecting brain morphology.

The strengths in Study II include the relatively large sample size with prospective design. The MRI data is analysed thoroughly with quality protocol according to Study I and, therefore, we provide state-of-the-art brain imaging data. Most demographics were collected from the national birth registry, which makes the variables reliable and standardized. Some variables were missing in our study compared to the reference work by Jha et al. (2019), but generally the environmental and obstetric variables matched satisfactorily.

A few limitations can be addressed. Some demographics were obtained from questionnaires completed by mothers during pregnancy (including paternal information), which creates minor but important reliability considerations. In addition, some variables included a modest sample size both absolutely and in relation to the whole data (e.g., maternal smoking and the use of medication during pregnancy). Despite the large variable pool, only a few of the candidate variables were found to predict cortical structure. However, the aim of this study was not to find the best predictive model for five-year-olds, but to explore similarities and differences of the variables in comparison with earlier infant studies.

The five-year time gap between the questionnaire data collection and imaging sessions was intended as per study design, as we were initially interested in whether a set of demographics that have been shown to be important determinants of infant brain structure have similar associations later in life. Full longitudinal data would be optimal for this purpose and such studies are warranted in the light of our results.
Even though we were satisfied with the matching of environmental and obstetric variables, the brain variables differed partly between the studies. We analysed cortical SA and volume, whereas Jha et al. (2019) studied CT and SA, which is in our view age appropriate. Most of these limitations can likely be addressed in large data sets such as the developing and baby Human Connectome Project (Eggebrecht

et al., 2018; Fenchel et al., 2020) and the Healthy Brain and Child Development Study (https://heal.nih.gov/research/infants-and-children/healthy-brain).

We inspected the influence of environmental factors and family demographics on cortical brain anatomy in healthy five-year-olds and examined whether our current findings are in line with similar studies carried out with newborns. The study was carried out to evaluate possible age-related and developmental differences in factors explaining the variance in brain cortical structure. We found that apart from child sex, different variables predicted cortical morphology in infants than in five-year-olds. This could be due to variation in methodology or study population characteristics, but it can also be interpreted as preliminary evidence on early childhood postnatal environmental factors influencing cortical development over prenatal factors as children mature. These findings also suggest that although closely related to each other, cortical SA and volumes have different developmental patterns with differential associations with observed variables. Overall, the effects of prenatal and early life variables on cortical development are not standardised in the current literature, and it is therefore important to report these factors if available. Future researchers will likely benefit from including similar environmental, parental and delivery related variables to sensitivity analyses in studies on cortical anatomy between 0–5 years of life, either by carrying out longitudinal studies or obtaining data retrospectively.

6.4.3 Language-related MRI studies

We provided a systematic literature search of the language-related MRI studies on the first two years of life, in which we sum up the current knowledge on the topic. Due to the modest number of the articles found, some exceptions were made. The study by O'Muircheartaigh et al., (2013) included participants aged from one to six years, which was over our age range from birth to two years. However, it was included in our review because there were a significant number of subjects within our scope of under two years of age. It was also justified to include studies that have dealt with cognitive development in general. Consequently, not all the reviewed studies were specifically interested in language outcomes, although the language scores were reported independently (Alcauter et al., 2014; O'Muircheartaigh et al., 2013, 2014; Short et al., 2013). Further, two functional MRI studies were also included despite the fact that the structural aspects of the language-related brain development were of high interest in the review. Due to methodological differences in both imaging techniques and analytical approaches, it is challenging to draw any

plausible synthesis of the findings. On the other hand, we are confident that the review will help investigators advance the field.

Future studies should seek to include repeated measures of both language skills and neuroimaging. For advancing the understanding of the MRI based metrics underlying language development, simultaneous measures of cortical structure and myelination would be helpful. Multimodal measures are needed to bridge the gap between relatively more advanced EEG studies of early language development and MR imaging – i.e., structure to function associations within brain metrics. Finally, one potential but previously unemployed study design would be to perform imaging after a certain level of language skills has been acquired, such as at the age of 12–18-months, when the subjects use a certain number of words or perform at a certain level in the used test scales. This would represent skill-based rather than age-based sampling, which would make it possible to assess what changes in the brain are crucial to a skill, and what parts are "a noise of normal brain growth".

7 Conclusions

The conclusions of the studies are the following:

I. With appropriate manual quality control, the automated segmentation FreeSurfer v6.0 tool can be reliably used to segment cortical brain areas of T1-weighted MRI images of healthy five-year-olds. The major challenges in the FreeSurfer output were detected in subcortical areas. We presented the semiautomated segmentation and quality control protocol for T1-weighted images that was applied in the FinnBrain Birth Cohort Study. The protocol is based on FreeSurfer v.6.0 and ENIGMA consortium tools. As there is a considerable methodological variation in paediatric neuroimaging studies, we are confident the presented quality control protocol will help this field of research ensure high-quality data and also make the results of the paediatric neuroimaging studies easier to compare between.

II. Males were found to have significantly larger cortical volume and SA than females in five-year-olds in line with prior reports. Maternal prenatal health profile and obstetric history variables associated with cortical development in the same time point. In comparison with the newborn MRI study, apart from child sex, different environmental factors were associated with cortical development in our study at the age of five years. The variance of the results between studies can be affected by different study population characteristics and methodology but could also suggest that the postnatal environment is a driving factor for cortical development rather than genetics. We found that the cortical lateralization at five years of age was mostly right-sided, except for temporal lobes showing left-sided asymmetry. This differed significantly from lateralization during infancy, where temporal lobes showed rightward and occipital and parietal lobes showed leftward asymmetry.

III. Many established components of the classic language-related cortical areas based on the adult brain were detected as predictors of language skills already during the first two years of life, including frontal and temporal regions (language-related Broca's and Wernicke's areas). Leftward asymmetry of the temporal regions was also reported, with considerable variation between studies. The development of the language-related WM trajectories showed also positive correlations with language skills. As there was a modest number of studies and great variance in the used methods, more research is still needed to determine systematic neural correlates of language skills and also to draw a consistent summary of the early life language-related cortical maturation.

Acknowledgements

I want to express my gratitude to my supervisors. Thank you, Jetro for the excellent guidance and encouragement throughout this process. Jetro has been my mentor since the early years of medical school and contributed significantly to all my publications. Obviously, this thesis would not have been completed without his crucial help. I want to thank Hasse and Linnea for letting me be a part of the FinnBrain study group. The way you have navigated this boat (which is more of a yacht today) called FinnBrain is something very special and admirable. Eeva-Leena, thank you for your supervision with the research work. Your help with the statistical analyses was crucial. I want to thank Jani Saunavaara and Jussi Tohka for your guidance in the follow-up group.

Research nurse Susanne along with Krisse and Ekaterina have been the ones responsible for the unique study visits that we have offered for the participating families. Their expertise resulted in high quality imaging data, which laid foundation to my thesis. All in all, I have been privileged to work with so many talented individuals in our research group, special thanks to all "FinnBrainers".

I want to thank associate professors Tuula Hurtig and Juha Salmitaival for their expertise in the preliminary examination. Your comments improved this thesis significantly.

I want to thank my peer PhD students Elmo, Venla and Anni. It's been a pleasant scientific journey with you from the very beginning from 2016 to this day. Working with you was easy even during the long Sunday evenings, which I do not take for granted. Special thanks to Elmo for your efforts with the imaging data and scientific reports, that are both applied in this thesis.

I also thank my boss Aleksi Lahti and other colleagues from the department of anesthesiology in Pori, for giving me time off from my clinical work during the past two years to finish this thesis.

Family comes first. I want to thank my parents Timo and Merja, and siblings Aino and Aaron for enabling optimal circumstances for "healthy, normal brain development". Special thanks to my father Timo for being an example that scientific interests can be combined with clinical work and, in fact, makes it even more meaningful. I also want to thank my uncle, aunts and cousins that have also been very

close to me. I want to express special gratitude to my late grandfather Lasse, who taught me the importance of pursuing dreams through hard work and persistence.

I have been privileged to share this life with so many dear friends, who I consider the greatest achievement in my life so far. As there are too many important ones from different occasions that simply cannot be all listed here, I want to highlight a few names especially regarding my academic efforts. Antti Saarinen and rest of the academic group *Pistiäiset*, thank you for showing an interest to my research work throughout the years. Matias Hilska, a research fellow from Itäharju, your feedback to my academic work and leading by example of how thesis work is carried out has been essential.

Lastly and most importantly, I want to thank Krista for her support to this work. As my life is and has been mostly very abundant and fruitful, yet also stressful and even anxious at some points, I am deeply grateful for and touched by the love and calmness you add to it. Thank you for sharing this journey with me.

I am grateful for the research funding from the following institutions: Department of Anesthesiology (Satasairaala Central Hospital, Satakunta Hospital District), Finnish Brain Foundation, Juho Vainio Foundation, Maire Taponen foundation, Signe and Ane Gyllenberg Foundation, Turunmaan Duodecim-seura, Turku University Foundation.

Pori, September 2023

Eero Silver

Appendices

Appendix A: Participant demographics and maternal medical history variables in Study I (n=146)

Continuous variables		Mean	Sd	Min	Max
Age from birth at scan (years)		5.35	0.06	5.08	5.52
Gestational age at scan (years)		6.11	0.07	5.85	6.33
Gestational age at birth (days)		278	12	237	296
Birth weight (grams)		3517	522	1790	4980
Maternal age at term (years)		31.2	4.7	18.3	42.0
Maternal BMI before pregnancy		24.2	4.3	17.5	42.0

Categorical variables	Number	Percent
Sex		
Male	79	54.1
Female	67	45.9
Maternal education level		
Upper secondary school or vocational school or lower	38	26.0
University of Applied Sciences	37	25.3
University	66	45.2
Missing	5	3.4
Maternal monthly income, estimated after taxes (euros)		
≤ 1 500	44	30.1
1 501 – 2 500	83	56.8
2 501 – 3 500	13	8.9
≥ 3 501	1	0.7
Missing	5	3.4
Maternal background		
Finnish	141	96.6
Other	5	3.4
Alcohol use during pregnancy		
Yes, continued to some degree after learning about	13	8.9
Yes, stopped after learning about the pregnancy	27	18.5
No	99	67.8
Missing	7	4.8
Tobacco smoking during pregnancy		
Yes, continued to some degree after learning about	3	2.1
Yes, stopped after learning about the pregnancy	7	4.8
No	129	88.4
Missing	7	4.8
Illicit drug use during pregnancy		
No	139	95.2
Missing	7	4.8
Maternal history of disease, yes (n = 141)		
Allergies	59	41.8
Depression	22	15.6
Asthma	13	9.2

Anxiety disorder	13	9.2
Eating disorder	12	8.5
Chronic urinary tract infection	10	7.1
Autoimmune disorder	9	6.4
Hypercholesterolemia	4	2.8
Hypertension	4	2.8
Celiac disease	2	1.4
Hypothyroidism	2	1.4
Emphysema	1	0.7
Chronic bacterial or viral infection	1	0.7
Psychosis	1	0.7
Epilepsy	1	0.7
Type 2 diabetes	1	0.7
Drug dependency	1	0.7
Migraine	1	0.7
Other chronic disease	6	4.3
Maternal medication at gestational week 14, yes (n = 135)		
Thyroxin	11	8.1
SSRI/SNRI	6	4.4
Corticosteroid	5	3.7
Hypertension medication	2	1.5
Other mood medication	2	1.5
Other medication affecting the CNS	2	1.5
Other medication	9	6.7
Maternal medication at gestational week 34, yes (n = 140)		
Thyroxin	9	6.4
SSRI/SNRI	7	5.0
Corticosteroid	6	4.3
Blood pressure medication	6	4.3
NSAID	1	0.7
Other mood medication	2	1.4
Other medication affecting the CNS	3	2.1
Other medication	27	19.3

Abbreviations: SD = standard deviation, BMI = body mass index, NSAID = non-steroidal anti-inflammatory drug, SSRI = selective serotonin reuptake inhibitor, SNRI = selective noradrenalin reuptake inhibitor, CNS = central nervous system, GA = gestational age.

Appendix B: Regression models for total cortical surface area and volumes of Study II.

Appendix Table 1. Regression model for total cortical surface area (original Table 4 in Study II.)

Region of interest	**Predictors**	**Unstandardized Coefficients**		**Standardized Coefficients**		
		B	Std.err	Beta	t	p
Total	ΔR2 = 0.242					
	Intercept	186097.140	1721.1		108.123	<0.001
	Sex	-13180.327	2474.3	-0.463	-5.327	<0.001
	NICU admission	8766.857	4340.7	0.175	2.020	0.046
	Excluded predictors			Beta In	t	p
	Ponderal Index at MRI			0,111	1,284	0,202
	Gestational age at birth			-0,049	-0,557	0,579
	Maternal age at child's birth			0,096	1,093	0,277
	Paternal age at child's birth			0,068	0,779	0,438
	Maternal BMI before gestation			0,077	0,87	0,386
	Birth weight			0,026	0,3	0,765
	Apgar 5min			-0,113	-1,171	0,244
	SSRI/SNRI medication during gestation			0,060	0,695	0,489
	Diagnosis of gestation diabetes			-0,013	-0,142	0,887
	Postnatal age at MRI			0,091	1,035	0,303
	Maternal smoking during gestation			-0,165	-1,904	0,06
	Maternal education			0,040	0,462	0,645
	Paternal education			-0,051	-0,582	0,562
	Mode of delivery			0,058	0,662	0,509

BMI= body mass index, MRI= magnetic resonance imaging, NICU= neonatal intensive care unit

Appendix Table 2. Regression models for total cortical volume (original table 5 in Study II)

Region of interest	**Predictor**	**Unstandardized coefficients**		**Standardized coefficients**	.	
		B	Std.error	Beta	T	P
Total volume	Δr2 = 0.176					
	Intercept	594860.313	13058.506		45.553	<0.001
	Sex	-31553.852	7745.375	-0.369	-4.074	<0.001
	Maternal education	18052.615	7713.996	0.212	2.340	0.021
	Excluded predictors			Beta in	T	P
	Ponderal index at MRI			0,089	0,983	0,328
	Gestational age at birth			-0,079	-0,859	0,392
	Maternal age at child's birth			0,169	1,863	0,065
	Paternal age at child's birth			0,114	1,221	0,225
	Maternal bmi before gestation			0,048	0,533	0,596
	Birth weight			-0,011	-0,117	0,907
	Apgar 5min			-0,052	-0,574	0,567
	NICU admission			0,109	1,205	0,231
	SSRI/SNRI medication during gestation			0,039	0,428	0,669
	Diagnosis of gestation diabetes			0,026	0,291	0,771
	Postnatal age at MRI			0,100	1,108	0,270
	Maternal smoking during gestation			-0,112	-1,23	0,221
	Paternal education			-0,117	-1,21	0,229
	Mode of delivery			0,093	1,024	0,308

BMI = body mass index, MRI = magnetic resonance imaging, NICU = neonatal intensive care unit, SSRI/SNRI = serotonin and norepinephrine reuptake inhibitors

Appendix C: Tables of the reviewed articles of Study III

Appendix Table 3. Summary of the reviewed structural MRI studies (original table 1 in Study III)

Region of interest	Outcome metrics	Scan resolution	Field Strength (Tesla,T)	MRI sequences	Age at scan	Study design (Imaging)	Excluded *N*	*N*	STUDY
Whole brain	Cortical volumes	1 x 1 x 1 mm^3	3.0T	T2	1-6 weeks	Cross-sectional	15	33 infants (14f, 19m)	Spann et al., 2014
Whole brain	Tissue density (VBM)	0.94 x 0.94 x 1 mm^3	1.5T	T1	7 months	Cross-sectional	10	19 infants (12f, 7m)	Deniz can et al., 2013
Whole brain	Whole brain VFM maps, quantitative	N/A	N/A	T1, T2	2-12 months (n=58), 12-24 months	Cross-sectional	N/A	183 infants (74f, 109m), 112 infants	O' Muirchertaigh et al., 2014
Whole brain	Whole brain VFM maps, quantitative	1.8 × 1.8 × 1.8 mm^3	N/A	T2	0-2 years: 6-month intervals, 2-	Longitudinal and cross-sectional	N/A	257 infants (109f, 148m) from 3	Deoni et al., 2015
Whole brain, asymmetry	Whole brain VFM maps, quantitative	1.7 × 1.7 × 1.7 mm^3	1.5T	T1, T2	1-6 years (mean 3.11 years)	Cross-sectional	N/A	108 infants (48f, 60m), 34 infants	O'Muirchertaigh et al., 2013

GM = grey matter, MSEL = mullen scales of early learning, SES = sosioeconomic status, VBM = voxel based morphometry, VFM = myelin volume fraction, WM = white matter

Appendix Table 3. Summary of the reviewed structural MRI studies (original table 1 in Study III) (continued)

Information about environmental surroundings	Main Results	Timing of cognitive tests, type of correlations	Study design (cognitive tests)	Cognitive tests	STUDY
Maternal info: age, race, years of education, SES, IQ, and previous deliveries	Posterior parietal regions and cingulate gyrus associated with subsequent language scores, and the most significant correlations were inverse. Inverse correlations between subsequent language scores and posterior parietal regions at 6, 12 and 24 months, and bilateral cingulate regions from 12 to 24 months.	6, 12, 18 and/or 24 months, concurrent correlations	Longitudinal	Bayley-III	Spann et al., 2014
Family SES	Strong positive correlations between GM intensities and language abilities in right posterior cerebellum and right hippocampus. Positive correlations to language skills and WM intensities in the right inferior cerebellar peduncle and left posterior limb of the internal capsule and cerebral peduncle	12 months, predictive correlations	Cross-sectional	MSEL	Deniz can et al., 2013
No	Temporal and frontal regions showed significant interactions with cognitive tests specific to expressive and receptive language scores, from 3 months to 4 years. No correlation were found in other cognitive scales.	within 1 week of MRI acquisition, concurrent correlations	Cross-sectional	MSEL	O' Muircheartaigh et al., 2014
Ethnicity, maternal SES	Cross-sectionally, significant positive correlations between MWF and MSEL cognitive scores in the splenium. Longitudinally, no significant correlations during the first year. Positive correlations between MWF and MSEL language scores in the left Broca' s area across the cohort (P <0.09), and in the right temporal WM postcentral gyrus, and supramarginal gyrus (P < 0.1)	within 1 week of the MRI acquisition, concurrent correlations	Longitudinal	MSEL,	Deoni et al., 2015
No	leftward asymmetry in anterior caudate, thalamus, medial frontal, posterior parietal and partly in temporal and occipital lobe. Rightward asymmetry in dorsal extreme capsule, white matter under lateral motor cortex and ventral frontal cortex. Age and gender did not have an effect on asymmetries. Leftward asymmetry in anterior thalamus and caudate correlated positively with receptive language and visual reception scores.	within 1 week of the MRI acquisition, concurrent correlations	Cross-sectional	MSEL	O'Muircheartaigh et al., 2013

GM = grey matter, MSEL= Mullen scales of early learning, SES = sosioeconomic status, VBM = voxel-based morphometry, VFM = Myelin water fraction, WM = white matter

Appendix Table 4. Summary of the reviewed functional MRI studies (original Table 2 in Study III)

Region of interest	Outcome metrics	Scan resolution	Field Strength (Tesla,T)	MRI sequences	Age at scan	Study design (Imaging)	Excluded *N*	*N*	Study
IFG, STG, intraparietal sulcus, dorsolateral prefrontal cortex, lingual and postcentral gyrus.	tissue density (VBM), functional connectivity	4 x 4 x 4 mm^3	3.0T	fMRI	1,3,6,9, 12,18 and 24 months	Longitudinal	N/A	71 infants (37f 34m) imaged, 23 infants cognitively tested	Emerson et al., 2016
9 cortical networks arising from thalamus	Tissue density (VBM), BOLD responses	4 x 4 x 4 mm^3	3.0T	fMRI	neonates, 1 and 2 years	Longitudinal	N/A	143 infants (70f, 77m)	Alcauter et al., 2014

BOLD = blood oxygen level-dependent, fMRI = functional magnetic resonance imaging, IFG = inferior frontal gyrus, STG = superior temporal gyrus, VBM = voxel based morphometry,

Appendix Table 4. Summary of the reviewed functional MRI studies (original Table 2 in Study III) (continued)

Information about environmental surroundings	Main Results	Timing of cognitive tests, type of correlations	Study design (cognitive tests)	Cognitive tests	STUDY
No	In IFG and STG, the asymmetrisation took place at the end of the first year and continued into the second year. Increased asymmetry was accompanied by increased unilateral connectivity between language regions. Early transition toward asymmetry correlated positively to language outcome scores in the IFG, but not in the STG	4 years, predictive correlations	Cross-sectional	MSEL	Emerson et al., 2016
No	All higher-order networks (lateral visual, default mode, and bilateral frontoparietal networks) developed most dynamically in the first year, with minor refinement in the second year. Thalamus–sensorimotor and thalamus–salience networks exist already in neonates, whereas thalamus–medial–visual and thalamus–default mode networks emerge in 1-year and 2-year-olds. Thalamus-salience network correlated with cognitive measures at 2 years of age.	1 and 2 years, predictive correlations	Longitudinal	MSEL, the Mullen Early Learning Composite Standard Score (ELCSS)	Alcauter et al., 2014

IFG = inferior frontal gyrus, MSEL= Mullen scales of early learning, STG = superior temporal gyrus

Appendix Table 5. Summary of the reviewed DTI studies (original Table 3 in Study III)

Region of interest	Outcome metrics	Scan resolution	Field Strength (Tesla,T)	MRI sequences	Age at scan	Study design (imaging)	Excluded *N*	*N*	Study
Whole brain	FA	1.8 x 1.8 x 1.8 mm^3	1.5T	DTI and T1	14.3 ± 1.6 days	Cross-sectional	8 infants	38 infants (18f, 20m)	Feng et al., 2019
AF	FA	2.0 × 2.0 × 2.0 mm^3	3.0T	DTI and T1	FHD+: 333 ± 118 days, FHD-: 298 ± 99 days	Cross-sectional	N/A	32 infants (17f, 15m). 14 infants with FHD+, 18 infants with no risk (FHD-).	Langer et al., 2015
Corpus callosum, AF, UF and ILF	FA	2.0 x 2.0 x 2.0 mm^3	3.0T	DTI	6, 12 and 24 months	Longitudinal	N/A	77 infants (32f, 45m)	Swanson et al., 2015

AF = arcuate fasciculus, DTI = diffusion tensor imaging, FA = fractional anisotropy, FHD = high familial risk of dyslexia, ILF= inferior longitudinal fasciculus, M-CDI = Macarthur-Bates communicative development inventory, UF = uncinate fasciculus, VASB-II = the Vineland adaptive behavior scales-ii motor subscale score.

Appendix Table 5. Summary of the reviewed DTI studies (original table 3 in Study III) (continued).

Information about environmental surroundings	Main results	Timing of cognitive tests, type of correlations	Study design (cognitive tests)	Cognitive tests	Study
Comprehensive information: parental IQ, education, income	Positive correlations between FA values and language scores in the left peduncle, right optic radiation, left and right radiation of the corpus callosum, and genu of the corpus callosum.	2 years, predictive correlations	Cross-sectional	Bayley-III	Feng et al., 2019
Comprehensive information: the highest degree, income, homeowner status, current activity (working etc.) and home literacy	Individuals with FHD+ had lower mean FA in left AF in comparison with FHD- subjects. In addition, FA values of left AF correlated positively with MSEL Expressive language subtest scores.	Simultaneously with scannings, concurrent correlations	Cross-sectional	MSEL (visual reception and receptive language subscales excluded)	Langer et al., 2015
Maternal and paternal age, educational level and family household income	Positive correlations between language production and FA values in corpus callosum and in right ILF from 6 to 24 months. FA values did not correlate to language production at 6 months, but by 24 months higher language production showed greater FA values. No significant associations between FA values and language production in bilateral AF, UF or left ILF.	MSEL : 6, 12 and 24 months, M-CDI: 24 months, VASB-II: 24 months, concurrent correlations	Longitudinal	MSEL , M-CDI and VASB-II.	Swanson et al., 2015

AF = arcuate fasciculus, DTI = diffusion tensor imaging, FA = fractional anisotropy, FHD = high familial risk of dyslexia, ILF = inferior longitudinal fasciculus, M-CDI = Macarthur-Bates communicative development inventory, UF = uncinate fasciculus, VASB-II = the Vineland adaptive behavior scales-ii motor subscale score.

References

Ai, L., Craddock, R. C., Tottenham, N., Dyke, J. P., Lim, R., Colcombe, S., Milham, M., & Franco, A. R. (2021). Is it time to switch your T1W sequence? Assessing the impact of prospective motion correction on the reliability and quality of structural imaging. *NeuroImage*, *226*. https://doi.org/10.1016/j.neuroimage.2020.117585

al Harrach, M., Rousseau, F., Groeschel, S., Wang, X., Hertz-pannier, L., Chabrier, S., Bohi, A., Lefevre, J., & Dinomais, M. (2019). Alterations in Cortical Morphology after Neonatal Stroke: Compensation in the Contralesional Hemisphere? *Developmental Neurobiology*, *79*(4), 303–316. https://doi.org/10.1002/dneu.22679

Alcauter, S., Lin, W., Smith, J. K., Short, S. J., Goldman, B. D., Reznick, J. S., Gilmore, J. H., & Gao, W. (2014). Development of Thalamocortical Connectivity during Infancy and Its Cognitive Correlations. *Journal of Neuroscience*, *34*(27), 9067–9075. https://doi.org/10.1523/JNEUROSCI.0796-14.2014

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

Allen, J. S., Damasio, H., Grabowski, T. J., Bruss, J., & Zhang, W. (2003). Sexual dimorphism and asymmetries in the gray-white composition of the human cerebrum. *NeuroImage*, *18*(4), 880–894. https://doi.org/10.1016/S1053-8119(03)00034-X

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

Barnes-Davis, M. E., Williamson, B. J., Merhar, S. L., Holland, S. K., & Kadis, D. S. (2020). Extremely preterm children exhibit altered cortical thickness in language areas. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-67662-7

Beelen, C., Phan, T. V., Wouters, J., Ghesquière, P., & Vandermosten, M. (2020). Investigating the Added Value of FreeSurfer's Manual Editing Procedure for the Study of the Reading Network in a Pediatric Population. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00143

Betancourt, L. M., Avants, B., Farah, M. J., Brodsky, N. L., Wu, J., Ashtari, M., & Hurt, H. (2016). Effect of socioeconomic status (SES) disparity on neural development in female African-American infants at age 1 month. *Developmental Science*, *19*(6), 947–956. https://doi.org/10.1111/desc.12344

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*. https://doi.org/10.1038/s41586-022-04554-y

Binder, J. et al., (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex Mon* 2000;10:512–528; 1047–3211/00

Blumenthal, J. D., Zijdenbos, A., Molloy, E., & Giedd, J. N. (2002). Motion artifact in magnetic resonance imaging: Implications for automated analysis. *NeuroImage*, *16*(1), 89–92. https://doi.org/10.1006/nimg.2002.1076

Botdorf, M., & Riggins, T. (2018). When less is more: Thinner fronto-parietal cortices are associated with better forward digit span performance during early childhood. *Neuropsychologia*, *121*, 11–18. https://doi.org/10.1016/j.neuropsychologia.2018.10.020

Boutzoukas, E. M., Crutcher, J., Somoza, E., Sepeta, L. N., You, X., Gaillard, W. D., Wallace, G. L., & Berl, M. M. (2020). Cortical thickness in childhood left focal epilepsy: Thinning beyond the seizure focus. *Epilepsy and Behavior*, *102*. https://doi.org/10.1016/j.yebeh.2019.106825

Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). Dorsal and ventral pathways in language development. *Brain and Language*, *127*(2), 289–295. https://doi.org/10.1016/j.bandl.2013.03.001

Brain Development Cooperative Group. (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

Capron, L. E., Glover, V., Pearson, R. M., Evans, J., O'Connor, T. G., Stein, A., Murphy, S. E., & Ramchandani, P. G. (2015). Associations of maternal and paternal antenatal mood with offspring anxiety disorder at age 18 years. *Journal of Affective Disorders*, *187*, 20–26. https://doi.org/10.1016/j.jad.2015.08.012

Catani, M., Jones, D. K., Daly, E., Embiricos, N., Deeley, Q., Pugliese, L., Curran, S., Robertson, D., & Murphy, D. G. M. (2008). Altered cerebellar feedback projections in Asperger syndrome. *NeuroImage*, *41*(4), 1184–1191. https://doi.org/10.1016/j.neuroimage.2008.03.041

Cattarinussi, G., Aarabi, M. H., Sanjari Moghaddam, H., Homayoun, M., Ashrafi, M., Soltanian-Zadeh, H., & Sambataro, F. (2021). Effect of parental depressive symptoms on offspring's brain structure and function: A systematic review of neuroimaging studies. In *Neuroscience and Biobehavioral Reviews* (Vol. 131, pp. 451–465). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2021.09.046

Choe, M. S., Ortiz-Mantilla, S., Makris, N., Gregas, M., Bacic, J., Haehn, D., Kennedy, D., Pienaar, R., Caviness, V. S., Benasich, A. A., & Ellen Grant, P. (2013). Regional infant brain development: An MRI-based morphometric analysis in 3- to 13-month-olds. *Cerebral Cortex*, *23*(9), 2100–2117. https://doi.org/10.1093/cercor/bhs197

Copeland, A., Silver, E., Korja, R., Lehtola, S. J., Merisaari, H., Saukko, E., Sinisalo, S., Saunavaara, J., Lähdesmäki, T., Parkkola, R., Nolvi, S., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2021). Infant and Child MRI: A Review of Scanning Procedures. In *Frontiers in Neuroscience* (Vol. 15). Frontiers Media S.A. https://doi.org/10.3389/fnins.2021.666020

Croteau-Chonka, E. C., Dean, D. C., Remer, J., Dirks, H., O'Muircheartaigh, J., & Deoni, S. C. L. (2016). Examining the relationships between cortical maturation and white matter myelination throughout early childhood. *NeuroImage*, *125*, 413–421. https://doi.org/10.1016/j.neuroimage.2015.10.038

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). *Cortical Surface-Based Analysis I. Segmentation and Surface Reconstruction*. http://www.idealibrary.com

de Macedo Rodrigues, K., Ben-Avi, E., Sliva, D. D., Choe, M. S., Drottar, M., Wang, R., Fischl, B., Grant, P. E., & Zöllei, L. (2015). Afreesurfer-compliantconsistentmanualsegmentationofinfantbrainsspanningthe0–2yearagerange. *Frontiers in Human Neuroscience*, *9*(FEB). https://doi.org/10.3389/fnhum.2015.00021

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*, *223*(4), 1953–1970. https://doi.org/10.1007/s00429-017-1600-2

DeCasper, A., & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voices. *Science*, *208*(June), 1174–1176.

del Tufo, S. N., Earle, F. S., & Cutting, L. E. (2019). The impact of expressive language development and the left inferior longitudinal fasciculus on listening and reading comprehension. *Journal of Neurodevelopmental Disorders*, *11*(1). https://doi.org/10.1186/s11689-019-9296-7

Dellatolas, G., & Câmara-Costa, H. (2020). The role of cerebellum in the child neuropsychological functioning. *Handbook of Clinical Neurology*, *173*, 265–304. https://doi.org/10.1016/B978-0-444-64150-2.00023-X

Deniz Can, D., Richards, T., & Kuhl, P. K. (2013). Early gray-matter and white-matter concentration in infancy predict later language skills: A whole brain voxel-based morphometry study. *Brain and Language*, *124*(1), 34–44. https://doi.org/10.1016/j.bandl.2012.10.007

Deoni, S. C. L., Dean, D. C., Remer, J., Dirks, H., & O'Muircheartaigh, J. (2015). Cortical maturation and myelination in healthy toddlers and young children. *NeuroImage*, *115*, 147–161. https://doi.org/10.1016/j.neuroimage.2015.04.058

Donald, K. A., Fouche, J. P., Roos, A., Koen, N., Howells, F. M., Riley, E. P., Woods, R. P., Zar, H. J., Narr, K. L., & Stein, D. J. (2016). Alcohol exposure in utero is associated with decreased gray matter volume in neonates. *Metabolic Brain Disease*, *31*(1), 81–91. https://doi.org/10.1007/s11011-015-9771-0

Dosman, C. F., Faap, F., Andrews, D., & Goulden, K. J. (2012). Evidence-based milestone ages as a framework for developmental surveillance. In *Paediatr Child Health* (Vol. 17). www.mchatscreen.com

Duan, C., Hare, M. M., Staring, M., & Deligiannidis, K. M. (2019). Examining the relationship between perinatal depression and neurodevelopment in infants and children through structural and functional neuroimaging research. In *International Review of Psychiatry* (Vol. 31, Issue 3, pp. 264–279). Taylor and Francis Ltd. https://doi.org/10.1080/09540261.2018.1527759

Dubois, J., Dehaene-Lambertz, G., Kulikova, S., Poupon, C., Hüppi, P. S., & Hertz-Pannier, L. (2014). The early development of brain white matter: A review of imaging studies in fetuses, newborns and infants. In *Neuroscience* (Vol. 276, pp. 48–71). Elsevier Ltd. https://doi.org/10.1016/j.neuroscience.2013.12.044

Dubois, J., Poupon, C., Thirion, B., Simonnet, H., Kulikova, S., Leroy, F., Hertz-Pannier, L., & Dehaene-Lambertz, G. (2016). Exploring the Early Organization and Maturation of Linguistic Pathways in the Human Infant Brain. *Cerebral Cortex*, *26*(5), 2283–2298. https://doi.org/10.1093/cercor/bhv082

Duerden, E. G., Chakravarty, M. M., Lerch, J. P., & Taylor, M. J. (2020). Sex-Based Differences in Cortical and Subcortical Development in 436 Individuals Aged 4-54 Years. *Cerebral Cortex*, *30*(5), 2854–2866. https://doi.org/10.1093/cercor/bhz279

Eggebrecht, A. T., Elison, J. T., Feczko, E., Todorov, A., Wolff, J. J., Kandala, S., Adams, C. M., Snyder, A. Z., Lewis, J. D., Estes, A. M., Zwaigenbaum, L., Botteron, K. N., Mckinstry, R. C., Constantino, J. N., Evans, A., Hazlett, H. C., Dager, S., Paterson, S. J., Schultz, R. T., … Pruett, J. R. (2018). Joint Attention and Brain Functional Connectivity in Infants and Toddlers. *March 2017*, 1709–1720. https://doi.org/10.1093/cercor/bhw403

el Marroun, H., Tiemeier, H., Franken, I. H. A., Jaddoe, V. W. V., van der Lugt, A., Verhulst, F. C., Lahey, B. B., & White, T. (2016). Prenatal Cannabis and Tobacco Exposure in Relation to Brain Morphology: A Prospective Neuroimaging Study in Young Children. *Biological Psychiatry*, *79*(12), 971–979. https://doi.org/10.1016/j.biopsych.2015.08.024

Emerson, R. W., Gao, W., & Lin, W. (2016). Longitudinal study of the emerging functional connectivity asymmetry of primary language regions during infancy. *Journal of Neuroscience*, *36*(42), 10883–10892. https://doi.org/10.1523/JNEUROSCI.3980-15.2016

Encyclopædia Britannica, 2023. https://www.britannica.com/science/Wernicke area#/media/1/639879/100577

Epstein, J. N., Casey, B. J., Tonev, S. T., Davidson, M., Reiss, A. L., Garrett, A., Hinshaw, S. P.,

Greenhill, L. L., Vitolo, A., Kotler, L. A., Jarrett, M. A., & Spicer, J. (n.d.). *Assessment and Prevention of Head Motion During Imaging of Patients with Attention Deficit Hyperactivity Disorder*.

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility*, *44*, 33–43. https://doi.org/10.1016/j.rssm.2016.01.003

Fenchel, D., Dimitrova, R., Seidlitz, J., Robinson, E. C., Batalle, D., Hutter, J., Christiaens, D., Pietsch, M., Brandon, J., Hughes, E. J., Allsop, J., O'Keeffe, C., Price, A. N., Cordero-Grande, L., Schuh, A., Makropoulos, A., Passerat-Palmbach, J., Bozek, J., Rueckert, D., … O'Muircheartaigh, J. (2020). Development of Microstructural and Morphological Cortical Profiles in the Neonatal Brain. *Cerebral Cortex*, *30*(11), 5767–5779. https://doi.org/10.1093/cercor/bhaa150

Feng, K., Rowell, A. C., Andres, A., Bellando, B. J., Lou, X., Glasier, C. M., Ramakrishnaiah, R. H., Badger, T. M., & Ou, X. (2019). Diffusion tensor MRI of white matter of healthy full-term newborns: Relationship to neurodevelopmental outcomes. *Radiology*, *292*(1), 179–187. https://doi.org/10.1148/radiol.2019182564

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

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, *89*(2), 267–276. https://doi.org/10.1016/S0093-934X(03)00351-1

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

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. M. (2006). Working memory in children with reading disabilities. *Journal of Experimental Child Psychology*, *93*(3), 265–281. https://doi.org/10.1016/j.jecp.2005.08.003

Gaudet, I., Hüsser, A., Vannasing, P., & Gallagher, A. (2020). Functional Brain Connectivity of Language Functions in Children Revealed by EEG and MEG: A Systematic Review. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00062

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., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. In *J. Comput. Assist. Tomogr* (Vol. 51, Issue 2). Plenum. http://neurosci.nature.com

Gilmore, J. H., Langworthy, B., Girault, J. B., Fine, J., Jha, S. C., Kim, S. H., Cornea, E., & Styner, M. (2020). Individual Variation of Human Cortical Structure Is Established in the First Year of Life. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(10), 971–980. https://doi.org/10.1016/j.bpsc.2020.05.012

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

Goldstein JM, L J Seidman, N J Horton, N Makris, & D N Kennedy, V. S. C. J. S. V. F. M. T. T. (2001). Normal Sexual Dimorphism of the Adult Human Brain Assessed by In Vivo

Magnetic Resonance Imaging. *Cereb Cortex*, *11*(6), 490–497.

Greene, D. J., Black, K. J., & Schlaggar, B. L. (2016). Considerations for MRI study design and implementation in pediatric and clinical populations. *Developmental Cognitive Neuroscience*, *18*, 101–112. https://doi.org/10.1016/j.dcn.2015.12.005

Hanson, J. L., Hair, N., Shen, D. G., Shi, F., Gilmore, J. H., Wolfe, B. L., & Pollak, S. D. (2013). Family poverty affects the rate of human infant brain growth. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0080954

Hazlett, H. C., Poe, M., Gerig, G., Styner, M., Chappell, C., Smith, R. G., Vachet, C., & Piven, J. (2012). Early Brain Overgrowth in Autism Associated with an Increase in Cortical Surface Area Before Age 2. *Bone*, *23*(1), 1–7. https://doi.org/10.1001/archgenpsychiatry.2011.39.Early

Heck, D., & Sultan, F. (2002). Cerebellar structure and function: Making sense of parallel fibers. In *Human Movement Science* (Vol. 21). www.elsevier.com/locate/humov

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(May), 1–10. papers2://publication/uuid/2AD83C67-014F-4AD1-B383-7F2EDF3D58F7

Highnam, C. L., & Bleile, K. M. (2011). Language in the Cerebellum. *American Journal of Speech-Language Pathology*, *20*(4), 337–347. https://doi.org/10.1044/1058-0360(2011/10-0096)

Innocenti, G. M., & Price, D. J. (2005). Exuberance in the development of cortical networks. In *Nature Reviews Neuroscience* (Vol. 6, Issue 12, pp. 955–965). https://doi.org/10.1038/nrn1790

Jha, S. C., Xia, K., Ahn, M., Girault, J. B., Li, G., Wang, L., Shen, D., Zou, F., Zhu, H., Styner, M., Gilmore, J. H., & Knickmeyer, R. C. (2019). Environmental influences on infant cortical thickness and surface area. *Cerebral Cortex*, *29*(3), 1139–1149. https://doi.org/10.1093/cercor/bhy020

Jolly, M., Sebire, N., Harris, J., Robinson, S., & Regan, L. (2000). The risks associated with pregnancy in women aged 35 years or older. *Human Reproduction*, *15*(11), 2433–2437. https://doi.org/10.1093/humrep/15.11.2433

Kamson, D. O., Pilli, V. K., Asano, E., Jeong, J. W., Sood, S., Juhász, C., & Chugani, H. T. (2016). Cortical thickness asymmetries and surgical outcome in neocortical epilepsy. *Journal of the Neurological Sciences*, *368*, 97–103. https://doi.org/10.1016/j.jns.2016.06.065

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

Klapwijk, E. T., van de Kamp, F., van der Meulen, M., Peters, S., & Wierenga, L. M. (2019). Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *NeuroImage*, *189*, 116–129. https://doi.org/10.1016/j.neuroimage.2019.01.014

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, R. C., Xia, K., Lu, Z., Ahn, M., Jha, S. C., Zou, F., Zhu, H., Styner, M., & Gilmore, J. H. (2017). Impact of Demographic and Obstetric Factors on Infant Brain Volumes: A Population Neuroscience Study. *Cerebral Cortex*, *27*(12), 5616–5625. https://doi.org/10.1093/cercor/bhw331

Koutra, K., Chatzi, L., Bagkeris, M., Vassilaki, M., Bitsios, P., & Kogevinas, M. (2013). Antenatal and postnatal maternal mental health as determinants of infant neurodevelopment at 18 months of age in a mother-child cohort (Rhea Study) in Crete, Greece. *Social Psychiatry and Psychiatric Epidemiology*, *48*(8), 1335–1345. https://doi.org/10.1007/s00127-012-0636-0

Koziol, L. F., Budding, D., Andreasen, N., D'Arrigo, S., Bulgheroni, S., Imamizu, H., Ito, M., Manto, M., Marvel, C., Parker, K., Pezzulo, G., Ramnani, N., Riva, D., Schmahmann, J., Vandervert, L., & Yamazaki, T. (2014). Consensus paper: The cerebellum's role in movement and

cognition. In *Cerebellum* (Vol. 13, Issue 1, pp. 151–177). Springer New York LLC. https://doi.org/10.1007/s12311-013-0511-x

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. https://doi.org/10.1111/j.1467-7687.2006.00468.x

Kuhl, U., Friederici, A. D., Emmrich, F., Brauer, J., Wilcke, A., Neef, N., Boltze, J., Skeide, M., Kirsten, H., Schaadt, G., Müller, B., Kraft, I., Czepezauer, I., Dörr, L., & Skeide, M. A. (2020). Early cortical surface plasticity relates to basic mathematical learning. *NeuroImage*, *204*. https://doi.org/10.1016/j.neuroimage.2019.116235

Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S. J., Doria, V., Serag, A., Gousias, I. S., Boardman, J. P., Rutherford, M. A., Edwards, A. D., Hajnal, J. v., & Rueckert, D. (2011). A dynamic 4D probabilistic atlas of the developing brain. *NeuroImage*, *54*(4), 2750–2763. https://doi.org/10.1016/j.neuroimage.2010.10.019

Kumpulainen, V., Lehtola, S. J., Tuulari, J. J., Silver, E., Copeland, AnniKorja, R., Karlsson, H., Karlsson, L., Merisaari, H., Parkkola, R., Saunavaara, J., Lähdesmäki, T., & Scheinin, N. M. (2020). Prevalence and Risk Factors of Incidental Findings in Brain MRIs of Healthy Neonates—The FinnBrain Birth Cohort Study. *Frontiers in Neurology*, *10*(January). https://doi.org/10.3389/fneur.2019.01347

Kuperberg, G. R., Broome, M. R., McGuire, P. K., David, A. S., Eddy, M., Ozawa, F., Goff, D., Caroline West, W., R Williams, S. C., W van der Kouwe, A. J., Salat, D. H., Dale, A. M., & Fischl, B. (2003). Regionally Localized Thinning of the Cerebral Cortex in Schizophrenia. *Arch Gen Psychiatry* Sep;60(9):878-88

Lambers, D. S., & Clark, K. E. (1996). The maternal and fetal physiologic effects of nicotine. *Seminars in Perinatology*, *20*(2), 115–126. https://doi.org/10.1016/S0146-0005(96)80079-6

Langer, N., Peysakhovich, B., Zuk, J., Drottar, M., Sliva, D. D., Smith, S., Becker, B. L. C., Grant, P. E., & Gaab, N. (2015). White Matter Alterations in Infants at Risk for Developmental Dyslexia. *Cerebral Cortex*, *February*, bhv281. https://doi.org/10.1093/cercor/bhv281

Lautarescu, A., Craig, M. C., & Glover, V. (2020). Prenatal stress: Effects on fetal and child brain development. In *International Review of Neurobiology* (Vol. 150, pp. 17–40). Academic Press Inc. https://doi.org/10.1016/bs.irn.2019.11.002

Lebel, C., & Deoni, S. (2018). The development of brain white matter microstructure. In *NeuroImage* (Vol. 182, pp. 207–218). Academic Press Inc. https://doi.org/10.1016/j.neuroimage.2017.12.097

Lehtola, S. J., Tuulari, J. J., Karlsson, L., Parkkola, R., Merisaari, H., Saunavaara, J., Lähdesmäki, T., Scheinin, N. M., & Karlsson, H. (2019). Associations of age and sex with brain volumes and asymmetry in 2–5-week-old infants. *Brain Structure and Function*, *224*(1), 501–513. https://doi.org/10.1007/s00429-018-1787-x

Lehtola, S. J., Tuulari, J. J., Scheinin, N. M., Karlsson, L., Parkkola, R., Merisaari, H., Lewis, J. D., Fonov, V. S., Louis Collins, D., Evans, A., Saunavaara, J., Hashempour, N., Lähdesmäki, T., Acosta, H., & Karlsson, H. (2020). Newborn amygdalar volumes are associated with maternal prenatal psychological distress in a sex-dependent way. *NeuroImage: Clinical*, *28*. https://doi.org/10.1016/j.nicl.2020.102380

Leonard, J. A., Romeo, R. R., Park, A. T., Takada, M. E., Robinson, S. T., Grotzinger, H., Last, B. S., Finn, A. S., Gabrieli, J. D. E., & Mackey, A. P. (2019). Associations between cortical thickness and reasoning differ by socioeconomic status in development. *Developmental Cognitive Neuroscience*, *36*. https://doi.org/10.1016/j.dcn.2019.100641

Levman, J., MacDonald, P., Lim, A. R., Forgeron, C., & Takahashi, E. (2017). A pediatric structural MRI analysis of healthy brain development from newborns to young adults. *Human Brain Mapping*, *38*(12), 5931–5942. https://doi.org/10.1002/hbm.23799

Lewis, J. D., Fonov, V. S., Collins, D. L., Evans, A. C., & Tohka, J. (2019). Cortical and subcortical T1 white/gray contrast, chronological age, and cognitive performance. *NeuroImage*, *196*, 276–288.

https://doi.org/10.1016/j.neuroimage.2019.04.022

Li, G., Nie, J., Wang, L., Shi, F., Lyall, A. E., Lin, W., Gilmore, J. H., & Shen, D. (2014). Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age. *Cerebral Cortex*, *24*(5), 1289–1300. https://doi.org/10.1093/cercor/bhs413

Li, G., Wang, L., Shi, F., Lyall, A. E., Ahn, M., Peng, Z., Zhu, H., Lin, W., Gilmore, J. H., & Shen, D. (2016). Cortical thickness and surface area in neonates at high risk for schizophrenia. *Brain Structure and Function*, *221*(1), 447–461. https://doi.org/10.1007/s00429-014-0917-3

Lidauer, K., Pulli, E. P., Copeland, A., Silver, E., Kumpulainen, V., Hashempour, N., Merisaari, H., Saunavaara, J., Parkkola, R., Lähdesmäki, T., Saukko, E., Nolvi, S., Kataja, E. L., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2022). Subcortical and hippocampal brain segmentation in 5-year-old children: Validation of FSL-FIRST and FreeSurfer against manual segmentation. *European Journal of Neuroscience*, *56*(5), 4619–4641. https://doi.org/10.1111/ejn.15761

Liu, Y., Zhi, M., & Li, X. (2011). Parental age and characteristics of the offspring. *Ageing Research Reviews*, *10*(1), 115–123. https://doi.org/10.1016/j.arr.2010.09.004

Luby, J., Belden, A., Botteron, K., Marrus, N., Harms, M. P., Babb, C., Nishino, T., & Barch, D. (2013). The effects of poverty on childhood brain development: The mediating effect of caregiving and stressful life events. *JAMA Pediatrics*, *167*(12), 1135–1142. https://doi.org/10.1001/jamapediatrics.2013.3139

Lyall, A. E., Shi, F., Geng, X., Woolson, S., Li, G., Wang, L., Hamer, R. M., Shen, D., & Gilmore, J. H. (2015). Dynamic Development of Regional Cortical Thickness and Surface Area in Early Childhood. *Cerebral Cortex*, *25*(8), 2204–2212. https://doi.org/10.1093/cercor/bhu027

Malm, H., Sourander, A., Gissler, M., Gyllenberg, D., Hinkka-Yli-Salomäki, S., McKeague, I. W., Artama, M., & Brown, A. S. (2015). Pregnancy complications following prenatal exposure to SSRIs or maternal psychiatric disorders: Results from population-based national register data. *American Journal of Psychiatry*, *172*(12), 1224–1232. https://doi.org/10.1176/appi.ajp.2015.14121575

Mariën, P., & Borgatti, R. (2018). Language and the cerebellum. In *Handbook of Clinical Neurology* (Vol. 154, pp. 181–202). Elsevier B.V. https://doi.org/10.1016/B978-0-444-63956-1.00011-4

Masouleh, S. K., Eickhoff, S. B., Zeighami, Y., Lewis, L. B., Dahnke, R., Gaser, C., Chouinard-Decorte, F., Lepage, C., Scholtens, L. H., Hoffstaedter, F., Glahn, D. C., Blangero, J., Evans, A. C., Genon, S., & Valk, S. L. (2020). Influence of processing pipeline on cortical thickness measurement. *Cerebral Cortex*, *30*(9), 5014–5027. https://doi.org/10.1093/cercor/bhaa097

McCarthy, C. S., Ramprashad, A., Thompson, C., Botti, J. A., Coman, I. L., & Kates, W. R. (2015). A comparison of FreeSurfer-generated data with and without manual intervention. *Frontiers in Neuroscience*, *9*(OCT). https://doi.org/10.3389/fnins.2015.00379

Merisaari, H., Tuulari, J. J., Karlsson, L., Scheinin, N. M., Parkkola, R., Saunavaara, J., Lähdesmäki, T., Lehtola, S. J., Keskinen, M., Lewis, J. D., Evans, A. C., & Karlsson, H. (2019). Test-retest reliability of Diffusion Tensor Imaging metrics in neonates. *NeuroImage*, *197*, 598–607. https://doi.org/10.1016/j.neuroimage.2019.04.067

Moog, N. K., Heim, C. M., Entringer, S., Simhan, H. N., Wadhwa, P. D., & Buss, C. (2022). Transmission of the adverse consequences of childhood maltreatment across generations: Focus on gestational biology. *Pharmacology Biochemistry and Behavior*, *215*. https://doi.org/10.1016/j.pbb.2022.173372

Muzik, O., Chugani, D. C., Juhász, C., Shen, C., & Chugani, H. T. (2000). Statistical parametric mapping: Assessment of application in children. *NeuroImage*, *12*(5), 538–549. https://doi.org/10.1006/nimg.2000.0651

Myrskylä, M., & Fenelon, A. (2012). Maternal Age and Offspring Adult Health: Evidence From the Health and Retirement Study. *Demography*, *49*(4), 1231–1257. https://doi.org/10.1007/s13524-012-0132-x

Nassar, A. H., & Usta, I. M. (2009). Advanced maternal age. part II: Long-term consequences. *American Journal of Perinatology*, *26*(2), 107–112. https://doi.org/10.1055/s-0028-1090593

Natu, V. S., Gomez, J., Barnett, M., Jeska, B., Kirilina, E., Jaeger, C., Zhen, Z., Cox, S., Weiner, K. S., Weiskopf, N., & Grill-Spector, K. (2019). Apparent thinning of human visual cortex during childhood is associated with myelination. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(41), 20750–20759. https://doi.org/10.1073/pnas.1904931116

Nwosu, E. C., Robertson, F. C., Holmes, M. J., Cotton, M. F., Dobbels, E., Little, F., Laughton, B., van der Kouwe, A., & Meintjes, E. M. (2018). Altered brain morphometry in 7-year old HIV-infected children on early ART. *Metabolic Brain Disease*, *33*(2), 523–535. https://doi.org/10.1007/s11011-017-0162-6

Oishi, K., Mori, S., Donohue, P. K., Ernst, T., Anderson, L., Buchthal, S., Faria, A., Jiang, H., Li, X., Miller, M. I., van Zijl, P. C. M., & Chang, L. (2011). Multi-contrast human neonatal brain atlas: Application to normal neonate development analysis. *NeuroImage*, *56*(1), 8–20. https://doi.org/10.1016/j.neuroimage.2011.01.051

Olson, I. R., Heide, R. J. V. der, Alm, K. H., & Vyas, G. (2015). Development of the uncinate fasciculus: Implications for theory and developmental disorders. In *Developmental Cognitive Neuroscience* (Vol. 14, pp. 50–61). Elsevier Ltd. https://doi.org/10.1016/j.dcn.2015.06.003

O'Muircheartaigh, J., Dean, D. C., Dirks, H., Waskiewicz, N., Lehman, K., Jerskey, B. A., & Deoni, S. C. L. (2013). Interactions between White Matter Asymmetry and Language during Neurodevelopment. *Journal of Neuroscience*, *33*(41), 16170–16177. https://doi.org/10.1523/JNEUROSCI.1463-13.2013

O'Muircheartaigh, J., Dean, D. C., Ginestet, C. E., Walker, L., Waskiewicz, N., Lehman, K., Dirks, H., Piryatinsky, I., & Deoni, S. C. L. (2014). White matter development and early cognition in babies and toddlers. *Human Brain Mapping*, *35*(9), 4475–4487. https://doi.org/10.1002/hbm.22488

Pechtel, P., & Pizzagalli, D. A. (2011). Effects of early life stress on cognitive and affective function: An integrated review of human literature. In *Psychopharmacology* (Vol. 214, Issue 1, pp. 55–70). https://doi.org/10.1007/s00213-010-2009-2

Perani, D., Saccuman, M. C., Scifo, P., Awander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(38), 16056–16061. https://doi.org/10.1073/pnas.1102991108

Peterson, R. L., & Pennington, B. F. (2012). Developmental Dyslexia. *Lancet*, *29*(9830), 997–1003. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Phan, T. V., Sima, D. M., Beelen, C., Vanderauwera, J., Smeets, D., & Vandermosten, M. (2018). Evaluation of methods for volumetric analysis of pediatric brain data: The childmetrix pipeline versus adult-based approaches. *NeuroImage: Clinical*, *19*, 734–744. https://doi.org/10.1016/j.nicl.2018.05.030

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. In *Developmental Cognitive Neuroscience* (Vol. 33, pp. 206–223). Elsevier Ltd. https://doi.org/10.1016/j.dcn.2017.08.009

Piccolo, L. R., Merz, E. C., He, X., Sowell, E. R., & Noble, K. G. (2016). Age-related differences in cortical thickness vary by socioeconomic status. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0162511

Poldrack, R., Juliana Pare´-Blagoev, E. J., & Grant, P. E. (2002). Pediatric Functional Magnetic Resonance Imaging:Progress and Challenges. *Topics in Magnetic Resonance Imaging*, *13*(1), 61–70.

Pulli, E. P., Kumpulainen, V., Kasurinen, J. H., Korja, R., Merisaari, H., Karlsson, L., Parkkola, R., Saunavaara, J., Lähdesmäki, T., Scheinin, N. M., Karlsson, H., & Tuulari, J. J. (2018). Prenatal exposures and infant brain: Review of magnetic resonance

imaging studies and a population description analysis. In *Human Brain Mapping* (Vol. 40, Issue 6, pp. 1987–2000). John Wiley and Sons Inc. https://doi.org/10.1002/hbm.24480

Pulli, E. P., Silver, E., Kumpulainen, V., Copeland, A., Merisaari, H., Saunavaara, J., Parkkola, R., Lähdesmäki, T., Saukko, E., Nolvi, S., Kataja, E. L., Korja, R., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2022). Feasibility of FreeSurfer Processing for T1-Weighted Brain Images of 5-Year-Olds: Semiautomated Protocol of FinnBrain Neuroimaging Lab. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.874062

Qiu, A., Rifkin-Graboi, A., Chen, H., Chong, Y. S., Kwek, K., Gluckman, P. D., Fortier, M. v., & Meaney, M. J. (2013). Maternal anxiety and infants' hippocampal development: timing matters. *Translational Psychiatry*, *3*. https://doi.org/10.1038/tp.2013.79

Remer, J., Croteau-Chonka, E., Dean, D. C., D'Arpino, S., Dirks, H., Whiley, D., & Deoni, S. C. L. (2017). Quantifying cortical development in typically developing toddlers and young children, 1–6 years of age. *NeuroImage*, *153*, 246–261. https://doi.org/10.1016/j.neuroimage.2017.04.010

Rosen, S., Wise, R. J. S., Chadha, S., Conway, E.-J., & Scott, S. K. (2011). *Hemispheric Asymmetries in Speech Perception: Sense, Nonsense and Modulations*. https://doi.org/10.1371/journal.pone.0024672

Ross, M. C., Dvorak, D., Sartin-Tarm, A., Botsford, C., Cogswell, I., Hoffstetter, A., Putnam, O., Schomaker, C., Smith, P., Stalsberg, A., Wang, Y., Xiong, M., & Cisler, J. M. (2021). Gray matter volume correlates of adolescent posttraumatic stress disorder: A comparison of manual intervention and automated segmentation in FreeSurfer. *Psychiatry Research - Neuroimaging*, *313*. https://doi.org/10.1016/j.pscychresns.2021.111297

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., Morris, J. C., Dale, A. M., & Fischl, B. (2004). Thinning of the cerebral cortex in aging. *Cerebral Cortex*, *14*(7), 721–730. https://doi.org/10.1093/cercor/bhh032

Salmi J, Pallesen KJ, Neuvonen T, Brattico E, Korvenoja A, Salonen O, Carlson S. Cognitive and motor loops of the human cerebro-cerebellar system. J Cogn Neurosci. 2010 Nov;22(11):2663-76. doi: 10.1162/jocn.2009.21382. PMID: 19925191.

Salminen, L. E., Wilcox, R. R., Zhu, A. H., Riedel, B. C., Ching, C. R. K., Rashid, F., Thomopoulos, S. I., Saremi, A., Harrison, M. B., Ragothaman, A., Knight, V., Boyle, C. P., Medland, S. E., Thompson, P. M., & Jahanshad, N. (2019). Altered Cortical Brain Structure and Increased Risk for Disease Seen Decades after Perinatal Exposure to Maternal Smoking: A Study of 9000 Adults in the UK Biobank. *Cerebral Cortex*, *29*(12), 5217–5233. https://doi.org/10.1093/cercor/bhz060

Scheinost, D., Sinha, R., Cross, S. N., Kwon, S. H., Sze, G., Constable, R. T., & Ment, L. R. (2017). Does prenatal stress alter the developing connectome? In *Pediatric Research* (Vol. 81, Issues 1–2, pp. 214–226). Nature Publishing Group. https://doi.org/10.1038/pr.2016.197

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

Schonwiesner, M., Rübsamen, R., & von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *European Journal of Neuroscience*, *22*(6), 1521–1528. https://doi.org/10.1111/j.1460-9568.2005.04315.x

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, *22*(3), 1060–1075.

https://doi.org/10.1016/j.neuroimage.2004.03.032

Seiger, R., Ganger, S., Kranz, G. S., Hahn, A., & Lanzenberger, R. (2018). Cortical Thickness Estimations of FreeSurfer and the CAT12 Toolbox in Patients with Alzheimer's Disease and Healthy Controls. *Journal of Neuroimaging*, *28*(5), 515–523. https://doi.org/10.1111/jon.12521

Serag, A., Kyriakopoulou, V., Rutherford, M. A., Edwards, A. D., Hajnal, J. v, Aljabar, P., Counsell, S. J., Boardman, J. P., & Rueckert, D. (2012). A. Serag et al.: a 4d probabilistic atlas of the developing brain a Multi-channel 4D Probabilistic Atlas of the Developing Brain: Application to Fetuses and Neonates. In *Annals of the BMVA* (Vol. 2012, Issue 3). www.brain-development.org.

Shapiro, G. D., Bushnik, T., Sheppard, A. J., Kramer, M. S., Kaufman, J. S., Yang, S., & Shapiro, G. (2017). Paternal education and adverse birth outcomes in Canada. *Community Health*. https://doi.org/10.1136/jech

Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., Clasen, L., Evans, A., Giedd, J., & Rapoport, J. L. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. www.pnas.org/cgi/content/full/

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature*, *440*(7084), 676–679. https://doi.org/10.1038/nature04513

Shaw, P., Malek, M., Watson, B., Sharp, W., Evans, A., & Greenstein, D. (2012). Development of cortical surface area and gyrification in attention-deficit/hyperactivity disorder. *Biological Psychiatry*, *72*(3), 191–197. https://doi.org/10.1016/j.biopsych.2012.01.031

Sheridan, M. A., Sarsour, K., Jutte, D., D'Esposito, M., & Boyce, W. T. (2012). The impact of social disparity on prefrontal function in childhood. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0035744

Shi, F., Yap, P. T., Wu, G., Jia, H., Gilmore, J. H., Lin, W., & Shen, D. (2011). Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS ONE*, *6*(4). https://doi.org/10.1371/journal.pone.0018746

Short, S. J., Elison, J. T., Goldman, B. D., Styner, M., Gu, H., Connelly, M., Maltbie, E., Woolson, S., Lin, W., Gerig, G., Reznick, J. S., & Gilmore, J. H. (2013). Associations between white matter microstructure and infants' working memory. *NeuroImage*, *64*(1), 156–166. https://doi.org/10.1016/j.neuroimage.2012.09.021

Shultz, S., Vouloumanos, A., Bennett, R. H., & Pelphrey, K. (2014). Neural specialization for speech in the first months of life. *Developmental Science*, *17*(5), 766–774. https://doi.org/10.1111/desc.12151

Silver, E., Pulli, E. P., Kataja, E. L., Kumpulainen, V., Copeland, A., Saukko, E., Saunavaara, J., Merisaari, H., Lähdesmäki, T., Parkkola, R., Karlsson, L., Karlsson, H., & Tuulari, J. J. (2022). Prenatal and early-life environmental factors, family demographics and cortical brain anatomy in 5-year-olds: an MRI study from FinnBrain Birth Cohort. *Brain Imaging and Behavior*, *16*(5), 2097–2109. https://doi.org/10.1007/s11682-022-00679-w

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, *17*(1), 87–97. https://doi.org/10.1109/42.668698

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

Spann, M. N., Bansal, R., Rosen, T. S., & Peterson, B. S. (2014). Morphological features of the neonatal brain support development of subsequent cognitive, language, and motor abilities. *Human Brain Mapping*, *35*(9), 4459–4474. https://doi.org/10.1002/hbm.22487

Stiles, J., & Jernigan, T. L. (2010). The basics of brain development. In *Neuropsychology Review* (Vol. 20, Issue 4, pp. 327–348). https://doi.org/10.1007/s11065-010-9148-4

Stipdonk, L. W., Boumeester, M., Pieterman, K. J., Franken, M. C. J. P., Rosmalen, J. van, Reiss, I. K., & Dudink, J. (2021). Cerebellar volumes and language functions in school-aged children born very preterm. *Pediatric Research*, *90*(4), 853–860. https://doi.org/10.1038/s41390-020-01327-z

Susser, L. C., Sansone, S. A., & Hermann, A. D. (2016). Selective serotonin reuptake inhibitors for depression in pregnancy. *American Journal of Obstetrics and Gynecology*, *215*(6), 722–730. https://doi.org/10.1016/j.ajog.2016.07.011

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

Swanson, M. R., Wolff, J. J., Elison, J. T., Gu, H., Hazlett, H. C., Botteron, K., Styner, M., Paterson, S., Gerig, G., Constantino, J., Dager, S., Estes, A., Vachet, C., & Piven, J. (2015). Splenium development and early spoken language in human infants. *Developmental Science*, 1–13. https://doi.org/10.1111/desc.12360

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

Theys, C., Wouters, J., & 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*(4). https://doi.org/10.1371/journal.pone.0094019

Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolín, S., Bralten, J., Bruin, W. B., Bülow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyler, L. T., Faraone, S. v., Favre, P., … Zelman, V. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. In *Translational Psychiatry* (Vol. 10, Issue 1). Springer Nature. https://doi.org/10.1038/s41398-020-0705-1

Toga, A. W., & Thompson, P. M. (2003). Mapping brain asymmetry. *Nature Reviews Neuroscience*, *4*(1), 37–48. https://doi.org/10.1038/nrn1009

Tooley, U. A., Bassett, D. S., & Mackey, A. P. (2021). Environmental influences on the pace of brain development. In *Nature Reviews Neuroscience* (Vol. 22, Issue 6, pp. 372–384). Nature Research. https://doi.org/10.1038/s41583-021-00457-5

Travis, K. E., Curran, M. M., Torres, C., Leonard, M. K., Brown, T. T., Dale, A. M., Elman, J. L., & Halgren, E. (2013). Age-related changes in tissue signal properties within cortical areas important for word understanding in 12- to 19-month-old infants. *Cerebral Cortex*, *24*(7), 1948–1955. https://doi.org/10.1093/cercor/bht052

Tusor, N., Arichi, T., Counsell, S. J., & Edwards, A. D. (2014). Brain Development in Preterm Infants Assessed Using Advanced MRI Techniques. In *Clinics in Perinatology* (Vol. 41, Issue 1, pp. 25–45). https://doi.org/10.1016/j.clp.2013.10.001

Vanderauwera, J., Altarelli, I., Vandermosten, M., de Vos, A., Wouters, J., & Ghesquière, P. (2018). Atypical Structural Asymmetry of the Planum Temporale is Related to Family History of Dyslexia. *Cerebral Cortex*, *28*(1), 63–72. https://doi.org/10.1093/cercor/bhw348

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental Science*, *10*(2), 159–164. https://doi.org/10.1111/j.1467-7687.2007.00549.x

Walhovd, K. B., Fjell, A. M., Giedd, J., Dale, A. M., & Brown, T. T. (2016). Through Thick and Thin: A Need to Reconcile Contradictory Results on Trajectories in Human Cortical Development. *Cerebral Cortex*, *27*(2), 1–10. https://doi.org/10.1093/cercor/bhv301

Wang, F., Lian, C., Wu, Z., Zhang, H., Li, T., Meng, Y., Wang, L., Lin, W., Shen, D., & Li, G. (2019). Developmental topography of cortical thickness during infancy. *Proceedings of the National Academy of Sciences of the United States of America*,

116(32), 15855–15860. https://doi.org/10.1073/pnas.1821523116

Waters, A. B., Mace, R. A., Sawyer, K. S., & Gansler, D. A. (2019). Identifying errors in Freesurfer automated skull stripping and the incremental utility of manual intervention. *Brain Imaging and Behavior*, *13*(5), 1281–1291. https://doi.org/10.1007/s11682-018-9951-8

Wedderburn, C. J., Subramoney, S., Yeung, S., Fouche, J. P., Joshi, S. H., Narr, K. L., Rehman, A. M., Roos, A., Ipser, J., Robertson, F. C., Groenewold, N. A., Gibb, D. M., Zar, H. J., Stein, D. J., & Donald, K. A. (2020). Neuroimaging young children and associations with neurocognitive development in a South African birth cohort study. *NeuroImage*, *219*. https://doi.org/10.1016/j.neuroimage.2020.116846

Weiss-Croft, L. J., & Baldeweg, T. (2015). Maturation of language networks in children: A systematic review of 22years of functional MRI. In *NeuroImage* (Vol. 123, pp. 269–281). Academic Press Inc. https://doi.org/10.1016/j.neuroimage.2015.07.046

Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, *87*, 120–126. https://doi.org/10.1016/j.neuroimage.2013.11.010

Wilke, M., Krägeloh-Mann, I., & Holland, S. K. (2007). Global and local development of gray and white matter volume in normal children and adolescents. *Experimental Brain Research*, *178*(3), 296–307. https://doi.org/10.1007/s00221-006-0732-z

Wilke, M., Schmithorst, V. J., & Holland, S. K. (2003). Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data. *Magnetic Resonance in Medicine*, *50*(4), 749–757. https://doi.org/10.1002/mrm.10606

Winkler, A. M., Sabuncu, M. R., Yeo, B. T. T., Fischl, B., Greve, D. N., Kochunov, P., Nichols, T. E., Blangero, J., & Glahn, D. C. (2012). Measuring and comparing brain cortical surface area and other areal quantities. *NeuroImage*, *61*(4), 1428–1443. https://doi.org/10.1016/j.neuroimage.2012.03.026

Wu, Y., Lu, Y. C., Jacobs, M., Pradhan, S., Kapse, K., Zhao, L., Niforatos-Andescavage, N., Vezina, G., du Plessis, A. J., & Limperopoulos, C. (2020). Association of Prenatal Maternal Psychological Distress with Fetal Brain Growth, Metabolism, and Cortical Maturation. *JAMA Network Open*, *3*(1). https://doi.org/10.1001/jamanetworkopen.2019.19940

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*, *44*(7), 1633–1640. https://doi.org/10.1007/s10803-014-2033-x

Yakovlev, P.A., & Lecours, I.R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minkowski (Ed.), Regional development of the brain in early life (pp. 3–70). Oxford: Blackwell.

Yang, D. Y. J., Beam, D., Pelphrey, K. A., Abdullahi, S., & Jou, R. J. (2016). Cortical morphological markers in children with autism: A structural magnetic resonance imaging study of thickness, area, volume, and gyrification. *Molecular Autism*, *7*(1). https://doi.org/10.1186/s13229-016-0076-x

Yang, M., Li, Y., Li, J., Yao, D., Liao, W., & Chen, H. (2017). Beyond the Arcuate Fasciculus: Damage to Ventral and Dorsal Language Pathways in Aphasia. *Brain Topography*, *30*(2), 249–256. https://doi.org/10.1007/s10548-016-0503-5

Yoon, U., Fonov, V. S., Perusse, D., & Evans, A. C. (2009). The effect of template choice on morphometric analysis of pediatric brain data. *NeuroImage*, *45*(3), 769–777. https://doi.org/10.1016/j.neuroimage.2008.12.046

Zhang, J., Zhong, S., Zhou, L., Yu, Y., Tan, X., Wu, M., Sun, P., Zhang, W., Li, J., Cheng, R., Wu, Y., Yu, Y., Ye, X., & Luo, B. (2021). Correlations between Dual-Pathway White Matter Alterations and Language Impairment in Patients with Aphasia: A Systematic Review and Meta-analysis. In *Neuropsychology Review* (Vol. 31, Issue 3, pp. 402–418). Springer.

https://doi.org/10.1007/s11065-021-09482-8

Zöllei, L., Iglesias, J. E., Ou, Y., Grant, P. E., & Fischl, B. (2020). Infant FreeSurfer: An automated segmentation and surface extraction pipeline for T1-weighted neuroimaging data of infants 0–2 years. *NeuroImage*, *218*. https://doi.org/10.1016/j.neuroimage.2020.1169

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2023

ISBN 978-951-29-9423-6 (PRINT)
ISBN 978-951-29-9424-3 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)