

Childhood images

Childhood experiences and the relation to brain structure and function in pre-adolescence



Elizabeth Buimer

UMC Utrecht Brain Center

Childhood images

Childhood experiences and the relation to
brain structure and function in pre-adolescence

Elizabeth Eilidh Laetitia Buimer

Childhood images

Childhood experiences and the relation to brain structure and function in pre-adolescence

Cover by: Elizabeth Buimer using modified illustrations from undraw.co to integrate outlines of FreeSurfer segmentations from the MRI data collected in YOUth: Child & Adolescent. The back cover is a mosaic of outlines of FreeSurfer segmentations from most of the MRI data collected in YOUth: Child & Adolescent. The outlines are in random order and not linked to any identifiable information.

Printed by: Ridderprint BV, www.ridderprint.nl on 100% recycled paper.

ISBN: 978-94-6483-418-5

DOI: <https://doi.org/10.33540/1936>

Childhood images

Childhood experiences and the relation to
brain structure and function in pre-adolescence

Beelden uit de kindertijd

Ervaringen tijdens de kindertijd en de relatie met
de structuur en functie van de hersenen
(*met een samenvatting in het Nederlands*)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

dinsdag 31 oktober 2023 des middags te 2.15 uur

door

Elizabeth Eilidh Laetitia Buimer

geboren op 29 januari 1993
te Aberdeen, Verenigd Koninkrijk

Promotor:

Prof. dr. H.E. Hulshoff Pol

Copromotor:

Dr. R.M. Brouwer

Beoordelingscommissie:

Prof. dr. S. Durston (voorzitter)

Prof. dr. C. Kemner

Prof. dr. J.L. Kenemans

Prof. dr. L. Nauta-Jansen

Prof. dr. N.F. Ramsey

Dit proefschrift werd (mede) mogelijk gemaakt met financiële steun van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO; beursnummer 024.001.003) voor het Consortium on Individual Development via het Zwaartekrachtprogramma.

Because I know there is strength in the differences between us,
and I know there is comfort where we overlap.

Lyrics by Angela Maria "Ani" DiFranco
from the song *Overlap* (1994).

Table of contents

Chapter 1 - Introduction.....	9
Chapter 2 - The YOUth cohort study: MRI protocol and test-retest reliability in adults.....	17
Chapter 3 - De-identification procedures for magnetic resonance images and the impact on structural brain measures at different ages.....	65
Chapter 4 - Adverse childhood experiences and fronto-subcortical brain structures in YOUth.....	95
Chapter 5 - Inter-individual differences in facial emotion processing in pre-adolescence.....	133
Chapter 6 - Summary and conclusion.....	169
Chapter 7 - References.....	183
A1 - Nederlandse wetenschappelijke samenvatting.....	215
A2 - Open science statement.....	222
A3 - Scientific publications.....	223
A4 - Societal publications.....	225
A5 - Samenvatting voor jongeren.....	226
A6 - Dankwoord.....	239
A7 - Curriculum Vitae.....	243

Chapter 1 - Introduction

How did you become the person you are today? The person you are today has been shaped, in part, by your environment and life experiences. Did you feel safe and supported? Did your parents have access to sufficient financial resources? What is your socio-cultural background? Traces of these childhood experiences can be found not only in childhood photo albums but also in magnetic resonance images of the developing brain.

Brain plasticity refers to the inherent ability of the brain to adapt its structure and function. It is highest in the developing brain and is driven by interactions between genes and the environment. The intricate interplay among genes, the environment, and the brain forms the foundation for the variations observed between individuals. Studying images of the developing brain aids researchers in gaining a deeper understanding of human development.



1 Studying brain development

Over the past decades developmental cognitive neuroscience emerged as an interdisciplinary scientific field to study environmental and biological influences on brain development (Blakemore et al., 2011). Technological advances enabled a rapid expansion of this field. Imaging the structure and functioning of the brain became possible in larger groups and younger children. This led to the initiation of multiple longitudinal studies investigating brain development using magnetic resonance imaging (MRI) (Bjork et al., 2017; Braams et al., 2015; Brown et al., 2015; Evans, 2006; Giedd et al., 1999; Herting et al., 2014; Schumann et al., 2010; Tamnes et al., 2013; van Soelen et al., 2012a; Wendelken et al., 2017; White et al., 2013; Yap et al., 2011). These cohorts provide rich datasets that can yield important insights on the concept of inter-individual differences in development. This PhD thesis will focus on one of these cohorts, the YOUTH study (Onland-Moret et al., 2020). But before we dive into the cohort specifics, I will introduce some of the main concepts that will help you to understand my research.

2 Prior studies

2.1 Typical brain development

At the macroscopic level, the brain grows rapidly during pregnancy and the first postnatal years with a steep increase in brain volume, cortical surface area and cortical folding (Dubois et al., 2021). During childhood and adolescence, the brain undergoes considerable developmental changes, including growth of subcortical volumes (Mills et al., 2021), a thinning of the cortex (Teeuw et al., 2019), an increase followed by a decrease in cortical surface area and continued growth of volume of the white matter connections (Frangou et al., 2022; Giedd et al., 1999; Koenis et al., 2015; Tamnes et al., 2017). Mapping neurodevelopmental trajectories is important as these trajectories help to explain inter-individual differences later in life, for example differences in cognitive functioning (Schnack et al., 2015) or psychiatric vulnerability (Paus et al., 2008).

2.2. Genetic influences on brain development

Brain structure is for a substantial part heritable (Brouwer et al., 2010, 2012; den Braber et al., 2013; Hulshoff Pol et al., 2012; Koenis et al., 2015; Lenroot et al., 2009; Lenroot & Giedd, 2008; Panizzon et al., 2009; Peper et al., 2009; Schmitt et al., 2014; van Soelen et al., 2011). Heritability is an estimate of the proportion of variation in a population that can be explained by genetic variation. Heritability is depending on the brain region and on the type of brain measure. For example, cortical surface area and cortical thickness are thought to have different sources of genetic influences (Panizzon et al., 2009). Importantly, the heritability of structural brain measures also depends on age and thus changes throughout development (Lenroot et al., 2009; Lenroot & Giedd, 2008; Schmitt et al., 2014). The speed and rate of brain development is also heritable (Brouwer et al., 2017, 2020; van Soelen et al., 2012b), with limited genetic overlap between variants involved in cross-sectional structural brain measures compared to the change in structural brain measures (Brouwer et al., 2022). Less studies focused on heritability of brain function, but available studies suggest lower heritability for brain function compared to brain structure (Jansen et al., 2015). Still, there are plenty examples of studies showing considerable heritability for resting-state and task-based functional MRI (Achterberg et al., 2018; Adhikari et al., 2018; Blokland et al., 2008; Teeuw et al., 2019).

2.3 The importance of the environment

Throughout development genetic and environmental factors interact. Sensitive periods of development involve complex experience-expectant learning mechanisms and encompass periods of heightened neuroplasticity (Garbard-Durnam & McLaughlin, 2020). Neuroplasticity is the ability of the brain to respond to stimuli by reorganizing the structure and connectivity of the brain. Neuroplasticity can be described as a double-edged sword (Brunson et al., 2003), positive experiences during sensitive periods of development can have long-lasting positive effects while negative experiences can have adverse effects on developmental outcomes. Adverse childhood experiences (ACEs), such as maltreatment, parental divorce, exposure to violence or substance abuse in the family, are a risk factor for developing mental health problems

later in life (Green et al., 2010; Kessler et al., 2010; McLaughlin, 2016). ACEs are associated with decreased life expectancy, for example via the effects of ACEs on toxic stress, increased adult health risk behavior, increased suicidality or socioeconomic inequality (Felitti et al., 1998; Hughes et al., 2017; Kalmakis & Chandler, 2015; Merrick et al., 2019). ACEs are also associated with aberrant brain structure and function (Calem et al., 2017; Cassiers et al., 2018; Daniels et al., 2013; Hart & Rubia, 2012; Kraaijevanger et al., 2020; Lim et al., 2020; McCrory et al., 2010; McLaughlin et al., 2019; Paquola et al., 2016; Teicher & Samson, 2016). The effect of ACEs on the brain and on outcomes later in life is thought to depend on the timing of the ACE in relation to sensitive periods of brain development (Andersen et al., 2008; Gee & Casey, 2015; Heim & Binder, 2012; Kuhn et al., 2016; Tottenham & Sheridan, 2009). Thus, there are clear interrelations between ACEs, (mental) health outcomes and brain development.

2.4 Mechanisms of risk and resilience

Because of the high prevalence of ACEs in society, it is important to better understand mechanisms that confer risk or resilience in individuals previously exposed to childhood adversity. Different neurobiological theories explain the relation between ACEs and adverse outcomes later in life. The theory of latent vulnerability (Figure 1) argues that vulnerability for psychiatric outcomes in adults exposed to childhood adversity arises from neurocognitive calibrations to neglectful or maltreating environments (McCrory & Viding, 2015).

These neurocognitive calibrations can increase the occurrence (stress generation) or the impact (stress susceptibility) of new stressful experiences (Goemans et al., 2023). Furthermore, altered social functioning in individuals previously exposed to childhood adversity may result in less social support (social thinning) (McCrory et al., 2022). For example, aberrant facial emotion processing is one of the most consistent neuroimaging findings in the childhood maltreatment literature (Hein & Monk, 2017). Also on a behavioral level childhood adversity is related to aberrant emotion recognition (Assed et al., 2020; Bérubé et al., 2023). Unfortunately, atypical

emotion recognition could hinder meaningful social interactions with others later in life (Trentacosta & Fine, 2010).

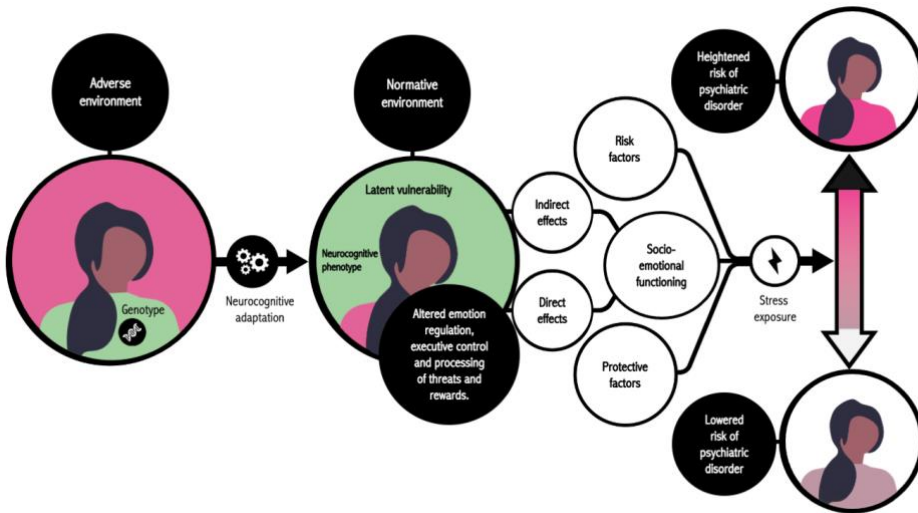


Figure 1. Schematic illustration of the theory of latent vulnerability. According to the theory of latent vulnerability, adverse early environments result in neurocognitive alterations through gene-environment interactions. The changes in response to the environment may help the child early in life but may prove maladaptive in normative environments. The maladaptive nature of the neurocognitive adaptations results in increased risk for mental health problems later in life. Inter-individual differences in mental health outcomes in individuals exposed to childhood adversity are explained by genetic factors as well as exposure to adverse or protective environmental factors in emerging adulthood. This illustration is adapted and slightly simplified from McCrory et al., (2017).

Importantly, we also know that social support can enhance resilience after childhood adversity (van Harmelen et al., 2017, 2021). The resilience framework approaches resilience as a dynamic process of adaptation resulting in a quick recovery of mental health after a stressful life event (Kalisch et al., 2017). Resilience is best explained as a complex, dynamic network of biological, psychological and social factors that help explain why some individuals do better than expected after childhood adversity (Ioannidis et al., 2020; Kalisch et al., 2019). Understanding neurobiological

mechanisms of risk and resilience can ultimately help to prevent negative outcomes after childhood adversity.

3 This thesis

3.1 The YOUth cohort study

The YOUth (Youth Of Utrecht) study is an ongoing longitudinal cohort study that comprises two independent cohorts: YOUth Baby & Child and YOUth Child & Adolescent. Together these cohorts provide a complete overview of development from 20 weeks of gestational age to adolescence. The aim of the YOUth study is to map variation in typical neurocognitive development and investigate why some children develop problematic behavior while others show resilient functioning (Onland-Moret et al., 2020). In this thesis, I will use the data collected in over 1000 children participating in the YOUth: Child & Adolescent cohort. More specific, I will use MRI data, neurocognitive measures, social competence measures and assessments of adverse childhood experiences.

3.2 Techniques

The YOUth MRI protocol comprises different types of MRI scans (Figure 2). Each type of scan provides other information about the developing brain. Structural T1-weighted images have high spatial resolution and can be used to estimate anatomical brain measures, such as (sub)cortical volume, cortical thickness, and cortical surface area. Diffusion-weighted images (DWI) can be used to map the integrity of white matter connections between brain regions. Resting-state functional MRI (rs-fMRI) scans are used to study brain function without an explicit task. Task-based functional MRI (fMRI) scans are acquired during a task and are used to elicit specific task-related functional activity. The YOUth study specifically focuses on self-regulation and social competence. Therefore, two fMRI tasks were chosen to match these themes: the inhibition task as a proxy of self-regulation and the emotion task as a proxy of social competence.

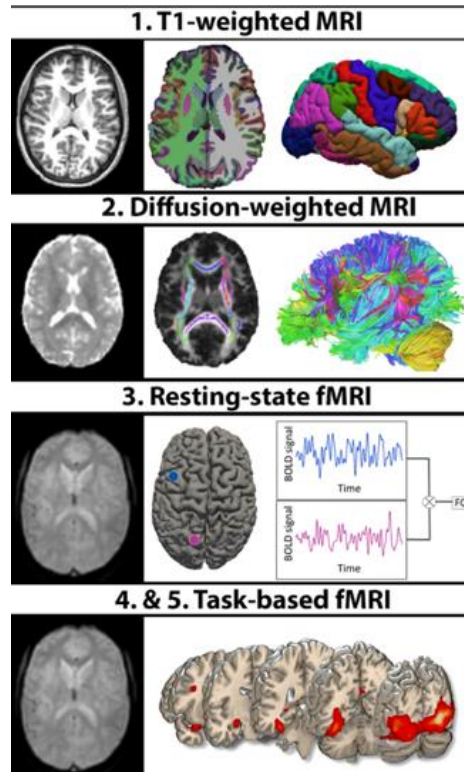


Figure 2. Scan types collected in YOUNG in order of acquisition. 1) Original T1-weighted scan (left), with subcortical and cortical brain tissue segmentation (middle) and the cortical regions of interest (right). 2) Diffusion unweighted volume after preprocessing (left); the intersection of the white matter regions (colored) and the skeleton plotted on the FA map (middle); the reconstructed fiber tracts used to create the connectivity maps (right). 3) One dynamic volume of the fMRI scan (left) and a schematic representation of how functional connectivity is computed (right). 4) One dynamic volume of the fMRI scan (left) and task-related activity during the face-processing in the emotion task (right). Reproduced from (Buimer et al., 2020).

3.3 Reliability of brain measures

Studying subtle inter-individual differences in the development of brain structure and function requires reliable brain measures. One way to assess reliability is by using a test-retest design. Therefore, as part of the quality control procedure in the

YOUth study, 17 young adults were scanned twice. To assess reliability of the YOUth MRI protocol, I compared structural and functional brain measures derived from the same subjects between the two scan sessions.

3.4 Sharing MRI data

YOUth is designed to facilitate data sharing with internal and external researchers guided by the FAIR (Findable, Accessible, Interoperable and Reusable) data principles (Wilkinson et al., 2016; Zondergeld et al., 2020). However, brain scans also contain privacy-sensitive facial characteristics. For these reasons, more and more open-access datasets contain MRI scans that were subjected to some type of de-identification method. I studied the impact of these methods on subsequent processing of the MRI scans in different age groups.

4. Aim and outline

The aim of this thesis is to study sources of inter-individual variation in brain structure and function in pre-adolescence. Each chapter focuses on a different source of variation. One source of variation in any type of measure is noise. **Chapter 2** focuses on the test-retest reliability of structural and functional brain measures derived from the YOUth MRI protocol. **Chapter 3** answers the question to what extent de-identification methods can introduce noise. After I mapped noise as potential source of variation in neuroimaging data, I move on to factors that are important for brain development. I specifically focus on the social environment and pre-adolescent brain structure and function. In **Chapter 4** I show the association between anatomical brain measures and adverse childhood experiences. **Chapter 5** closes with the interrelations between the ability to label emotions on facial expressions, neural activity during the processing of emotional faces and social competence. **Chapter 6** summarizes the significance of the main findings in this thesis and discusses future directions.

Chapter 2 - The YOUth cohort study: MRI protocol and test-retest reliability in adults.

Elizabeth E.L. Buimer*, Pascal Pas*, Rachel M. Brouwer, Martijn Froeling, Hans Hoogduin, Alexander Leemans, Peter Luijten, Bastiaan J. van Nierop, Mathijs Raemaekers, Hugo G. Schnack, Jalmar Teeuw, Matthijs Vink, Fredy Visser, Hilleke E. Hulshoff Pol, René C.W. Mandl

*both authors contributed equally

Published in 2020

Developmental cognitive neuroscience, 45, 100816.

<https://doi.org/10.1016/j.dcn.2020.100816>

Abstract

The YOUth cohort study is a unique longitudinal study on brain development in the general population. As part of the YOUth study, 2000 children will be included between 8, 9 or 10 years of age and planned to return every three years during adolescence. Every three years magnetic resonance imaging (MRI) brain scans are collected, including structural T1-weighted imaging, diffusion-weighted imaging (DWI), resting-state functional MRI and task-based functional MRI. Here, we provide a comprehensive report of the MR acquisition in YOUth Child & Adolescent including the test-retest reliability of brain measures derived from each type of scan. To measure test-retest reliability, 17 adults were scanned twice with a week between sessions using the full YOUth MRI protocol. Intraclass correlation coefficients were calculated to quantify reliability. Global brain measures derived from structural T1-weighted and DWI scans were reliable. Resting-state functional connectivity was moderately reliable, as well as functional brain measures for both the inhibition task (stop versus go) and the emotion task (face versus house). Our results complement previous studies by presenting reliability results of regional brain measures collected with different MRI modalities. YOUth facilitates data sharing and aims for reliable and high-quality data. Here we show that using the state-of-the art YOUth MRI protocol brain measures can be estimated reliably.

1. Introduction

To quantify and understand atypical brain development, we need to first understand typical brain development. In the past two decades multiple longitudinal magnetic resonance imaging (MRI) studies investigating brain development have been initiated around the world (Bjork et al., 2017; Braams et al., 2015; Brown et al., 2015; Evans, 2006; Giedd et al., 1999; Herting et al., 2014; Schumann et al., 2010; Tamnes et al., 2013; van Soelen et al., 2012a; Wendelken et al., 2017; White et al., 2013; Yap et al., 2011). These cohorts provide rich datasets that can yield important insights on the concept of optimal brain development and individual developmental trajectories.

Studying subtle inter-individual differences in the development of brain structure and function requires reliable brain measures. One way to assess reliability is by using a test-retest design, in which subjects are scanned repeatedly in a short time period. Although, a between-scan interval of a month or less seems appropriate, this data is rarely collected in children and the shortest time intervals found in fMRI test-retest literature are between 3 to 6 months (Herting et al., 2018). Short time intervals ensure that changes related to plasticity or development are negligible and therefore intra-individual variation between these scan sessions can be regarded as noise. Test-retest reliability can be quantified with the intraclass correlation coefficient (ICC) (Bartko & Carpenter, 1976; McGraw & Wong, 1996; Shrout & Fleiss, 1979), a widely-used statistic in both structural and functional MRI studies.

YOUth (Youth of Utrecht) is an ongoing longitudinal cohort study that comprises two independent cohorts: YOUth Baby & Child and YOUth Child & Adolescent. Together these cohorts should provide a complete overview of development from 20 weeks of gestational age to adolescence. The aim of the YOUth study is to map variation in typical neurocognitive development and investigate why some children develop problematic behavior and others show resilience. To this end, an extensive dataset is collected, including MRI, eye tracking, parent-child observations, computer tasks, cognitive measurements and questionnaires on behavior, personality, health, lifestyle, parenting, child development, use of (social) media and more. Furthermore, blood samples, buccal swabs, saliva and hair samples are collected. More information

about the study design and a full overview of the collected data can be found at the website: www.uu.nl/en/research/youth-cohort-study (Onland-Moret et al., 2020). The current paper focuses on the MRI data collected in the YOUth Child & Adolescent cohort.

The YOUth MRI protocol comprises different types of MRI scans, i.e. structural T1-weighted images, diffusion-weighted images (DWI), resting-state functional MRI (rs-fMRI) scans and task-based functional MRI (fMRI) scans. YOUth specifically focuses on self-regulation and social competence. Therefore, two fMRI tasks were chosen to match these themes: the inhibition task as a proxy of self-regulation and the emotion task as a proxy of social competence.

YOUth is designed to facilitate data sharing with internal and external researchers guided by the FAIR (Findable, Accessible, Interoperable and Reusable) data principles (Wilkinson et al., 2016). In this paper we provide a transparent report of the collected MRI data. The aim of this paper is two-fold: First, to describe the full YOUth MRI protocol including its state-of-the-art MRI acquisition protocol. Second, to quantify the test-retest reliability of the included MRI acquisitions. To assess test-retest reliability of the YOUth MRI protocol, we included a sample of 17 healthy adult volunteers.

2. Materials and methods

2.1 YOUth child & adolescent

2.1.1 Sample and recruitment

YOUth Child & Adolescent aims to include a total of 2000 children from the general population and their parents or caregivers. Children are recruited mostly at primary schools in the province of Utrecht, the Netherlands. At the first measurement children are 8, 9 or 10 years old. Follow-up measurements are planned every three years during adolescence.

2.1.2 In- and exclusion criteria

All children in the specified age categories can be included as long as they are physically and mentally capable to participate. Furthermore, we exclude children if they or their parents do not master the Dutch language enough to give informed consent or participate in the different subparts of the study. Atypically developing children are not excluded but also not specifically selected. Children that do not meet MR safety criteria (absence of specific metal implants including most braces) were still welcome to participate in the other parts of the study.

2.2 The YOUth MRI protocol

2.2.1 Mock procedure

Prior to scanning, children undergo a practice session in a mock scanner. Implementing a mock procedure mimicking the actual experience in the scanner has been shown to decrease scanner-related distress in children (Durstun et al., 2009). For YOUth, an older MR scanner model, no longer operational, is reconstructed to be used as a mock scanner to make the experience as authentic as possible. Print-outs of T1-weighted scans with severe motion artefacts and negligible motion artefacts are shown to explain the importance of not moving in the scanner at the level of the child. During the simulation, children are positioned in a mock scanner with headphones on. To familiarize them to the noise of the different MRI sequences sound recordings of these sequences are played, while they practice the inhibition task that they will perform in the real scanner. Following the scanner simulation, the child, the parent or guardian and the research assistant rate the level of excitement and anxiety of the child in anticipation of the MRI scans. This is done using a Visual Analogue Scale where the rater indicates on two questions how excited the child feels and how tensed the child feels. These measurements are used as a proxy of scanner-related distress. If any of the three raters estimate high scanner-related distress, the MRI visit may be canceled. This procedure is repeated just before commencing the MRI session. Furthermore, the MRI session can be canceled at any time if the child or the parent/guardian indicates that the child does not feel comfortable continuing.

2.2.2 Acquisition

Scans are acquired on a Philips Ingenia 3.0T CX scanner with a 60 centimeter bore (Philips Medical Systems, Best, The Netherlands), using a 32-channel SENSE head-coil and operated using software version R530. First, a structural T1-weighted 3D gradient echo scan is acquired, followed by a diffusion-weighted multi-shell multi-band echo planar (EPI) acquisition including two short DWI scans with a reversed phase encoding readout to correct for susceptibility artefacts. Next, multi-band EPI acquisitions are acquired during resting-state, the inhibition task and the emotion task. During the acquisition, the structural T1-weighted scan is visually checked for motion artefacts. If the MR operator regards the scan as unusable due to severe motion artefact, the scan is repeated after emphasizing the instructions to lie still. Prior to the fMRI scans, a short EPI acquisition scan of one dynamic is acquired. MR operators use this scan to visually check the reconstruction for (shimming) artefacts or for replacement of the head outside of the field of view. If rescanning is needed, this can come at the expense of the last acquisition as we always ensure that the ethically approved maximal time in the MR scanner is not exceeded.

The main acquisition parameters are listed in Table 1. See Online Supplementary materials, for the complete set of acquisition parameters. An illustration of the scan types collected in YOUth can be found in Figure 1.

Table 1. Acquisition parameters YOUth MRI protocol.

Parameters	Structural T1-weighted	DWI	EPI		
			resting state	inhibition task	emotion task
Acquisition time (m:s)	10:02	8:05	8:07	9:22	6:40
Scan orientation	sagittal	transversal	transversal	transversal	transversal
TR (ms)	10	3500	1000	1000	1000
TE (ms)	4.6	99	25	25	25
Flip angle (degrees)	8	90	65	65	65
Number of slices	*	66	51	51	51
Slice thickness (mm)	*	2.0	2.5	2.5	2.5
Field of view (mm)	240x240x200	224x224	220x220	220x220	220x220
Acquisition matrix	304x304	112x112	88x88	88x88	88x88
Reconstructed voxel size (mm ³)	0.75x0.75x0.80	2.0x2.0x2.0	2.5x2.5x2.5	2.5x2.5x2.5	2.5x2.5x2.5
Multiband acceleration factor	Off	3	3	3	3
Parallel imaging factor	1.70 (AP) 1.40 (RL)	1.30 (AP)	1.80 (AP)	1.80 (AP)	1.80 (AP)
Diffusion directions		105			
		500 [15]			
		1000 [30]			
b-values (s/mm ²)		2000 [60]			
[directions]		every 10th scan is a B0-scan			

Abbreviations: m = minutes; s = seconds; TR = repetition time; TE = echo time; ms = milliseconds; mm = millimeter; AP = anterior-posterior axis; RL = right-left axis; *3D acquisition.

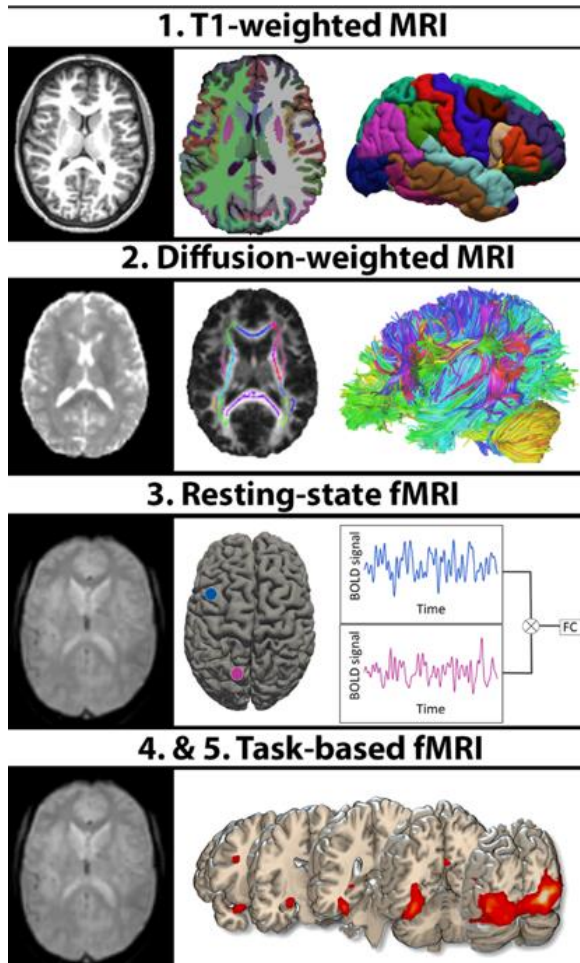


Figure 1. Scan types collected in YOUth in order of acquisition. 1) Original T1-weighted scan (left), with subcortical and cortical brain tissue segmentation (middle) and the cortical regions of interest (right). 2) Diffusion unweighted volume after preprocessing (left); the intersection of the white matter regions (colored) and the skeleton plotted on the FA map (middle); the reconstructed fiber tracts used to create the connectivity maps (right). 3) One dynamic volume of the fMRI scan (left) and a schematic representation of how functional connectivity is computed (right). 4) One dynamic volume of the fMRI scan (left) and task-related activity during the face-processing in the emotion task (right).

2.2.3 Stimulus presentation

During the scan session, stimuli for fMRI acquisitions are presented using an MRI-compatible 23-inch LCD screen with a resolution of 1080 by 1920 pixels (BOLDscreen, Cambridge Research Systems). During the rs-fMRI acquisition, lights inside the scanner room are turned off and participants are instructed to look at a white fixation cross on a grey screen.

2.2.4 Inhibition task

The stop-signal anticipation task for functional MRI (Zandbelt & Vink, 2010) aims to measure performance and brain activation during actual stopping as well as during the anticipation of stopping. Subjects are presented with three parallel horizontal lines. On each trial, a bar moves at a constant speed from the lower line towards the upper line, reaching the middle line in 800 milliseconds. The main task is to stop the bar as close to the middle line as possible, by pressing a button with the right thumb (i.e. Go trial). Stop trials are identical to Go trials, except that the bar stops moving automatically before reaching the middle line, indicating that a response has to be suppressed (i.e. stop-signal). The probability that such a stop-signal will appear is manipulated across trials and can be anticipated based on three different cues; '0' indicating 0% chance, '*' 22 percent and '**' 33 percent chance the bar will stop on its own. Task difficulty is adjusted to performance in a stepwise fashion, with a varying delay between the stop-signal and the target (i.e. the stop-line) depending on the success of the previous trial, thereby keeping the number of failed and successful trials comparable between subjects and sessions.

2.2.5 Emotion task

The emotion task is aimed at activating face processing areas in the brain. Participants are required to passively view pictures of faces (happy, fearful, or neutral expression) and pictures of houses. The pictures are presented in a pseudorandom order with blocks of face images interspersed with blocks of house images. The stimuli are taken from the Radboud Faces Database (Langner et al., 2010). Stimuli are presented in blocks of 18 seconds, with four blocks for each of the

four stimulus types. Rest periods are modeled as implicit baseline. Because of the short duration of the task, this block-design combined with passive viewing was chosen to ensure a strong contrast between conditions, without noise from behavioral responses. Behavioral data on emotion labeling in the children is measured in another part of YOUth (outside the scanner) during a computer task. To ensure that participants stay awake, they are instructed to press a button in between the block in response to a cue (red circle).

For more information on both fMRI tasks and their usage in the YOUth cohort study, see: www.uu.nl/en/research/youth-cohort-study.

2.3 The YOUth MRI protocol - Quality control

In the YOUth study, all children are scanned on the same scanner, with the acquisition parameters kept as stable as possible over the course of the study. Scanner soft- and firmware are only updated when it concerns essential updates with minimal impact on the acquisition. Scanner performance is monitored systematically throughout the YOUth study.

2.3.1 Monitoring scanner performance using human data

Collected MRI-scans of the children are processed immediately after data collection for quality control purposes, on a local server with scripted pipelines. Functional MRI scans are processed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>). The structural T1-weighted scans are processed using the CAT toolbox (<http://www.neuro.uni-jena.de/cat/>). DWI scans are processed using the SQUAD- tool running on FSL (Andersson & Sotiropoulos, 2016; Bastiani et al., 2019). Quality control (QC) measures are generated automatically after each scanning session and results are accessible through a web-portal on the local intranet for in-house viewing purposes. These reports consist of different slices generated from the T1-weighted scans to allow for a visual check, with additional statistics like noise- and inhomogeneity-contrast ratios from the CAT toolbox. A single researcher, experienced in quality control, visually checks these reports and this results in a list of scans that are

deemed unusable due to inhomogeneity and movement artefacts. In the future, we will also perform a QC on the outer surface reconstruction of the FreeSurfer output to have more information about which scans are unusable. We do not plan to provide quality information at the ROI-level as there is no golden standard for this type of QC yet and depending on the research question different processing software or parcellation atlases can be used. For DWI scans the reports are generated using QUAD (part of FSL's EDDY QC) and include information on the amount of spatial distortion and artefacts in the scans (Bastiani et al., 2019). For fMRI-scans statistics on movement and signal-to-noise ratio (SNR) are generated, including signal maps for visual inspection. Reports are checked manually after each scanning session and a qualitative assessment is saved as meta-data to the local XNAT storage server (Marcus et al., 2007a) together with the raw data. An example of a QC report, generated for each participant, is added in the Supplement.

2.3.2 Monitoring scanner performance using phantom data

Every other week a proton (demi water) spherical phantom (Philips sphere A fluid, doped with CuSO_4 1 ml + SH_2O 60mg; acetate 2.5 ml; ethanol 5.0 ml; H_3PO_4 4.4ml; total contents 524ml) fixed in a standard placeholder is used to acquire a series of scans. These scans include a B_0 map to determine the uniformity of the main magnetic field based on two gradient echo images with varying echo time; a B_1 map to determine the uniformity of the excitation field based on two gradient echo images with varying repetition time; a 3D gradient echo scan with, and without, the use of gradients and RF excitation; and a dynamic fast field EPI scan (2000 dynamics and 30 dummy scans). After each measurement, data is processed automatically. The output is accessible through a local server and results are inspected to monitor changes over time as well as temporary changes.

2.3.3 Example of data on scanner stability in YOUth

Signal-to-fluctuation-noise ratio (SFNR) is an important measure for estimating the presence of unwanted scanner-related variance in fMRI data (Bennett & Miller, 2010; Murphy et al., 2007) that can e.g. be used as covariate to calibrate multicenter studies

(Friedman et al., 2006). A stable scanner would have a high and stable SNR and SFNR. Figure 2 shows the SFNR calculated from resting-state human data (top row). The human data is derived from the rs-fMRI data collected in the YOUth cohort. The average human data is smoothed by filtering it with a 100-point gaussian window. Figure 2 also shows the SFNR (middle row) and the SNR (bottom row) derived from the dynamic fast field EPI scan in the phantom data (Friedman & Glover, 2006; Weisskoff, 1996).

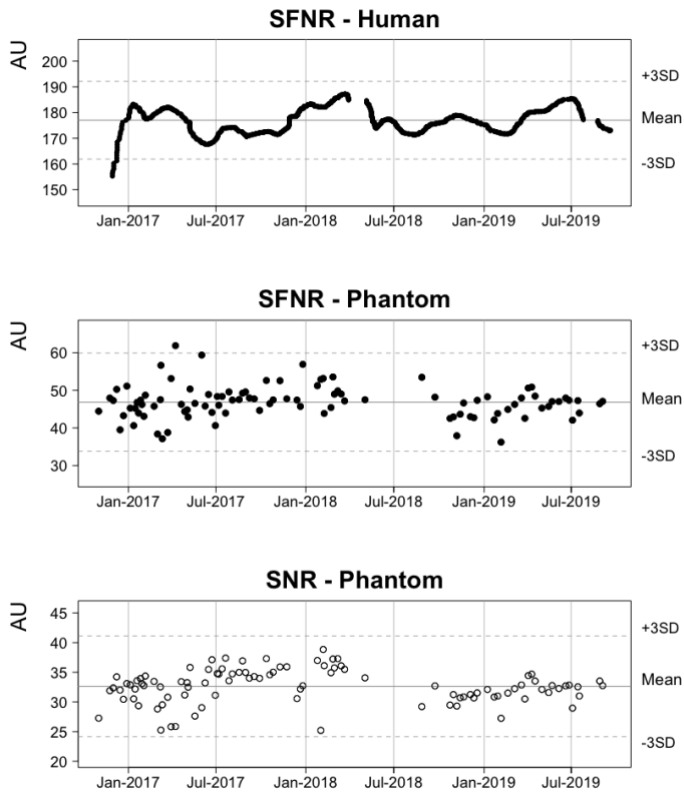


Figure 2. Monitoring scanner performance with human and phantom data using dynamic EPI scans. Data on scanner stability over the course of the study. The solid horizontal line indicates the mean of the signal and the dotted line indicates a threshold of ± 3 standard deviations from the mean.

2.4 The reliability study – Sample and recruitment of adults

To assess the test-retest reliability of the YOUth MRI protocol, we recruited healthy adult volunteers under the premise of MRI protocol development approved by the Medical Ethical Committee. All participants gave written informed consent prior to participation. Test-retest data was collected in adults, in the absence of ethical approval to include YOUth participants for this purpose. Participants were scanned twice with the MRI protocol used in the YOUth children’s cohort study described above. The scan-rescan interval was between 6 and 8 days. The test-retest sample consisted of 17 volunteers (7 male and 10 female) with a mean age of 23 years old (range: 19 to 31 years old). The participants, most of which were university students, were not given any restrictions regarding food or drink intake.

2.5 The reliability study - MRI processing

All scans were visually checked before starting the analyses. If a scan was excluded from analysis, both test and retest scans of the subject were excluded. Only those scans were excluded that had such obvious artefacts or anatomical anomalies that they would have been removed in regular practice. This resulted in sample sizes of 15 or 16 subjects depending on the type of scan. For the reliability-analyses of the T1-weighted scans, one male was excluded due to a structural anomaly. For the analyses of the DWI scans, one female was excluded due to motion artefacts and one female due to extensive spatial distortions. For the analysis of the resting-state MRI data, one female was excluded due to motion artefacts and one male due to an anatomical anomaly. For the task-based fMRI analyses, one male was excluded due to a local artefact and one female was excluded due to missing data.

2.5.1 Processing of structural T1-weighted scans

The T1-weighted test-retest scans were processed using FreeSurfer version 6.0 (freesurfer.net) for automatic brain segmentation and parcellation (Fischl et al., 2002). Global and regional brain measures of subcortical volume, cortical volume, cortical thickness and cortical surface area were extracted. The ROIs established

according to the Desikan-Killiany atlas were used for further analysis (Desikan et al., 2006). Besides atlas-based measures of cortical thickness, vertex-wise cortical thickness measures were extracted to include a measure that is independent of a parcellation atlas. For the vertex-wise analysis, cortical thickness of each scan was resampled to an average brain created with FreeSurfer by averaging the first scan of each participant in the test-retest dataset. After resampling, the cortical surface was smoothed with a 3D Gaussian kernel (FWHM = 10 mm).

2.5.2 Processing of DWI scans

FSL (version 6.01) in combination with MRtrix (version 3.0) was used to preprocess the DWI scans as described in detail here: B.A.T.M.A.N.: <https://osf.io/fkyht/>. Preprocessing included gradient direction correction (Leemans & Jones, 2009), eddy current (Andersson & Sotiropoulos, 2016) and susceptibility corrections (Andersson et al., 2003) as well as a correction for Gibbs ringing (Perrone et al., 2015). No correction for signal drift was needed because dynamic stabilization was applied in the acquisition. The results were visually checked and a QC check was performed (using squad, part of FSL). Tract-Based Spatial Statistics (TBSS) were used with the default settings to create a skeletonized version of the fractional anisotropy (FA) and mean diffusivity (MD) values computed from the single tensors (computed using FSL's DTIFit) that were fitted to the preprocessed multi-shell diffusion data. Global FA and MD values were computed for all skeleton voxels. In addition, average FA and MD values were computed over skeleton voxels from 48 regions of interest (ROIs) selected from the ICBM-DTI-81 white matter (WM) labels atlas (Mori & van Zijl, 2007) similar to (Svatkova et al., 2015).

Connectivity maps were constructed using MRtrix to perform test-retest analysis of the structural network analysis. Here the gray matter (GM) ROIs of the Desikan-Killiany atlas from the FreeSurfer output (generated while processing the T1-weighted scans) were used to define the nodes of the network. Fiber orientation distributions were estimated by deconvolution of the diffusion signal using 8th order spherical harmonics. The response function was obtained using the multi-shell-multi-tissue constrained spherical deconvolution algorithm. For each dataset 5,000,000

streamlines were generated within a seeding area covering the whole brain using deterministic tracking and a FOD-amplitude threshold of 0.05. The number of streamlines was then filtered down to 1,000,000 so that streamline densities better matched the fiber orientation distributions. Connectivity maps were generated by assigning streamlines to the closest node (ROI) found within a 2 mm radius of the streamlines' endpoints. Streamlines were stored only if they connected two different nodes. Connectivity maps were created based on the number of streamlines and their mean FA for each edge (connection between nodes). Only edges with at least four streamlines in 60% of the subjects were included in the analysis (de Reus & van den Heuvel, 2013). For these connectivity maps characteristic path length, global efficiency, mean local efficiency and mean strength were calculated (Dimitriadis et al., 2017).

2.5.3 Processing of rs-fMRI scans

Processing of rs-fMRI scans was performed using the CONN toolbox version 18a (Whitfield-Gabrieli & Nieto-Castanon, 2012) and SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) in MATLAB 2015b (The MathWorks Inc., Massachusetts, United States). The structural T1-weighted MRI scans were segmented into cerebrospinal fluid (CSF), GM and WM tissue maps, and registered to MNI-152 space using unified segmentation. The WM and CSF tissue maps were threshold at >50% and binarized to create tissue masks. The WM masks were eroded by two voxels to reduce the number of voxels at the white-gray matter tissue interface. The CSF tissue masks were constrained to contain only voxels inside the lateral ventricles. Motion correction was performed by realigning the volumes of the rs-fMRI scans to the mean functional volume using a rigid-body transformation in a two-stage approach. The transformation parameters were used to compute frame-wise displacement as an approximation of in-scanner head motion (Power et al., 2012). No slice-timing correction was performed to avoid temporal interpolation of the BOLD signal. Slice-timing correction provides little benefit with fast/short TR or multiband EPI sequences such as used in the current study (TR = 1 sec, multiband factor = 3), and has no effect on the reliability of functional connectivity estimates (Parker et al., 2017; Parker & Razlighi, 2019). The realigned rs-fMRI scans were co-

registered with the structural scans using a rigid-body transformation. The structural scans, tissue maps, and rs-fMRI scans were transformed into MNI-152 space and resampled to a 2.0 mm isotropic resolution in a single concatenated transformation step to minimize data-loss as a result of resampling. No spatial smoothing was applied.

Correction for confounding effects was performed using linear regression of the top ten principal components from the BOLD signal of WM and (ventricular) CSF maps (Behzadi et al., 2007; Chai et al., 2012), 24 head motion parameters (Friston et al., 1996; Yan et al., 2013), and scrubbing of a subject-dependent number of frames (Power et al., 2012). Scrubbing of frames with high motion ($FD > 0.30$ mm) or unusually large whole-brain BOLD signal changes (DVARs Z-score > 3.0) was performed by including a regressor for each of the flagged frames, the preceding frame, and the two following frames (Power et al., 2012). Linear regression was performed on the individual voxels of the brain after quadratic detrending of the BOLD time series to reduce the effects of scanner drift, followed by temporal bandpass filtering at the frequency range of 0.008 to 0.080 Hz (Waheed et al., 2016). All resting-state functional MRI scans were processed independently from each other.

2.5.4 Processing of task-based fMRI scans

Functional MRI scans were processed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) in MATLAB 2015b (The MathWorks Inc., Massachusetts, United States). Preprocessing involved realignment, slice timing correction, spatial normalization to MNI-152 space, and smoothing (8 mm full width at half maximum) to correct for inter-individual differences. Functional images were then submitted to a general linear model.

For both tasks two contrasts were created. For the inhibition task these were: 1) successful stops versus go trials with a stop-signal probability of zero percent, 2) successful stops versus go trials with a stop-signal probability of 20 and 33 percent (from here on referred to as $>0\%$ stop-signal probability). For the face processing task, we also created two contrasts: 1) images of faces versus rest, 2) images of faces versus

images of houses. Six realignment parameters were added as regressors of no interest to correct for head motion. All data were high-pass filtered with a cut-off of 128 seconds to control for low-frequency drifts. These analyses produced four (two contrasts per task) t-maps for each participant.

2.6 The reliability study – Statistical analysis

Test-retest reliability was quantified with ICCs and their 95% confidence intervals calculated with the irr package version 0.84.1 in R (<https://www.r-project.org/>). ICCs were computed using a single measure, absolute-agreement, 2-way random-effects model. Average ICCs were always computed after Fisher's Z transformation of the individual correlations. Percentage difference (PD) was calculated for each individual and the subsequent mean was calculated from the absolute values of the individual PDs.

2.6.1 Reliability of structural T1-weighted MRI

Global brain measures of cortical and cerebellar volume, cortical thickness and cortical surface area were used to compute mean absolute PDs and ICCs. Next, ICCs were calculated on atlas-based brain measures of subcortical volume, cortical volume, cortical surface area and cortical thickness. Additionally, ICCs were calculated for vertex-wise cortical thickness measures after resampling and smoothing.

2.6.2 Reliability of DWI

For each of the 48 WM ROIs, mean absolute PDs and were computed for FA and MD. To determine if there is a relation between certain QC characteristics and reliability of FA and MD, the mean absolute PDs were correlated with SNR (part of the QUAD results), average motion and mean displacement obtained from the QC data. For network analysis, ICCs for FA and the number of streamlines were calculated for each included edge. In addition, ICCs were calculated for the mean characteristic path length, global efficiency, mean local efficiency and mean strength (Dimitriadis et al., 2017).

2.6.3 Reliability of resting-state fMRI

The spatially-averaged BOLD signal was obtained from the unsmoothed and denoised time series for components of major resting-state networks defined in the networks atlas provided by the CONN toolbox (Whitfield-Gabrieli & Nieto-Castanon, 2012; <https://web.conn-toolbox.org/>; Supplementary Figure S1). Functional connectivity estimates were computed using full Pearson correlation between the BOLD signal of two regions. Fisher r -to- Z transformation of the functional connectivity estimates was performed prior to statistical analysis. Test-retest reliability of the Z -transformed functional connectivity estimates was assessed using the ICC as described before. For mean functional connectivity within and between resting-state networks, the ICCs were computed for the averaged Z -transformed functional connectivity estimates across all connections within or between the resting-state network(s).

2.6.4 Reliability of task-based fMRI

2.6.4.1 Behavioral reliability

For the stop-signal task behavioral ICCs were calculated for response times and accuracy. During the emotion task no behavioral data was collected.

2.6.4.2 Imaging reliability

ICCs were computed for each voxel of the brain using the unthresholded t -maps resulting from the statistical analysis in the processing phase. This voxel-wise analysis yielded a 3D matrix of Fisher transformed ICC values. An ROI-analysis was subsequently conducted using the automated anatomical labelling (AAL) template (Tzourio-Mazoyer et al., 2002), generating mean activation levels per AAL region. As these tasks were designed to elicit activation in specific regions of the brain, statistics for selected regions are reported. For the inhibition task, these are bilateral ROIs based on previous research (Vink et al., 2014; Zandbelt et al., 2013), spanning the putamen, motor cortex, and frontal and parietal lobe. As the face/house task is aimed at activating face processing areas in the brain, we report the reliability of occipital,

parietal and temporal regions of interest (Passarotti et al., 2003). In addition to statistics for specific ROIs, the mean of ICC values for all voxels across the whole brain are also reported per contrast.

2.7 The reliability study - Post-hoc analysis: Sample size estimations

To better understand the implications of our results for future studies, we did a post-hoc analysis, modelling samples size as a function of effect size Cohen's D. Power was set at 80% (beta=0.2) and the alpha level was set at 0.05. We assumed normally distributed brain measures. Cohen's D was varied between 0 and 0.5. For each scan type we used the main ICC findings as estimates of reliability, and computed sample size as $(z_{(1-\alpha/2)} + z_{(1-\beta)})^2 / (ICC * \text{Cohen's D})^2$.

3. Results

3.1 Reliability of structural T1-weighted MRI

The test-retest reliability of global structural brain measures was high (Table 2). Especially cortical and cerebellar GM volume, intracranial volume and total cortical surface area were highly replicable as indicated by a comparable mean and standard deviation between the two scan sessions, a small mean absolute PD (< 1.43%) and an excellent ICC (> 0.98). Global measures of cerebellar WM were highly reliable (mean absolute PD < 3.35%; ICC > 0.90). Average cortical thickness could be reliably measured as well (mean absolute PD < 1.25%; ICC > 0.74)

Figure 3 shows regional test-retest ICCs for subcortical and cortical brain measures. The ICCs for each region are also listed in Online Supplementary Table S1. Regional test-retest ICCs of subcortical volumes were high with an average of 0.95 (ICCs ranging from 0.84 to 0.99) over all regions in both hemispheres. Regional test-retest ICCs for cortical volumes were high with an average of 0.96 (ICCs ranging from 0.65 to 1). Regional test-retest ICCs for cortical surface area were high with an average of 0.98 (ICCs ranging from 0.53 to 1) with the lowest ICC in the left frontal pole.

Regional test-retest ICCs for cortical thickness were good with an average of 0.84 (ICCs ranging from 0.07 to 0.97) with the lowest values in the right hemisphere for the rostral middle frontal gyrus (ICC = 0.07), frontal pole (ICC = 0.48) and medial orbitofrontal gyrus (ICC = 0.51). Vertex-wise cortical thickness ICCs were high with an average ICC over all vertices of 0.88.

Taking a closer look at the low ICC in the right rostral middle frontal gyrus, we identified three participants with a large change in cortical thickness between the two scan sessions (0.16, -0.10 and -0.25 mm). We did not find artefacts in the raw scan nor segmentation errors. The vertex-wise analysis confirmed lower reliability in this region suggesting a regional effect unrelated to the parcellation atlas. We did not find evidence for an anterior-posterior gradient in vertex-wise reliability and did not find a pattern when looking at scan date or time. Focusing on the participant with the biggest change between sessions (-0.25 mm), recalculating the ICC without this participant increased the ICC in this region to 0.37 suggesting that the low ICC cannot be explained by a single outlier.

Table 2. Test-retest statistics of global brain measures.

Global brain measure (mm, mm ² or mm ³)	Mean (SD)	Mean (SD)	Mean absolute PD	ICC [95% CI]
	Test	Retest		
	ml	ml	%	
Intracranial volume	1484 (258)	1494 (262)	1.11 (1.82)	0.99 [0.98 to 1.00]
Brain without ventricles	1159 (124)	1158 (126)	0.67 (0.31)	1.00 [0.99 to 1.00]
Left cortical GM	250.6 (21.4)	249.9 (22.6)	1.09 (1.08)	0.98 [0.96 to 0.99]
Right cortical GM	252.6 (22.3)	251.9 (22.9)	1.43 (1.37)	0.98 [0.93 to 0.99]
Left cortical WM	227.4 (34.2)	227.8 (35.1)	0.73 (0.69)	1.00 [0.99 to 1.00]
Right cortical WM	228.4 (35.5)	228.8 (36.2)	0.79 (0.72)	1.00 [0.99 to 1.00]
Left cerebellum GM	55.82 (5.28)	55.82 (5.20)	0.94 (0.72)	0.99 [0.98 to 1.00]
Right cerebellum GM	54.79 (5.44)	54.78 (5.44)	0.72 (0.55)	1.00 [0.99 to 1.00]
Left cerebellum WM	15.25 (1.48)	15.15 (1.66)	3.28 (3.20)	0.90 [0.74 to 0.96]
Right cerebellum WM	14.48 (1.58)	14.42 (1.85)	3.35 (3.12)	0.93 [0.80 to 0.97]
	cm ²	cm ²	%	
Left total surface area	894.2 (95.3)	893.6 (95.6)	0.45 (0.43)	1.00 [0.99 to 1.00]
Right total surface area	895.3 (96.5)	894.9 (97.1)	0.42 (0.27)	1.00 [1.00 to 1.00]
	mm	mm	%	
Left average thickness	2.493 (0.056)	2.487 (0.062)	0.88 (0.75)	0.89 [0.72 to 0.96]
Right average thickness	2.521 (0.052)	2.514 (0.620)	1.25 (1.10)	0.74 [0.41 to 0.90]

Abbreviations: ml = milliliter; cm = centimeter; mm = millimeter; SD = standard deviation; PD = percentage difference; ICC = intraclass correlation; CI = confidence interval; GM = gray matter; WM = white matter.

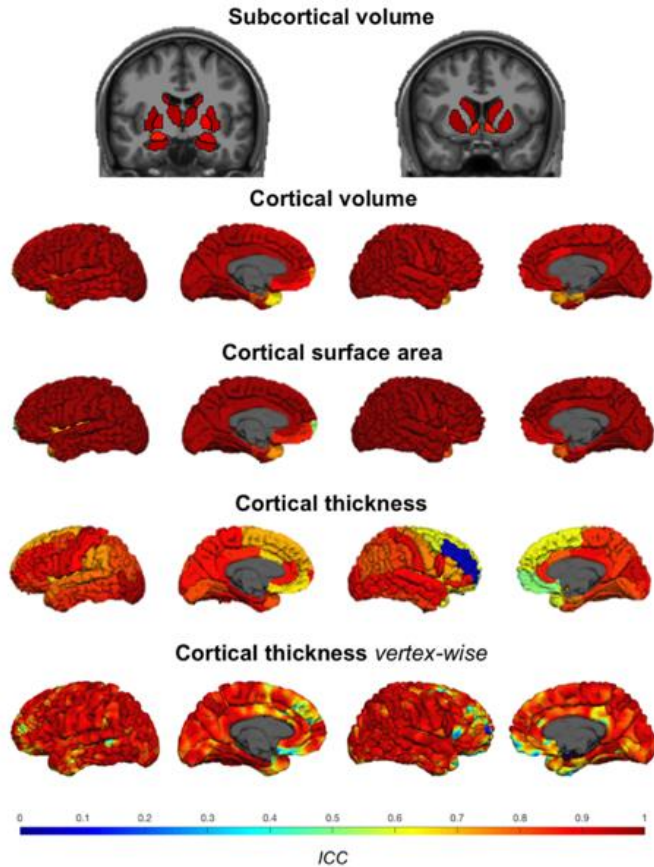


Figure 3. Test-retest ICCs of subcortical and cortical brain measures. The first row shows the ICCs of subcortical volumes on two coronal slices. The slice on the left cuts through the caudate nucleus, thalamus, putamen, pallidum, amygdala and hippocampus. The slice on the right cuts more anterior through the caudate nucleus, putamen and nucleus accumbens. The second, third and fourth row show ICCs of cortical volume, cortical surface area and cortical thickness respectively. The last row shows vertex-wise cortical thickness ICCs. The ICCs of cortical measures are shown on the surface from an outer and medial view with the left hemisphere on the left and the right hemisphere on the right. To visualize the regional test-retest reliability, a model brain was created using the first scan of each participant (Peper et al., 2009; supporting information) and segmented and parcellated with FreeSurfer. For each region or vertex, the ICC was recoded to an RGB color-code using colormap jet in MATLAB.

3.2 Reliability of DWI

3.2.1 FA and MD

The test-retest reliability and 95% confidence interval of global skeleton FA and MD was 0.94 (ICCs ranging from 0.83 to 0.98) and 0.87 (ICCs ranging from 0.65 to 0.95), respectively. The mean absolute PD for global FA was 0.86% and for global MD 1.33%. For the ROI-based test-retest analysis, the mean ICC for FA was 0.84 with values ranging from 0.51 (found in the pontine crossing tract (a part of MCP)) to 0.97 (left anterior corona radiata). The mean ICC for MD found in the test-retest analysis was 0.74, ranging from 0.09 (right cerebral peduncle) to 0.95 (fornix (column and body of fornix)). See Online Supplementary Table S2 for details.

3.2.2 Relation between scan quality and FA/MD

A significant Pearson correlation (0.60, $p = .02$) was found between the PD computed for SNR and the PD computed for global FA. For global MD the association was not significant (-0.35, $p = .19$). For relative motion, a significant negative correlation was found between the PD for relative motion and the PD for global FA (-0.51, $p = .05$) but not for MD (0.15, $p = .59$). No correlation was found between the PD computed for mean voxel displacement and the PD for FA (-0.12, $p = .67$) or MD (-0.33, $p = .23$). Online Supplementary Table S2 for test-retest results of ROIs from the JHU Atlas.

3.2.3 DWI network analysis

The ICCs computed on global network metrics with the connection-weight based on the number of streamlines and for connections weighted using FA are shown in Table 3. A total of 1053 edges were included in the connectivity maps. The mean ICC across edges was 0.52 for the number of streamlines, and 0.39 for the mean FA. Figure 4 shows the distribution of ICC's of the 1053 edges Figure 5 shows the ICCs for the mean FA (upper-left triangle) and for the number of streamlines (lower-right triangle) for each individual edge.

Table 3. Test-retest ICCs for global network metrics.

Network metric	Mean (SD)		ICC [95% CI]	
	Test	Retest		
# Streamlines	CPL	1238 (215)	1195 (289)	0.39 [-0.11 to 0.73]
	GE	0.0515 (0.007)	0.0528 (0.009)	0.88 [0.71 to 0.96]
	MLE	0.0662 (0.009)	0.0686 (0.012)	0.81 [0.56 to 0.93]
	MS	1.023 (0.130)	1.047 (0.164)	0.91 [0.76 to 0.97]
FA	CPL	2669 (471)	2745 (527)	0.64 [0.24 to 0.86]
	GE	0.0679 (0.008)	0.0680 (0.007)	0.58 [0.14 to 0.83]
	MLE	0.0799 (0.009)	0.0796 (0.009)	0.60 [0.18 to 0.84]
	MS	1.494 (0.177)	0.1498 (0.163)	0.69 [0.33 to 0.88]

Abbreviations: CPL = characteristic path length; GE =global efficiency; MLE = mean local efficiency; MS = mean strength; SD = standard deviation; ICC = intraclass correlation; CI = confidence interval; FA = fractional anisotropy.

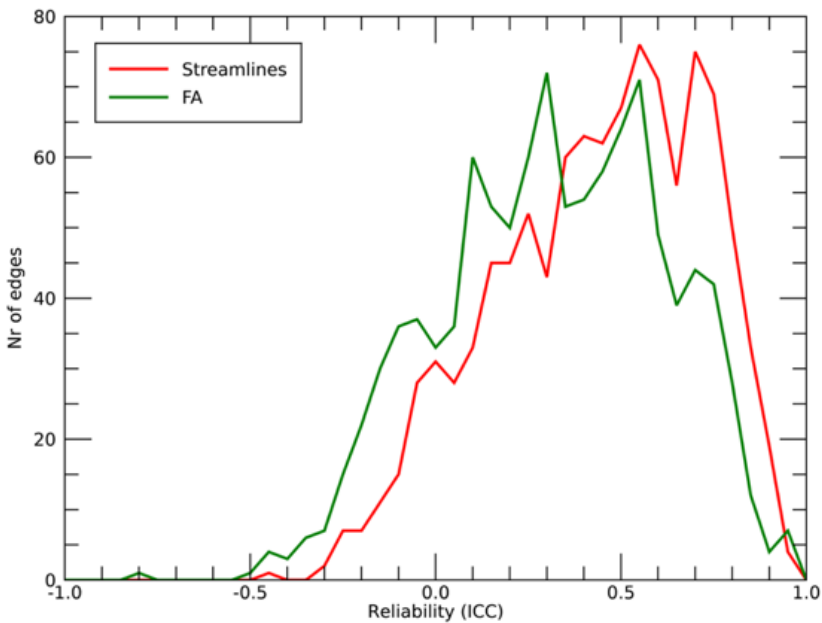


Figure 4. Histogram of the test-retest ICC's of the 1053 included edges. The bin size of the histogram is 0.05.

3.3 Reliability of resting-state fMRI

Group-mean functional connectivity was highly consistent between scan sessions as indicated by a high correlation between average connectivity at the first and second time point (Pearson's $r = 0.95$) with typical higher functional connectivity within resting-state networks and highest functional connectivity between contralateral homotopic regions (Figure 6A; Online Supplementary Table S3). Test-retest reliability of functional connectivity between regions of cortical resting-state networks was moderate (mean ICC = 0.36; ICCs ranging from -0.41 to 0.85; Figure 6B; Online Supplementary Table S4), with moderate to high test-retest reliability of average functional connectivity within cerebral cortical resting-state networks (ICCs ranging from 0.38 to 0.61; Table 4).

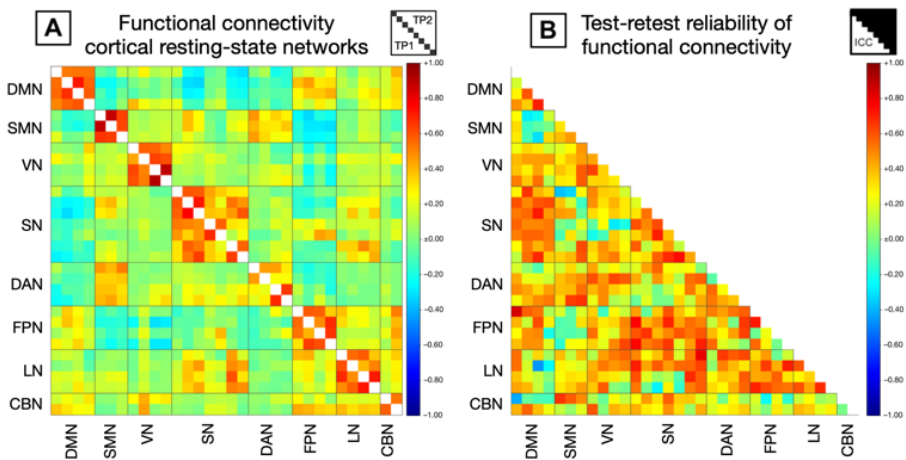


Figure 6. Group-mean functional connectivity (A) and test-retest reliability (B) of functional connectivity for connections between regions of cortical resting-state networks. Abbreviations: DMN = default mode network; SMN = sensorimotor network; VN = visual network; SN = salience network; DAN = dorsal attention network; FPN = frontoparietal network; LN = language network; CBN = cerebellar network; TP1 = estimates from test session; TP2 = estimates from retest session.

Table 4. Test-retest reliability of functional connectivity estimates within cortical resting-state networks.

Resting-state network	Mean FC-Z (SD) Test	Mean FC-Z (SD) Retest	Mean change FC-Z (SD)	ICC [95% CI]
Default mode	+0.66 (0.23)	+0.67 (0.25)	+0.01 (0.23)	0.61 [0.16 to 0.85]
Sensorimotor	+1.02 (0.39)	+0.95 (0.37)	-0.06 (0.27)	0.38 [-0.15 to 0.74]
Visual	+0.76 (0.41)	+0.79 (0.38)	+0.03 (0.29)	0.51 [0.02 to 0.80]
Saliience	+0.55 (0.29)	+0.46 (0.30)	-0.09 (0.24)	0.57 [0.10 to 0.83]
Dorsal attention	+0.39 (0.29)	+0.43 (0.29)	+0.04 (0.25)	0.48 [-0.03 to 0.79]
Frontoparietal	+0.58 (0.26)	+0.65 (0.26)	+0.07 (0.24)	0.41 [-0.11 to 0.76]
Language	+0.70 (0.29)	+0.59 (0.27)	-0.10 (0.23)	0.52 [0.02 to 0.81]
Cerebellar	+0.65 (0.26)	+0.60 (0.12)	-0.05 (0.28)	-0.01 ⁺⁺ [-0.50 to 0.49]

Abbreviations: FC-Z = r-to-Z-transformed functional connectivity; SD = standard deviation; ICC = intraclass correlation; CI = confidence interval, ⁺⁺ = lowest ICC.

3.4 Reliability of task-based fMRI

3.4.1 Behavioral reliability inhibition task

Only the inhibition task had behavioral measurements in addition to the fMRI data. The ICC for the reaction time, accuracy and response slowing measurements had an average ICC of 0.85 (Table 5). A paired-samples t-test was performed on each measure to test for possible learning effects between the two sessions. At the second session, subjects were slower in their incorrect responses, and an increase of the stop probability slope indicates that they slowed down more with increasing stop-signal probability.

Table 5. ICC values for behavioral measurements.

Contrast	ICC [95% CI]	M_1	M_2	SD_1	SD_2	t	sig
RT correct Go	0.95 [0.85 to 0.98]	851	856	36	35	-1.91	0.76
RT incorrect Stop	0.92 [0.77 to 0.97]	829	836	38	34	-2.39	0.03
Stop accuracy	0.71 ⁺⁺ [0.31 to 0.90]	0.59	0.59	0.4	0.3	0.77	0.46
Stop signal delay	0.82 [0.53 to 0.94]	211	209	28	26	0.56	0.58
Stop probability slope	0.91 [0.74 to 0.97]	91	119	61	64	-2.56	0.02

Abbreviations: ICC = intraclass correlation; CI = confidence interval, ++ = lowest ICC.

3.4.2 Imaging reliability inhibition task

Overall ICCs for the first contrast – stop versus go-trials with 0% stop-signal probability – averaged at 0.52. ICCs for the second contrast - stop versus go-trials with >0% stop-signal probability - were slightly lower, with an average of 0.44. The mean ICC of all voxels across the brain was 0.39 (range -0.76 to 0.92, median 0.47) for the first contrast, 0.37 (range -0.77 to 0.89, median 0.42) for the second. ROI ICCs can be found in Table 6.

Table 6. AAL ROI ICC statistics for the inhibition task.

AAL ROI	Stops versus go trials Stop-signal probability = 0	Stops versus go trials Stop-signal probability > 0
	ICC [95% CI]	ICC [95% CI]
Precentral gyrus	0.50 [-0.02 to 0.81]	0.49 [-0.03 to 0.80]
Superior frontal gyrus	0.54 [0.04 to 0.82]	0.48 [-0.04 to 0.80]
Middle frontal gyrus	0.60 [0.13 to 0.85]	0.48 [-0.04 to 0.80]
Inferior frontal gyrus	0.60 [0.13 to 0.85]	0.46 [-0.07 to 0.79]
Superior Temporal lobe	0.51 [0.00 to 0.81]	0.48 [-0.04 to 0.80]
Supplementary motor area	0.55 [0.05 to 0.83]	0.51 [0.00 to 0.81]
Paracentral Lobule	0.56 [0.07 to 0.83]	0.33 [-0.22 to 0.72]
Rolandic Operculum	0.50 [-0.02 to 0.81]	0.47 [-0.06 to 0.79]
Putamen	0.31++ [-0.24 to 0.71]	0.23++ [-0.32 to 0.66]

Abbreviations: ICC = intraclass correlation; CI = confidence interval, ++ = lowest ICC for each contrast.

3.4.3 Imaging reliability face processing task

For the contrast of face versus rest, the average ICC in the selected AAL regions was 0.54. For the contrast of face versus house, the average ICC in the selected AAL regions was 0.64. The mean ICC of all voxels across the brain was 0.34 (range -0.76 to 0.91, median 0.38) for the first contrast, 0.38 (range -0.55 to 0.96, median 0.43) for the second. ROI ICCs can be found in Table 7.

Table 7. AAL ROI ICC statistics for faces task.

AAL ROI	Faces versus rest	Faces versus houses
	ICC [95% CI]	ICC [95% CI]
Occipital (superior)	0.41++ [-0.13 to 0.76]	0.65 [0.21 to 0.87]
Occipital (middle)	0.53 [0.02 to 0.82]	0.63 [0.17 to 0.86]
Occipital (inferior)	0.65 [0.21 to 0.87]	0.77 [0.42 to 0.92]
Fusiform gyrus	0.47 [-0.06 to 0.79]	0.68 [0.26 to 0.88]
Inferior temporal gyrus	0.55 [0.05 to 0.83]	0.61 [0.14 to 0.85]
Superior parietal lobe	0.54 [0.04 to 0.82]	0.65 [0.21 to 0.87]
Inferior parietal lobe	0.58 [0.10 to 0.84]	0.43++ [-0.11 to 0.77]

Abbreviations: ICC = intraclass correlation; CI = confidence interval, ++ = lowest ICC for each contrast.

3.5 Post-hoc analysis: Sample size estimations

Figure S2 in shows the relationship between the reported ICCs and the sample size needed in future studies to detect an effect of interest with 80% power and an alpha level of 0.05.

4. Discussion

The YOUth MRI protocol was designed to study typical brain development longitudinally in children from 8 and up. In this paper we provide a detailed description of the MRI acquisition in YOUth and include the test-retest reliability of data collected with this protocol. Global structural brain measures could be estimated with high reliability. Regional structural and functional brain measures in ROIs or specific networks were within the ranges found in literature (outlined below per scan type).

4.1 Structural T1-weighted MRI

Regional test-retest ICCs had an average of 0.95 for subcortical volume, 0.96 for cortical volume and 0.98 for cortical surface area. Regional test-retest ICCs for cortical thickness were lower with an average of 0.84 including lower ICCs for specific regions, mostly in the right hemisphere. Vertex-wise cortical thickness ICCs were, on average, higher with an average ICC over all vertices of 0.88. For most regions, vertex-wise ICCs are comparable to those based on the parcellated region. However, in some regions the vertex-wise ICCs are on average higher than the atlas-based ICC. This difference can be explained by the fact that the between-subject variation for vertex-wise cortical thickness measures is higher than for atlas-based cortical thickness measures in these regions. Our results are in line with other studies that found higher reliability for cortical volume, compared to cortical thickness (Iskan et al., 2015; Liem et al., 2015; Wonderlick et al., 2009). One study also found lower reliability for vertex-wise cortical thickness in the right rostral middle frontal area (Wonderlick et al., 2009). In this study we wanted to have an honest and unbiased estimate of the noise in our brain measures. Therefore, we processed the T1-weighted scans and rescans separately using FreeSurfer's cross-sectional pipeline. This way, the reliability measures are valid for data obtained from only one measurement too. However, when processing YOUTH data, using FreeSurfer's longitudinal pipeline (Reuter et al., 2012) can improve reliability (Jovicich et al., 2013; Morey et al., 2010).

4.2 DWI

Reliable measures of global FA and MD were found. For the ROI-based analysis, the average ICC for FA was 0.84. The average ROI-based ICC for MD was 0.74. Another study also found FA to be more reliable than MD (Duan et al., 2015). At the network level, global network metrics were on average more reliable than metrics at the nodal level, as has been reported before (Dimitriadis et al., 2017). Global network metrics (characteristic path length, global efficiency, mean local efficiency and mean strength) were moderately reliable when weighted by FA, with ICCs between 0.58 and 0.69. The same network metrics were highly reliable when weighted by the number of streamlines, with ICCs between 0.81 and 0.91, with the exception of

characteristic path length that was unreliable, ICC=0.39, comparable to what was found in another study (Cheng et al., 2012). Reliability was lower at the nodal level, with a mean ICC across edges of 0.52 for the number of streamlines, and 0.39 for the mean FA. Numerous methodological choices exist for DWI data, which makes it difficult to directly compare our findings to literature (for an extensive review see: Welton et al., 2015).

4.3 Resting-state fMRI

Group-mean functional connectivity was consistent between scan sessions with higher functional connectivity within resting-state networks and highest functional connectivity between contralateral homotopic regions typically observed for cortical resting-state networks. Test-retest reliability of functional connectivity between regions of cortical resting-state networks was moderate with an average ICC over all networks of 0.36, partially due to poor reliability within the cerebellar network. When looking at only cerebral cortical resting-state networks, ICCs were in the range of 0.38 to 0.61. A recent meta-analysis reported an average reliability of 0.29 for functional connectivity on edge-level based on 25 studies (Noble et al., 2019).

4.4 Task-based fMRI

The inhibition task had highly reliable behavioral measurements with an average ICC of 0.85. MRI measures during this task had an average ICC over the ROIs of 0.44 and 0.52 for the two task contrasts. MRI measures during the emotion task had an average ICC over the ROIs of 0.54 or 0.64. The contrast between faces and houses generated a more reliable response than the contrast of faces versus rest. These results are in line with ICC values of pre-defined ROIs in other task-based fMRI studies. A meta-analysis of 13 fMRI studies between 2001 and 2009 reported ICCs values in a range from 0.16 to 0.88, with an average reliability of 0.50 (Bennett & Miller, 2010). Similar to our results, reliability generally tends to be best for occipital regions (Koolschijn et al., 2011; Vetter et al., 2015, 2017) and fair to poor for frontal and subcortical regions (Herting et al., 2018). Whole-brain average ICCs were lower than ROI ICCs for both tasks, suggesting that the task contrasts more accurately

modulate activity in the targeted ROIs than in other areas. Voxel-wise calculations are a stringent measure of reliability and indicate whether the level of activity in all voxels is consistent between test and retest (Bennett & Miller, 2010).

4.5 Factors that determine reliability

In literature, ICCs for functional MRI measures are generally deemed lower compared to structural MRI measures. Our findings are in line with other studies that show that structural MRI brain measures can be measured more reliably than fMRI brain measures. ICC is related to statistical power and therefore the threshold of an acceptable ICC depends on the included sample size and the size of the effect of interest. In MRI research, noise may arise from subject- and MRI-related factors, and their interaction. Effective processing methods can ensure that the effect of noise on the brain measures are kept to a minimum. The impact of methodological choices is reviewed for studies on structural (Mills & Tamnes, 2014; Vijayakumar et al., 2018) and functional brain development (Bennett & Miller, 2010; Herting et al., 2018; Telzer et al., 2018). In-depth investigation of the origin of the noise in our data is beyond the scope of this paper. However, based on the literature we can speculate on possible sources of the noise.

Our acquisition parameters were chosen to create an optimal tradeoff between acquisition duration and SNR/SFNR (e.g. high field strength, isotropic voxels, multiband, scan duration, validated fMRI tasks) and scans were processed using widely-used software. Still, MRI remains a very sensitive measurement technique that inherently has some degree of instability, which may vary per MRI scanner. Consequently, scanner performance is monitored using human and phantom data throughout the YOUth study. Variation is amongst others introduced by scanner drift due to gradient heating and differences between scan sessions with regard to the positioning of participants and variations in shimming (i.e. correcting inhomogeneities of main magnetic field). Therefore, reported results are specific to our scanner, acquisition, processing software and study sample.

Subject movement remains the foremost cause of low reliability of fMRI signals (Gorgolewski et al., 2013b). It has been shown before that residual movement contamination is left in the fMRI BOLD signal even after motion correction (Power et al., 2012). Similarly, our reliability study shows residual variation in DWI scans related to SNR even after correcting for motion. Motion can be a problematic source of variation in longitudinal research as it can be age-related and heritable (Achterberg & van der Meulen, 2019; Savalia et al., 2017; Teeuw et al., 2019; Van Dijk et al., 2012). Therefore, it is important to implement a stringent motion correction technique and QC. Additionally, QC measures, like SNR and SFNR may be included as covariates in DWI and fMRI studies, respectively (Farrell et al., 2007; Friedman et al., 2006; Friedman & Glover, 2006).

For task-based fMRI, additional sources of variation may be introduced by practice effects and compliance to the scanner procedure. Variation induced by the latter can be reduced by familiarizing participants with the MRI environment before the scanning session using a mock scanner as is done within the YOUth cohort. Other subject-related noise can occur due to dehydration (Duning et al., 2005; Kempton et al., 2009; Nakamura et al., 2014; Streitbürger et al., 2012), or caffeine intake (Laurienti et al., 2002). Finally, the type and complexity of the task used with an fMRI measurement can greatly affect reliability, with simple motor-movement tasks generally being more reliable than tasks requiring complex cognitive strategies (Gorgolewski et al., 2013a, 2013b).

Scan duration can also greatly affect reliability in fMRI (Birn et al., 2013; Shah et al., 2016; Termenon et al., 2016). A resting-state acquisition duration of approximately 8 minutes used in the YOUth cohort study is at the minimum recommended duration (Birn et al., 2013). However, the high temporal resolution (TR of 1 second) provides additional sampling points to still achieve a robust measurement within the limited time window. The quality assurance protocol of the YOUth cohort study ensures high temporal SNR (Figure 1), and might be further improved by early-stage denoising strategies (Adhikari et al., 2018). Denoising strategies to combat the influence of random fluctuations due to physiological noise can result in cleaner estimates of functional connectivity (Caballero-Gaudes & Reynolds, 2017; Parkes et al., 2018),

although no optimal strategy currently exists. In some cases, denoising procedures may decrease reliability statistics as reproducible artefacts are also removed (Noble et al., 2019). On a whole, fMRI measurements, such as functional connectivity, are dynamic and state-dependent (Poldrack et al., 2015). As such, longitudinal changes might be due to developmental changes intrinsic to the brain or due to extrinsic factors such as mood, sleep quality, or substance use (Poldrack et al., 2015).

4.6 Relevance of reliability results and the relation to power

First, the ICCs reported in this study can be useful to researchers that want to adopt our acquisition parameters (listed in online Appendix 1). Secondly, it shows how different modalities and processing methods relate to each other in terms of reliability (e.g. FA in ROIs versus FA on edge-level). Lastly, the results can inform researchers that want to apply for data collected in YOUth. Because researchers with all types of research questions can apply for data, in this study we aimed to show reliability measures for each scan using methods that are well-known and widely-used in the field. Our reliability results should not be used to refrain from studying certain brain measures as all of them can be relevant when studying brain development. However, the reliability results can provide guidance when making methodological choices. Accounting for exclusions due to MR safety criteria, scanner-related distress or artefacts, a sample size of 1500 for each type of scan seems sufficient to detect an effect size of 0.2 (Figure S2). Furthermore, this analysis shows that it is not advised to apply for small subsamples of the MR data in YOUth, particularly when one is interested in regional measures of DWI on network-level and (rs-)fMRI data.

4.7 Limitations

This test-retest study has several limitations. First, the test-retest sample consists of adults, while the YOUth study focuses on development in children. Therefore, the reliability of brain measures found in this study may be considered an overestimation since it does not reflect pediatric data. Consequently, the number of good quality pediatric scans needed to obtain enough power to detect a certain effect is likely

higher than estimated in Figure S2. In general, more in-scanner head motion is seen in children compared to adults (Poldrack et al., 2002; Satterthwaite et al., 2013; Thomas et al., 1999), but not in all studies (Alexander-Bloch et al., 2016; Koolschijn et al., 2011). Furthermore, processing pediatric data comes with challenges. For example, the processing pipelines used in this study use adult templates as reference for spatial normalization, registration and segmentation. Studies show that using adult templates for pediatric data rather than age-appropriate templates introduces bias in brain measures (Fonov et al., 2011; Poldrack et al., 2002; Wilke et al., 2002, 2008; Yoon et al., 2009). A second limitation can be that the practice effect (for task-fMRI) and compliance effect in this short test-retest period cannot be compared to the three-year scan interval in YOUth. A third limitation is that the test-retest sample size is in conformance with common practice but not big enough to mitigate the effect of regional outliers.

4.8 Conclusion

It has been shown that neuroimaging studies are often underpowered with the risk of false positive results (Button et al., 2013). Statistical power can be boosted by increasing reliability and sample size. In YOUth, the large sample size together with reasonable to good test-retest reliability increases the probability of finding subtle developmental effects. This paper provides a transparent report of the methodology used in YOUth from MRI acquisition to monitoring quality and reliability. The reliability study shows promising results for the studies that will be done using MRI data collected within the YOUth cohort.

5. Supplementary materials

Some of the supplementary files were too large to be incorporated here and can be downloaded online via this link <http://doi.org/10.17605/OSF.IO/M5R3U>. For Chapter 2, the online supplement includes one file with a printout of all acquisition parameters for the human and phantom YOUNG MRI protocol and one file with supplementary tables S1-S4 listing test-retest reliability in local regions for T1-weighted, DWI and resting-state functional connectivity measures.

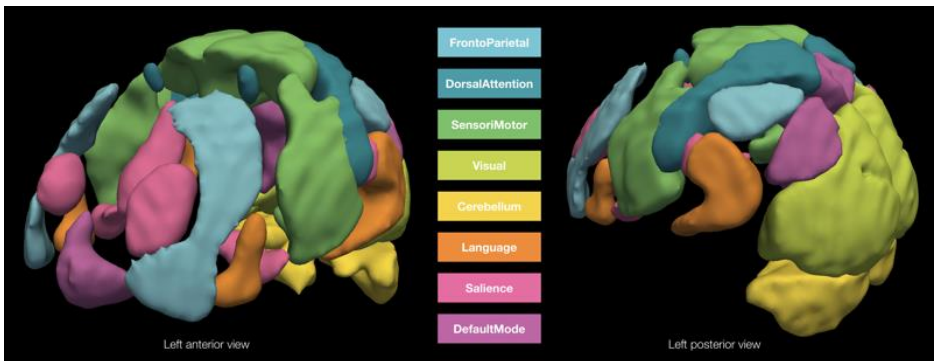


Figure S1. Atlas of canonical resting-state networks and their spatially distinct components provided by the CONN toolbox version 18.a based on ICA decomposition of 497 young adults from the Human Connectome Project.

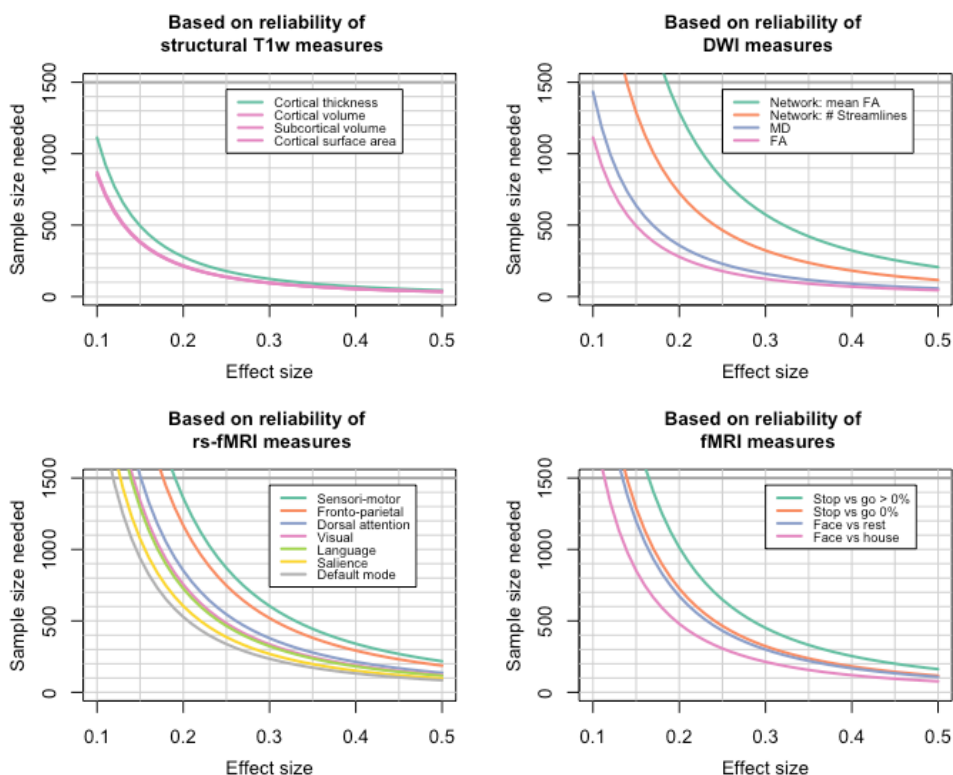


Figure S2. Sample size estimations for each type of scan based on the ICCs. The sample size (y-axis), needed to detect an effect of interest with 80% power ($\beta=0.2$ and $\alpha=0.05$), is modelled for different effect sizes (Cohen's D, x-axis). For each type of scan the test-retest ICCs are used to estimate the amount of error. All legends are ordered from lowest ICC to higher ICCs. Top left, for the structural T1-weighted scans the average of the regional ICCs is used for cortical thickness, cortical volume, subcortical volume and cortical surface area (ICCs lowest to highest: 0.84, 0.95, 0.96 and 0.98). Please note that the last three lines are nearly identical and therefore overlap in the plot. Top right, for the diffusion-weighted images the ICCs correspond to the average FA and number of streamlines over all edges and the average regional ICC of the MD and FA values (ICCs lowest to highest: 0.39, 0.52, 0.74 and 0.84). Bottom left, for the resting-state fMRI scans the ICCs used for the sample size estimation correspond to the reliability of the functional connectivity within cerebral cortical resting-state networks (ICCs lowest to highest: 0.38, 0.41, 0.48, 0.51, 0.52, 0.57 and 0.61). Bottom right, for the task-based fMRI the average regional ICCs used for the sample size estimation correspond to two contrasts for the inhibition task and two contrasts for the emotion task indicated in the legend (ICCs lowest to highest: 0.44, 0.52, 0.54 and 0.64).

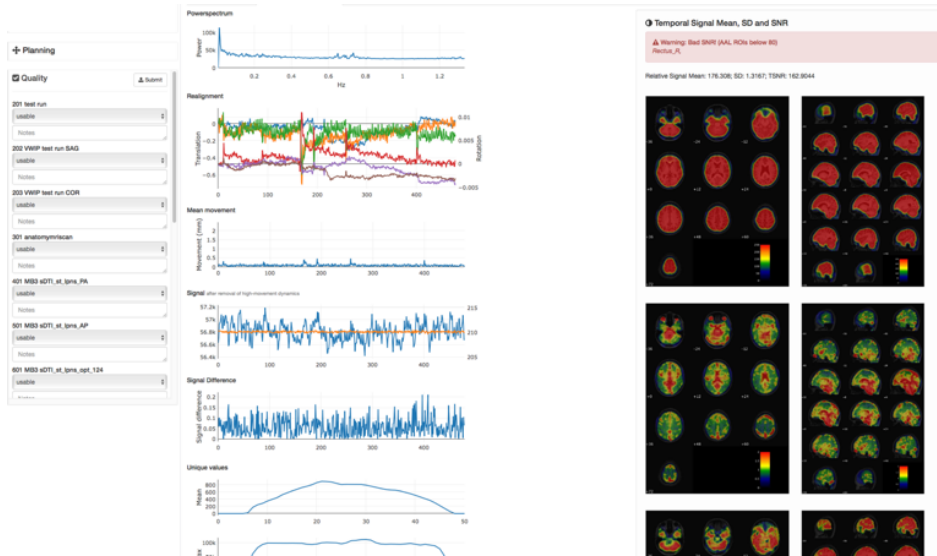
Example of quality control output that is generated in YOUth after each session

Examples are obtained from different reports and intended just as illustration. Participant identifiers and test dates are masked.

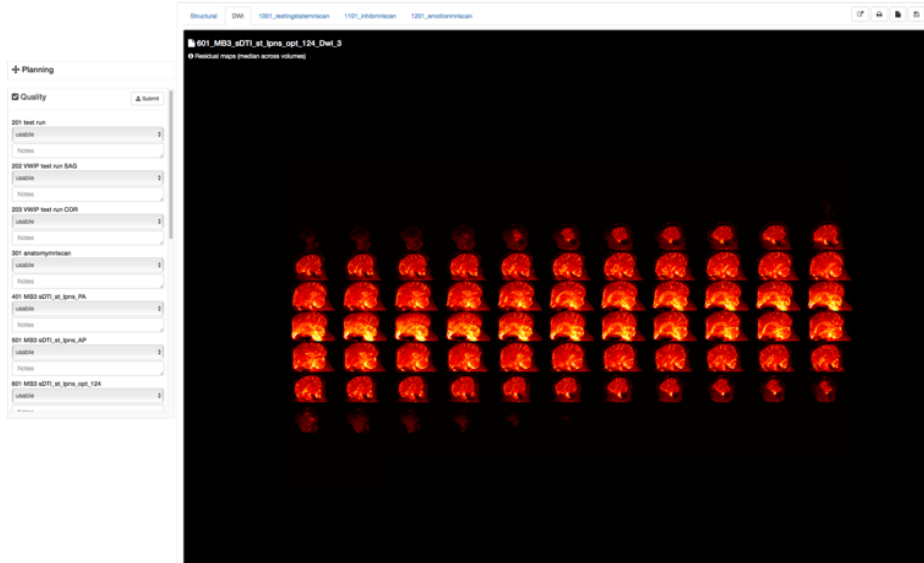
Quality control of T1-weighted scans

The screenshot displays the YOUth quality control interface. On the left, there is a 'Planning' table and a 'Quality' section. The 'Quality' section includes checkboxes for '201 test run', '202 VSWP test run SAG', '203 VSWP test run COR', '201 anatomy/mri/can', '201 M33 a2T1_wt_sara_PA', '201 M33 a2T1_wt_sara_AP', and '201 M33 a2T1_wt_sara_sag 124'. The main area shows a header with metrics: IQR (1.432), Noise Contrast Ratio (1.2134), Inhomogeneity Contrast Ratio (2.0403), and RMS resolution (1.5261). Below the header, the selected scan '201_anatomy/mri/can' is shown with three brain slices: an axial slice, a sagittal slice, and a coronal slice. Each slice has a 'Select' button and a progress indicator below it.

Quality control of Functional MRI scans



Quality control of DWI-scans



2

Including the Eddy quad report:



Single subject QC report generated using eddy quad v1.0.2

When using eddy and its QC tools, we ask you to please reference the papers describing the different aspects of the modelling and corrections. The following suggestion for a methods section and list of references has been tailored for you based on your eddy command line.

METHODS

The susceptibility induced off-resonance field was estimated from spin-echo EPI images acquired with different phase-encode directions (Andersson et al., 2003). This field was passed to "eddy", a tool that combined it with estimating gross subject movement and eddy current-induced distortions (Andersson & Sotiropoulos, 2016). The quality of the dataset was assessed using the eddy QC tools (Bastiani et al., 2019). Slices with signal loss caused by subject movement coinciding with the diffusion encoding were detected and replaced by predictions made by a Gaussian Process (Andersson et al., 2016).

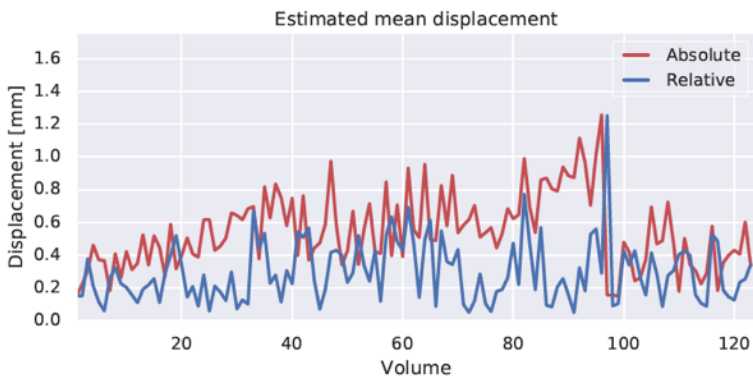
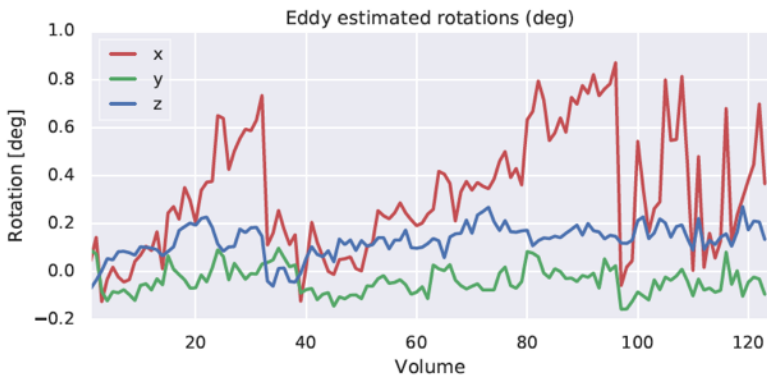
REFERENCES

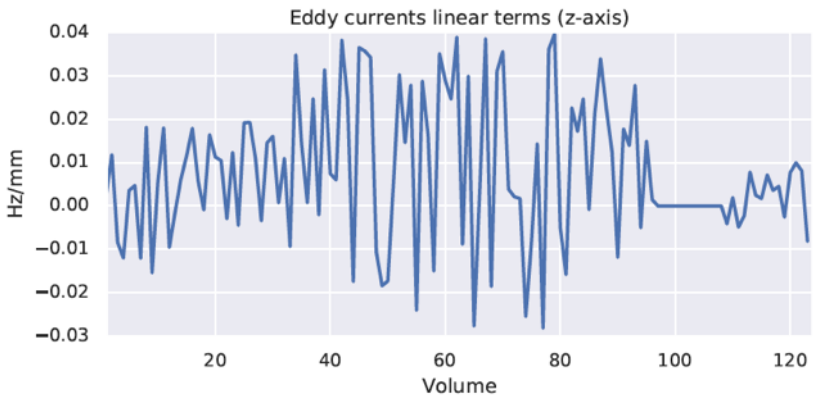
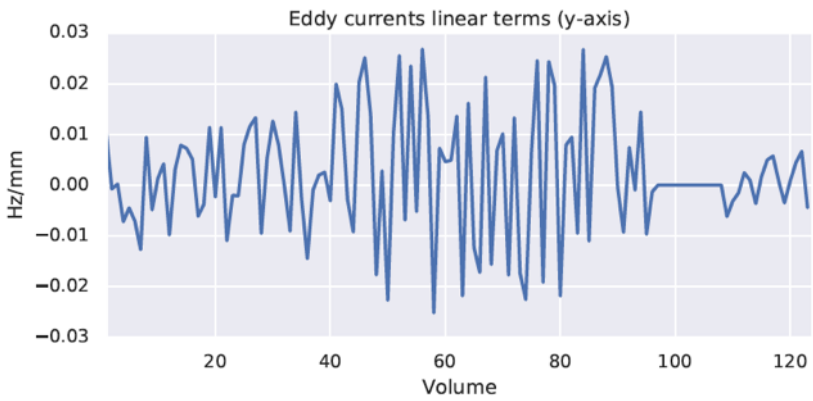
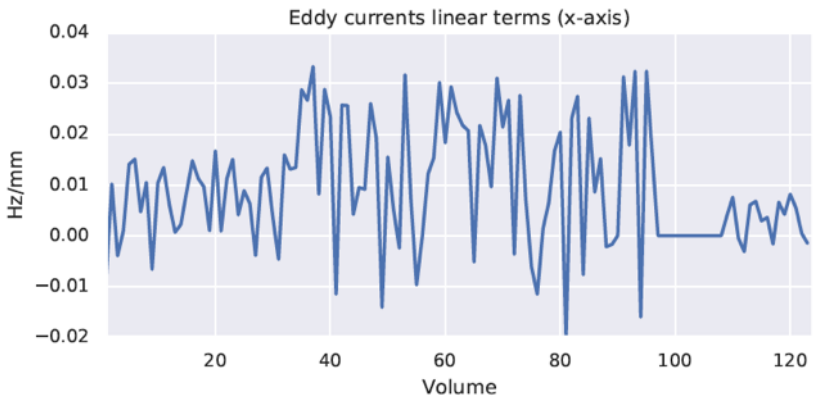
Jesper L.R. Andersson, Stefan Skare and John Ashburner. 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20:870-888

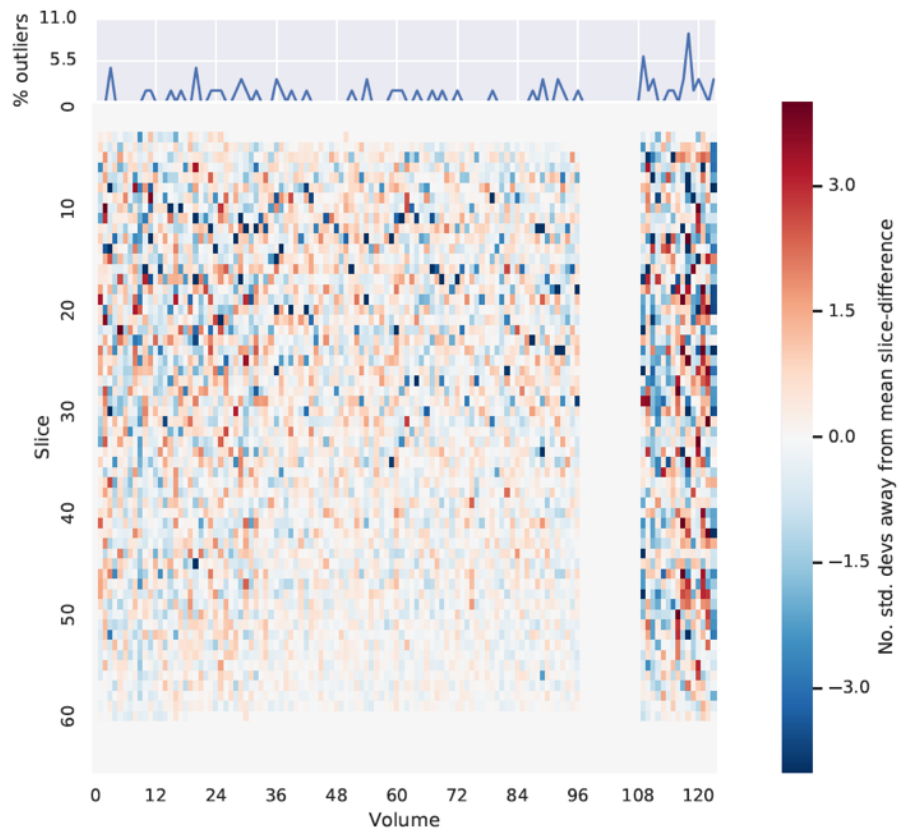
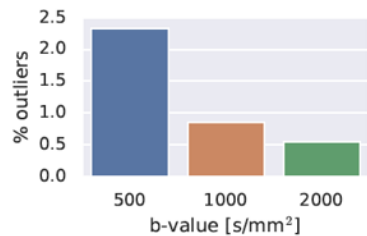
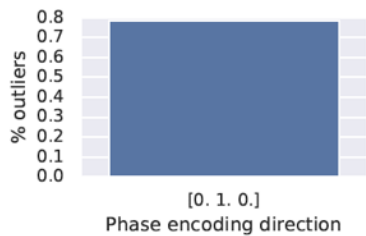
Jesper L.R. Andersson and Stamatios N. Sotiropoulos. 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* 125:1063-1078

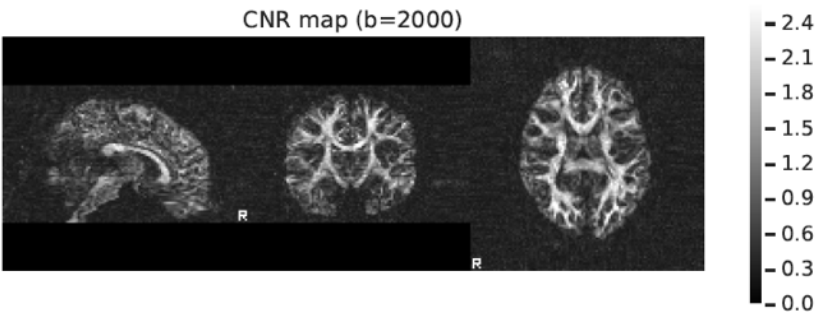
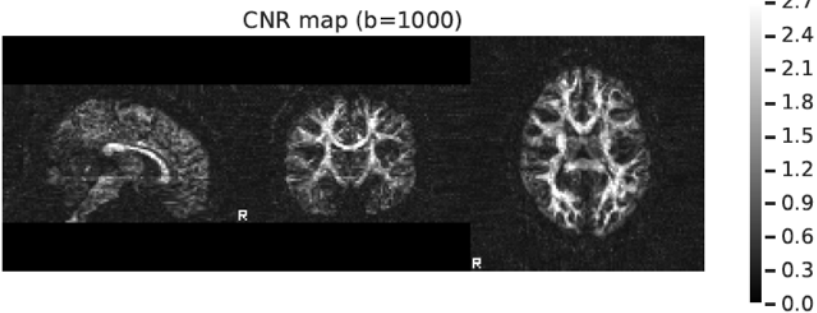
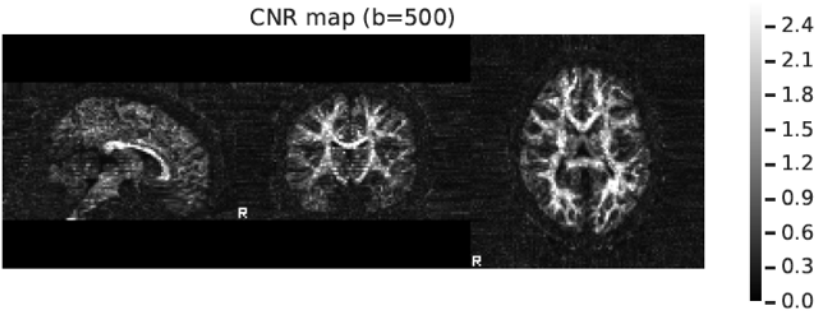
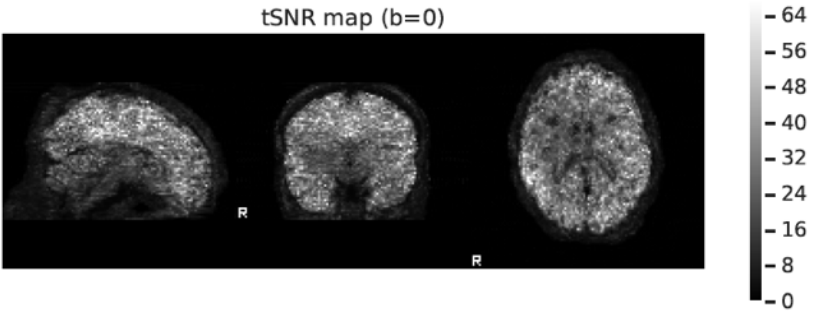
Matteo Bastiani, Michiel Cottaar, Sean P. Fitzgibbon, Sana Suri, Fidel Alfaro-Almagro, Stamatios N. Sotiropoulos, Saad Jbabdi and Jesper L.R. Andersson. 2019. Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage* 184:801-812

Jesper L.R. Andersson, Mark S. Graham, Eniko Zsoldos and Stamatios N. Sotiropoulos. 2016. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage* 141:556-572

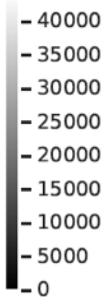
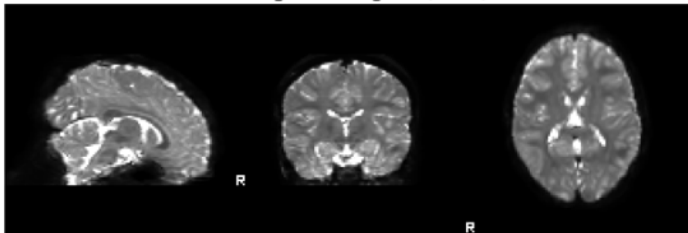




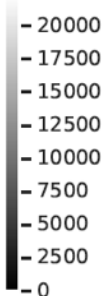
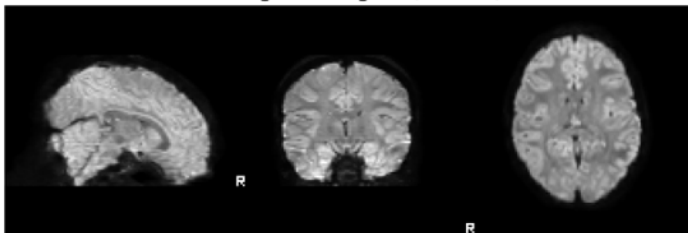




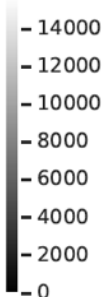
Average DW signal (b=0)



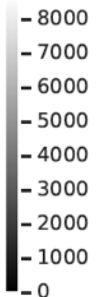
Average DW signal (b=500)

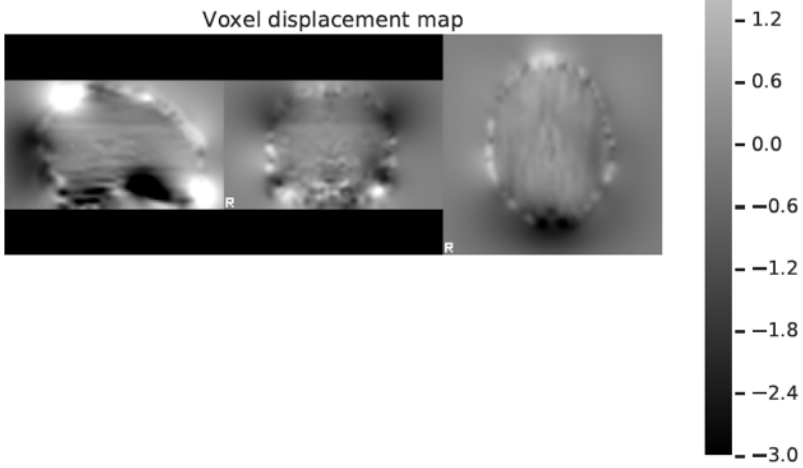
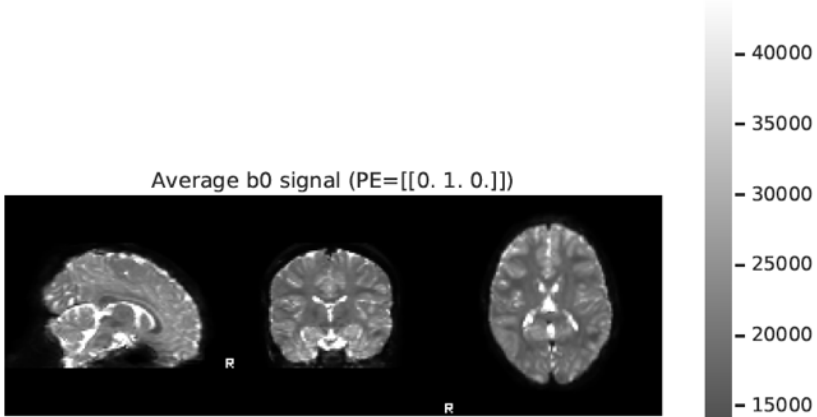


Average DW signal (b=1000)



Average DW signal (b=2000)





Chapter 3 - De-identification procedures for magnetic resonance images and the impact on structural brain measures at different ages.

Elizabeth E.L. Buimer, Hugo G. Schnack, Yaron Caspi, Neeltje E.M. van Haren, Mikhail Milchenko, Pascal Pas, Alzheimer's Disease Neuroimaging Initiative, Hilleke E. Hulshoff Pol, Rachel M. Brouwer

Published in 2021

Human brain mapping, 42 (11), 3643-3655

<https://doi.org/10.1002/hbm.25459>

Abstract

Surface rendering of MRI brain scans may lead to identification of the participant through facial characteristics. In this study we evaluate three methods that overwrite voxels containing privacy-sensitive information: Face Masking, FreeSurfer defacing and FSL defacing. We included structural T1-weighted MRI scans of children, young adults and older adults. For the young adults, test-retest data was included with a one-week interval. The effects of the de-identification methods were quantified using different statistics to capture random variation and systematic noise in measures obtained through the FreeSurfer processing pipeline. Face Masking and FSL defacing impacted brain voxels in some scans especially in younger participants. FreeSurfer defacing left brain tissue intact in all cases. FSL defacing and FreeSurfer defacing preserved identifiable characteristics around the eyes or mouth in some scans. For all de-identification methods regional brain measures of subcortical volume, cortical volume, cortical surface area and cortical thickness were on average highly replicable when derived from original versus de-identified scans with average regional correlations >0.90 for children, young adults and older adults. Small systematic biases were found that incidentally resulted in significantly different brain measures after de-identification, depending on the studied subsample, de-identification method and brain metric. In young adults, test-retest intraclass correlation coefficients (ICCs) were comparable for original scans and de-identified scans with average regional ICCs >0.90 for (sub)cortical volume and cortical surface area and ICCs >0.80 for cortical thickness. We conclude that apparent visual differences between de-identification methods minimally impact reliability of brain measures, although small systematic biases can occur.

1. Introduction

Advances in magnetic resonance imaging (MRI) technology enable researchers to collect good quality structural MRI scans of the brain. However, brain scans also contain privacy-sensitive facial characteristics. The participants' faces can be reconstructed with 3D rendering software that is part of most MRI viewers. Recently, this topic received attention from the scientific community as well as popular journals after the release of a study in which face recognition software was used to identify participants based on their MRI scan (Schwarz et al., 2019). This study complements an earlier study that shows how participants can be identified through their 3D renders by humans (Prior et al., 2009). In light of the increasing number of worldwide (public) neuroimaging collaborations (Poline et al., 2012) and technical improvements, the question is whether sharing raw anatomical MRI images is still in line with privacy regulations (White et al., 2022). From an ethical viewpoint sharing identifiable data may compromise the confidentiality participants consented to. For these reasons, more and more open-access datasets contain MRI scans that were subjected to some type of de-identification method.

Different efforts can be made to de-identify MRI scans (listed here: open-brain-consent.readthedocs.io/en/stable/anon_tools.html). First, brain extraction or skull-stripping removes non-brain tissue. Second, defacing algorithms remove facial and dental characteristics. Examples of defacing methods are: `fsl_deface` (Alfaro-Almagro et al., 2018; fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL), `mri_deface` (Bischoff-Grethe et al., 2007; freesurfer.net/fswiki/mri_deface), `pydeface` (github.com/poldracklab/pydeface), `QuickShear` (github.com/nipy/quickshear/), `mrdefacer` (github.com/mih/mrdefacer/). Third, Face Masking (Milchenko & Marcus, 2013; nrg.wustl.edu/software/face-masking) masks the face and, optional, the ear region and thus preserves more anatomical landmarks than defacing methods. An advantage of the latter toolbox is that it can be applied to raw DICOM images which limits the number of intermediate processing steps. In a recent study a new method is introduced, `mri_reface`, to replace voxels in the face and ear region with a population average (Schwarz et al., 2021). In this article we will focus on defacing with FSL, a method developed by and used in the UK biobank study (Alfaro-Almagro et al., 2018) defacing with FreeSurfer, for

example used in the BRAINS (Job et al., 2017), CamCAN (Taylor et al., 2017), ATLAS (Liew et al., 2018) and HID study (Ozyurt et al., 2010) and Face Masking which is implemented in XNAT (Marcus et al., 2007a), used in the repositories of OASIS (Marcus et al., 2007b), HCP (Marcus et al., 2013) and GSP (Holmes et al., 2015).

De-identification methods are typically optimized for healthy adults, but different results can be expected in other populations for example related to the amount of atrophy, i.e., the closeness of the brain to the skull. A recent study investigated the effects of different de-identification methods (QuickShear, Face Masking and FreeSurfer defacing) in patients with multiple sclerosis, Alzheimer's Disease (AD) and glioblastoma (de Sitter et al., 2020). In this study, we focus on age-related effects by including children, young adults and older adults (with and without AD). From an ethical perspective, children are an extra sensitive population. Possible age-related effects of de-identification procedures on brain measures are also relevant in the light of longitudinal studies investigating brain development or ageing.

An optimal de-identification method (i) prevents participant identification, (ii) leaves brain tissue intact and (iii) has a negligible effect on brain measures, i.e. the de-identified scan should approximately generate the same results as an unmasked scan. Previous research shows that these criteria are not always met and even after de-identification some participants can still be identified with facial recognition software (Abramian & Eklund, 2018; Schwarz et al., 2021). Furthermore, defacing can overwrite a small amount of brain voxels in some cases (Alfaro-Almagro et al., 2018). Lastly, de-identification procedures impact subsequent processing and outcome measures (de Sitter et al., 2020; Holmes et al., 2015; Schwarz et al., 2021). In this study we focus on the effect of the three de-identification procedures on brain measures, but we also describe visual aspects of the methods.

To evaluate the de-identification techniques, we started with a visual check to rate the invasiveness, i.e. whether brain voxels are preserved, and to rate the presence of eyes and mouth characteristics after de-identification. Secondly, we assessed whether de-identification altered brain measures differently in children, young adults and older adults. To this end, we extracted regional subcortical and cortical volumes,

cortical surface area and cortical thickness of original scans and de-identified scans. Lastly, we quantified how the effect of de-identification techniques on brain measures compared to test-retest reliability in young adults.

2. Materials and methods

2.1 Samples

2.1.1 Sample of children

The children's data consisted of 25 children from the general population (8 male) with a mean age of 9.5 (0.9) years within a range from 8 to 11 years. This data was collected as part of the first wave in a large longitudinal study on brain development in Utrecht, the Netherlands: The YOUth cohort study (Onland-Moret et al., 2020). The subset included in this study is labeled as pilot data. The children's parents or guardians gave written consent. Included MR images were of good quality after exclusion of scans with poor contrast or major motion artefacts such as ringing.

2.1.2 Sample of young adults

The sample of young adults consisted of 16 volunteers from the general population (6 male) with a mean age of 23.6 (3.3) years within a range from 19 to 31 years. To assess the test-retest reliability of the YOUth MRI protocol, the adults were scanned twice using the same acquisition parameters. The scan-rescan interval was between 6 and 8 days. The adult dataset was acquired in the context of protocol development. The adults signed written informed consent. All available scans were of good quality without major motion artefacts or other artefacts.

2.1.3 Sample of older adults

The elderly sample was selected from the large Alzheimer's Disease Neuroimaging Initiative (ADNI, adni.loni.usc.edu). ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal

of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Selected scans were of good quality without major motion or dental artefacts based on visual quality control in addition to provided quality assessment codes. Furthermore, the subset was created to ensure a heterogenous sample with regard to age, sex and patient-control status. The older adults sample consisted of 43 elderly participants: 22 participants (7 male) with Alzheimer's disease with a mean age of 72.6 (7.5) years ranging from 56 to 86 and 21 age-matched participants (9 male) without cognitive impairment with a mean age of 74.6 (5.9) years ranging from 65 to 85.

2.2 Acquisition parameters

The ADNI subset was selected with uniform acquisition parameters, that resembled the parameters of the YOUth MRI protocol, used for the acquisition of the child and adult data. All acquisition parameters can be found in Table 1.

Table 1. Acquisition parameters

Parameters	Children and young adults YOUth MRI protocol	Older adults ADNI 2 (Philips: MPRAGE)
Multicenter	No, a single MR scanner	Yes
Type of MR scanner	Philips Ingenia	Philips Achieva
Field strength (T)	3.0	3.0
Head-coil	32-channel SENSE head-coil	8-channel SENSE head-coil
Scan	Structural T1W 3D GRE	Structural T1W 3D GRE
Scan orientation	Sagittal	Sagittal
TR (ms)	10	6.8
TE (ms)	4.6	3.1
Flip angle (degrees)	8	9
Field of view (mm)	240 × 240 × 200	256 × 240 × 204
Acquisition matrix	304 × 304	256 × 240
Reconstructed voxel size (mm ³)	0.75 × 0.75 × 0.80	1.00 × 1.00 × 1.20

Abbreviations: T=Tesla; T1W=T1-weighted; GRE=gradient echo; TR = repetition time; TE = echo time.

2.3 De-identification methods

All scans were subjected to three de-identification methods. FreeSurfer defacing was applied using version 1.22 of the `mri_deface` function (Bischoff-Grethe et al., 2007; https://surfer.nmr.mgh.harvard.edu/fswiki/mri_deface) in FreeSurfer version 6.0 (Fischl et al., 2002). FSL defacing was applied using version 1.0.0 of the `fsl_deface` function (Alfaro-Almagro et al., 2018) in FSL 6.0.1 (Jenkinson et al., 2012; <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>). The Face Masking toolbox was applied with default coarseness and an ear mask using version 12/26/2017 of the `mask_face` function (Milchenko & Marcus, 2013; nrg.wustl.edu/software/face-masking). The toolbox is implemented in neuroinformatics platform XNAT (Marcus et al., 2007a), but in this article we used the offline toolbox.

For Face Masking, we additionally compared different mask settings in children and young adults by varying the coarseness of the mask. Furthermore, we investigated the effect of switching the ear mask option on or off (-e flag). The coarseness of the mask was varied between 0.1 to 1.2 in steps of 0.1 (default value is 1) by adjusting the grid step variable (-s flag). The coarseness variable applied to both the face and ear mask.

2.4 Visual inspection

A visual inspection of the de-identified scans was performed with MRICroGL (version 14 July 2017, www.mccauslandcenter.sc.edu/mricrogl/). A 3D rendering was created to inspect the effect of de-identification on the face and ear characteristics. Two raters rated the fraction of participants in each age group where the eyes or the mouth were preserved after de-identification. Next, all 2D axial brain slices were checked and coded by three independent raters. The raters rated whether brain tissue was left intact or whether more than a few brain voxels were removed or blurred by the de-identification method. For FSL defacing and Face Masking, the invasiveness of the method in the ear area was assessed separately. FreeSurfer defacing currently does not provide an option to remove ears.

Additionally, one rater assessed the invasiveness of different coarseness settings in the child sample and the adult sample (first time point). The mask setting that resulted in no overlap with brain tissue in any participant (child or adult) was labeled “the non-invasive mask”.

2.5 Visualization of de-identification

To show the effects of de-identification while preventing participant identification, we created average brains (Caspi et al., 2020; Peper et al., 2009). In short, the individual scans were registered to Talairach space and corrected for non-uniformity followed by a series of linear and nonlinear warpings of the scans (Collins et al., 1995). The average child brain was created by averaging the scans of all 25 children. The average young adult brain was created using the first scan of each young adult, 16 in

total. The average older brain was created by averaging the scan of all older participants, 43 in total. To visualize the effect of de-identification on the brain (Figure 1), the average brains were used as input to the masking and defacing tools.

2.6 MRI processing

Face Masking was applied to the raw DICOMs, after which they were converted from DICOM to NIfTI format (dcm2niix, <https://github.com/rordenlab/dcm2niix>) together with the original DICOM files. Face Masking could not be applied to NIfTI format. Both defacing tools accept only NIfTI input and were therefore applied to the original scans after conversion to NIfTI. FreeSurfer version 6.0 was used for automatic brain segmentation and parcellation (Fischl et al., 2002; freesurfer.net). Global and regional brain measures of subcortical volume, cortical volume, cortical surface area and cortical thickness were extracted. The Desikan-Killiany atlas was used for cortical parcellation (Desikan et al., 2006). No additional quality check procedures were performed on the segmentations and parcellations. Besides atlas-based measures of cortical thickness, vertex-wise cortical thickness was extracted. For the vertex-wise analysis, cortical thickness of each scan was resampled to the average brain. After resampling, the cortical surfaces were smoothed with a 3D Gaussian kernel (FWHM = 10 mm).

2.7 Statistical analysis

We computed several measures to assess the impact of de-identification on brain measures: the intraclass correlation coefficient (ICC) of absolute agreement (Bartko & Carpenter, 1976; Koo & Li, 2016; McGraw & Wong, 1996; Shrout & Fleiss, 1979) is traditionally used in test-retest context and captures random variation as well as systematic biases. To disentangle these two types of variation and get a better understanding of the effects of de-identification on brain measures, we computed Pearson correlation coefficients (Pearson's r) for coherence, and paired t-tests and signed percentage differences (PD) for assessment of systematic bias. As an indication of the variance introduced by de-identification we also computed the

coefficient of variation (CoV) and absolute PDs. All these statistics were computed for the comparisons described below.

First, we investigated the differences between the original and the de-identified scans in children, young adults (first time point) and older adults. The ADNI data was analyzed as a whole and in two separate groups based on diagnosis of Alzheimer's disease. Second, the scan-rescan adult data was used to calculate the test-retest reliability of brain measures from de-identified scans and compare this to the test-retest reliability of the original scans. Last, to separate the effect of de-identification from test-retest reliability, we calculated the agreement between brain measures from the first scan session and the second scan session, masking or defacing only the second scan. The test-retest reliability of the original scans in this study was also reported elsewhere (Buimer et al., 2020).

Additionally, the coarseness setting of the non-invasive mask was defined based on the visual inspection described above. To compare the mask with default coarseness and the non-invasive mask, both with and without ear mask, we computed ICCs between brain measures derived from the original scans versus one of the mask settings. This additional analysis was done in the child sample and the adult sample (first time point), because in older adults invasiveness was less of an issue likely due to an increased distance between the skull and the brain.

All statistics were computed in R (version 3.5.0, 2018-04-23). ICCs and their 95% confidence intervals were calculated with the "irr package" (version 0.84). based on a single measure, absolute-agreement, 2-way model. Local ICCs were visualized and average ICCs are reported over all regions or vertices. When averaging Pearson correlations or ICCs, the average was computed after Fisher's Z transformation of the individual values and then transformed back.

2.8 Visualization of reliability

For visualization of local reliability, ICCs were color-coded using colormap "jet" in MATLAB_R2017b. Next, region- or vertex-wise color-coded cortical ICCs were overlaid on the cortical surface of the average brain using FreeSurfer's tksurfer.

3. Results

3.1 Visual inspection

Figure 1 shows the effect of the different de-identification methods from an outer and sagittal view. The first column shows the original average scans for each age category. The 3D renders show facial and ear characteristics in great detail. The second column shows that defacing with FreeSurfer removes facial characteristics in a confined part of the face only, preserving ear characteristics and in some participants sensitive information in the area of the eyes or the mouth. Full preservation of the eyes occurred in up to 10% of the participants and full preservation of the mouth in up to 27% of the participants, but the inter-rater reliability for this assessment was low (Table 2). The third column shows that defacing with FSL removes most facial and ear characteristics, but some characteristics around the eyes remain. Full preservation of the eyes occurred often in children (18%) and only incidentally in young adults and older adults (Table 2). The fourth column shows that Face Masking results in blurring of the full face and ear area.

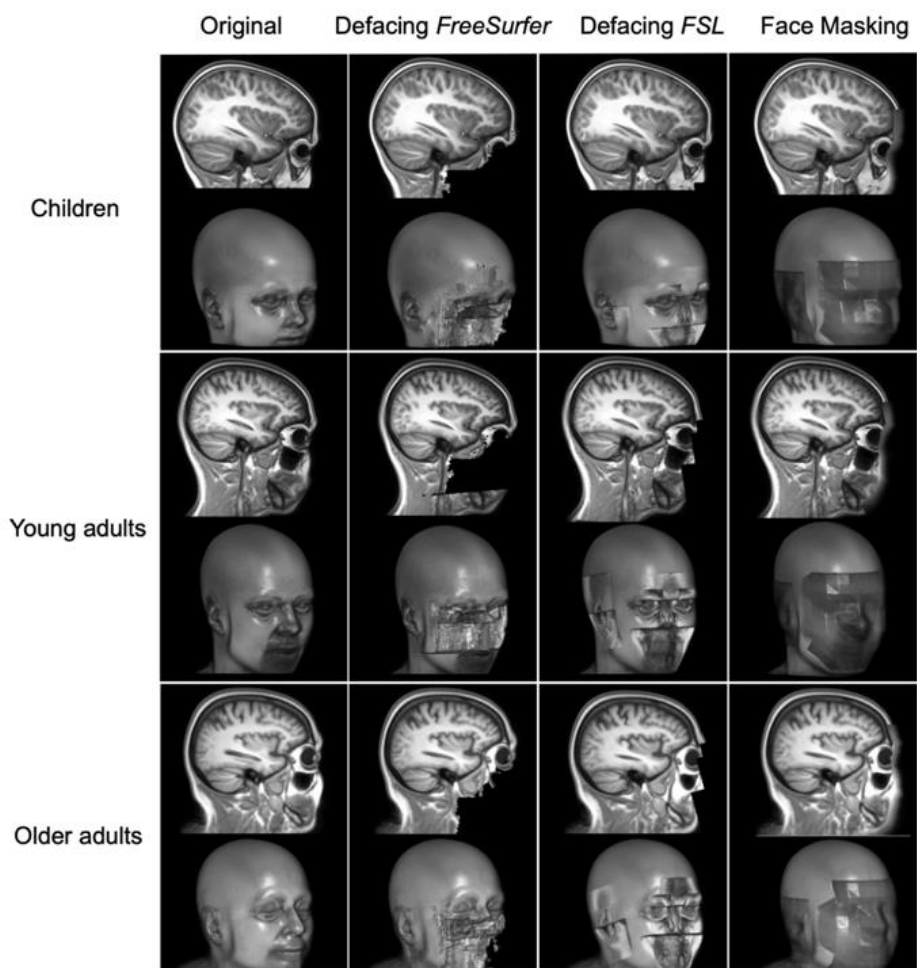


Figure 1. Visual appearance of scans before and after de-identification. The first column shows the original scan before de-identification and the other columns show the visual appearance of the scans after different de-identification options. Each row shows the de-identified average brain of a specific age group. Within each row a sagittal slice is shown on top and the 3D render below. To prevent identification, the face renderings shown here are renders of the average scans for each sample.

Table 2. Presence of eyes and mouth characteristics after de-identification. For each sample, the percentage of individuals is given for whom the eyes or mouth was preserved in the 3D render after de-identification. The percentages are the average of the percentages given by two raters. The inter-rater variability was 0.48 based on a one-way model of absolute agreement.

	Defacing FreeSurfer	Defacing FSL	Face Masking
Children – eyes	8%	18%	0%
Young adults – eyes	0%	6%	0%
Older adults – eyes	10%	5%	0%
Children – mouth	4%	4%	0%
Young adults – mouth	13%	0%	0%
Older adults – mouth	27%	0%	0%

Defacing with FreeSurfer was not invasive. With this method no brain tissue was removed across all scans. Defacing with FSL resulted in the removal of some brain tissue in the majority of children and in some young adults. Face Masking resulted in blurring of some brain tissue in all children and some young and older adults. Both FSL defacing and Face Masking were more invasive in younger participants compared to older participants. Furthermore, invasiveness was higher in proximity to the ear compared to the face region. Table 3 lists the percentages of scans in which brain tissue was affected by de-identification averaged over three ratings.

3.2 The effects of de-identification on brain measures

All original and de-identified scans were successfully processed using FreeSurfer. Brain measures were altered by de-identification procedures. However, the effect of de-identification on brain measures was small, i.e. absolute PDs were on average <5% for any age category or de-identification method. Means, standard deviations and corresponding CoVs were comparable before and after de-identification. Figure 2 shows the signed PDs for global brain measures. For most brain measures the signed PDs averaged out to around zero, but for cerebellar white matter volume and intracranial volume larger systematic biases were found depending on the age category and de-identification method. Systematic biases for average cortical

thickness, total cortical surface area and total cortical volume were very small, as suggested by the signed PDs (Figure 2) and scatterplots (Figure 3).

Table 3. Invasiveness of each de-identification method. For each sample, the percentage of individuals is given for whom the procedure was too invasive, i.e. brain tissue was blurred or removed due to de-identification of the face or ears. The percentages are the average of the percentages given by three raters. The inter-rater variability was 0.63 based on a one-way model of absolute agreement. * Currently, FreeSurfer does not provide an option to de-identify ears.

	Defacing FreeSurfer*	Defacing FSL	Face Masking
Children – face	0%	85%	100%
Young adults – face	0%	34%	10%
Older adults – face	0%	0%	6%
Children – ears	NA	64%	100%
Young adults – ears	NA	12%	56%
Older adults – ears	NA	0%	24%

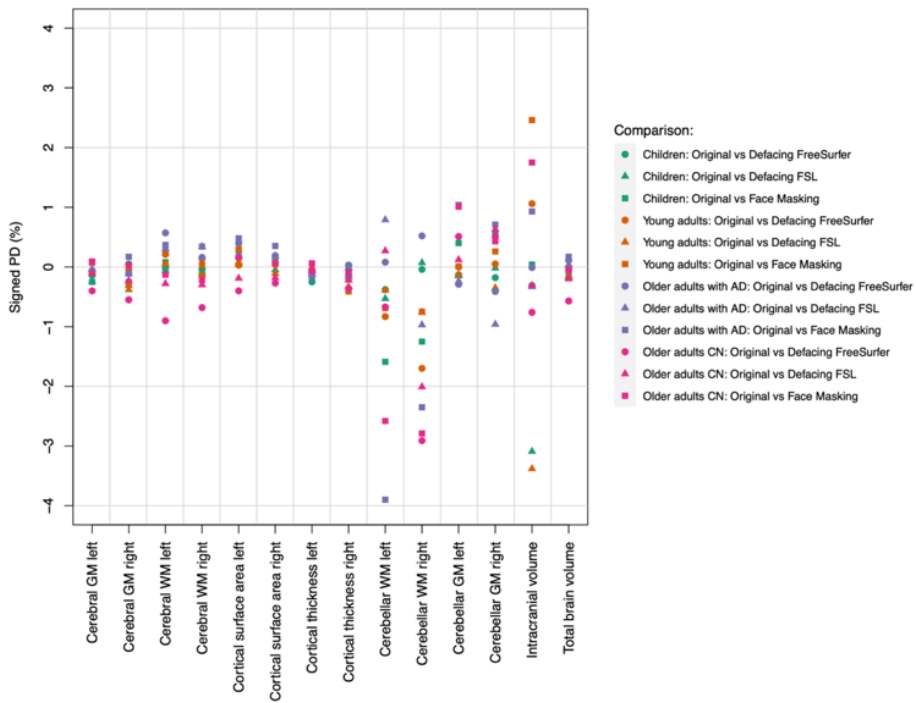


Figure 2. Average signed percentage differences for global brain measures after de-identification. WM=White matter; GM=Gray matter; AD=Alzheimer’s disease; CN=no cognitive impairment.

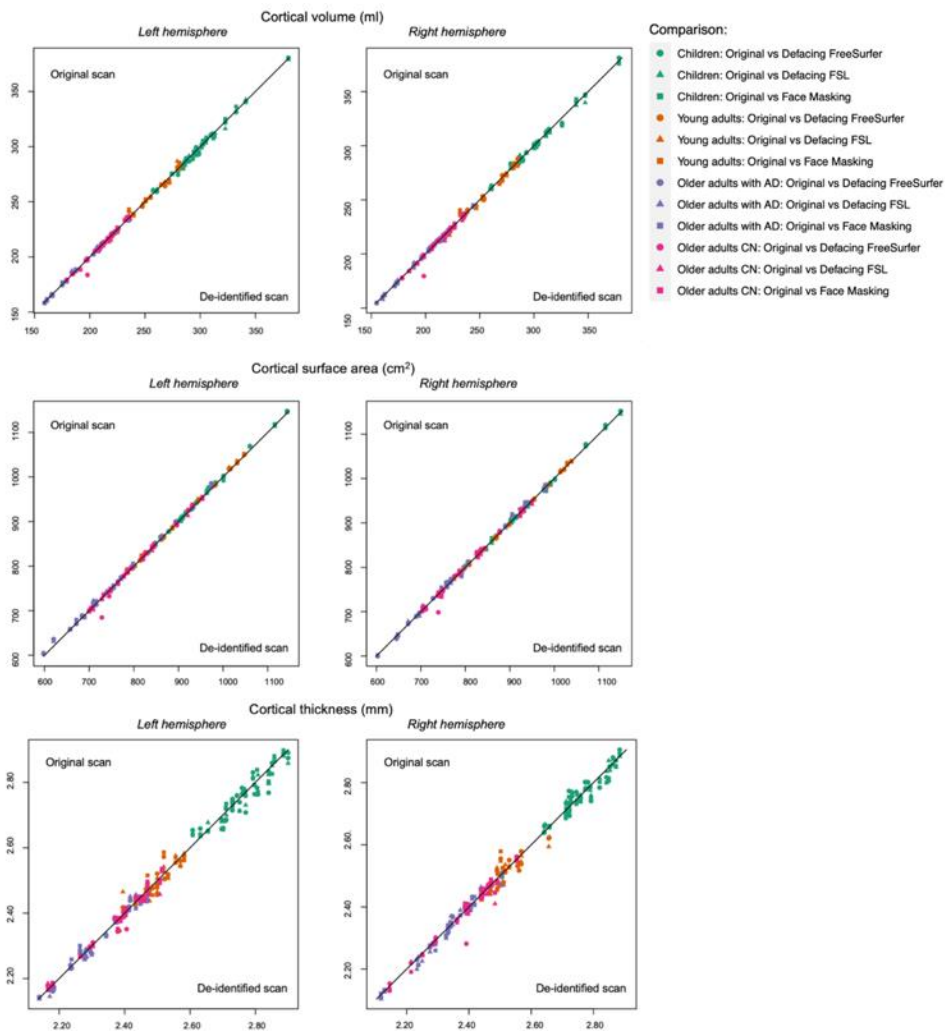


Figure 3. Individual global brain measures derived from original versus de-identified scans. AD=Alzheimer’s disease; CN=no cognitive impairment.

In children, all correlations were above 0.90 except for the ICC of intracranial volume with FSL defacing (Pearson's $r = 0.95$; ICC = 0.87). In young adults, all correlations for global brain measures were above 0.90 except right hemisphere cortical thickness for all de-identification techniques and left hemisphere cortical thickness with FSL defacing and Face Masking (Pearson's $r > 0.8$; ICCs > 0.8). In older adults, only the correlations for the cerebellar white matter were below 0.90 in the left hemisphere for all methods (Pearson's $r > 0.5$; ICCs > 0.5) and in the right hemisphere only for Face Masking (Pearson's $r > 0.8$; ICCs > 0.8). In addition to the lower correlations in the cerebellar white matter in the full group of older adults, lower correlations were found for right hemisphere cerebellar gray matter in older adults without Alzheimer's disease (Pearson's $r > 0.5$; ICCs > 0.5).

The correlation between regional brain measures derived from original scans and de-identified scans was on average higher than 0.90 using Pearson's r and ICCs. In general, similar correlations were found using Pearson's r or ICCs.

Regional ICCs for cortical volume were high in children, young adults and older adults (Figure S1). Average regional ICCs for cortical surface area in children, young adults and older adults were high for each method in each age category (Figure S2). Average regional ICCs for cortical thickness in children, young adults and older adults were on average also high (ICCs > 0.93) for each method in each age category (Figure 4, see Figure 5 for the vertex-wise ICCs). Despite these high ICCs, cortical volumes in children were significantly different in more regions than expected by chance after Face Masking or FSL defacing, with in general smaller volumes after de-identification. In older adults, cortical surface area was significantly different in more regions than expected by chance after Face Masking and in the AD subsample after both Face Masking and FreeSurfer defacing, with generally larger surface areas after de-identification.

All statistics comparing regional and global brain measures of original versus de-identified can be found in the Online Supplementary tables S1-S5.

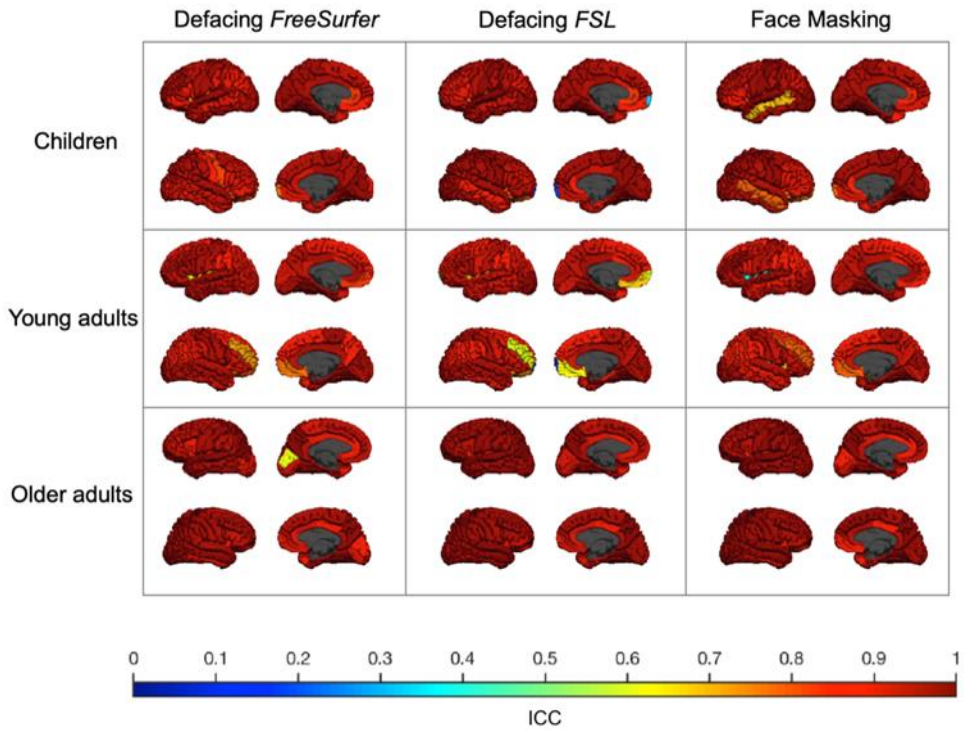


Figure 4. ICC of cortical thickness derived from original versus de-identified scans. The ICC for each sample (children, young adults and older adults) is plotted on the corresponding average scan. Each column shows a different de-identification technique. Within each square, the left hemisphere (top) and the right hemisphere (bottom) are shown from an outer and medial view. The lowest ICC (0.07) was found in the right frontal pole in young adults using FSL defacing.

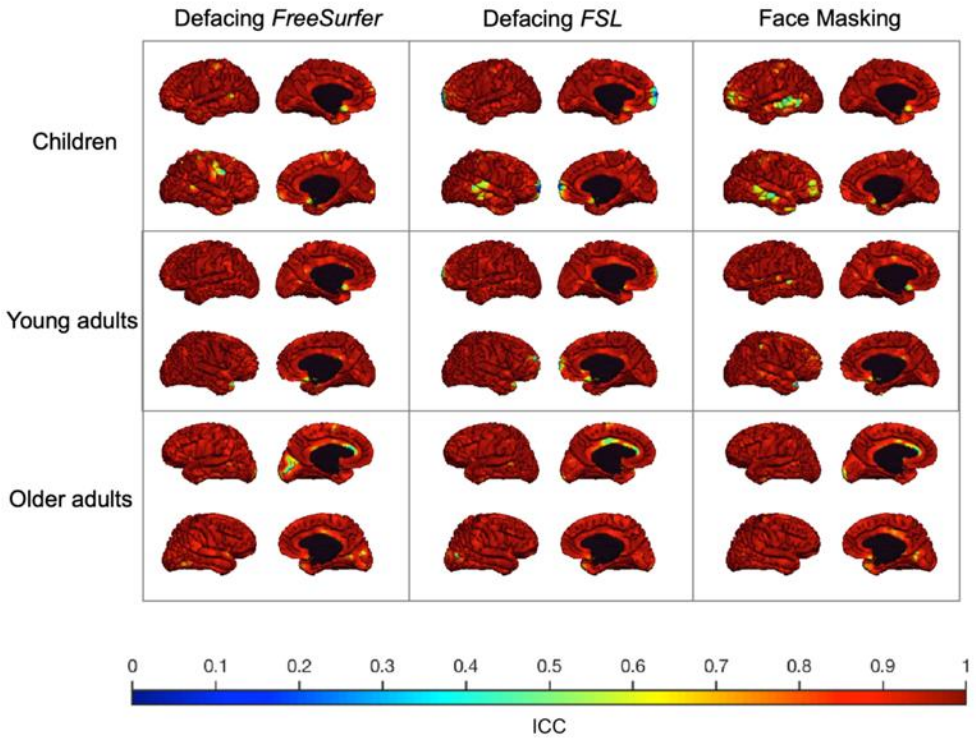


Figure 5. ICC of vertex-wise cortical thickness derived from original versus de-identified scans. The ICC for each sample (children, young adults and older adults) is plotted on the corresponding average scan. Each column shows a different de-identification technique. Within each square, the left hemisphere (top) and the right hemisphere (bottom) are shown from an outer and medial view.

3.3 Face Masking: Varying the coarseness of the mask in children and young adults

In order to better understand the effect of mask invasiveness on reliability of brain measures we varied the coarseness settings of the face and ear masks in children and young adults. In supplementary Figure S3 we define settings for a non-invasive mask based on visual inspection. The non-invasive mask had a coarseness setting (grid step value) of 0.6. At this value the brain tissue of all participants was untouched by both the face and the ear mask. Figure S4 and S5 show the ICCs for brain measures in cortical regions for the non-invasive and default mask (with and without ear mask) compared to the original scans in children and young adults respectively. We show that different coarseness settings generate similar reliability of brain measures. Furthermore, adding or removing the ear mask has minimal effect on reliability.

3.4 Effect of de-identification on test-retest reliability of brain measures

Independent of de-identification procedures, global brain measures were highly reliable, although reliability for cortical thickness was lower than for the other measures. The test-retest reliability for cortical thickness in the original scans was 0.89 for the left hemisphere and 0.74 for the right hemisphere. The other global brain measures had test-retest ICCs above 0.90. Test-retest reliability was similar for scans with or without de-identification for regional brain measures. Furthermore, de-identifying only the second scan did on average not result in different regional test-retest ICCs. Figure 6 shows the regional test-retest ICCs for cortical brain measures with or without de-identification. Figure 7 shows similar cortical test-retest ICCs when de-identifying only the second scan.

All test-retest statistics can be found in the Supplementary tables S6-S8.

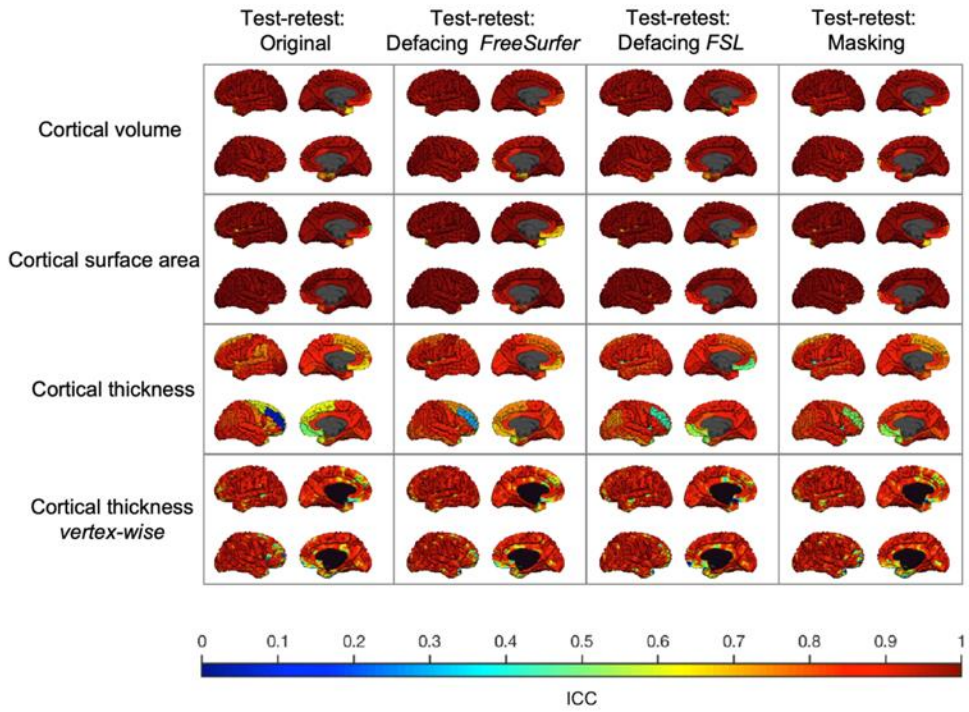


Figure 6. Test-retest reliability in young adults using different de-identification techniques. The test-retest ICC for each type of brain measure is plotted on the average adult scan. The first column shows the test-retest reliability of the original scans. The other columns show the test-retest reliability if both scans are subject to a de-identification technique. Within each square, the left hemisphere (top) and the right hemisphere (bottom) are shown from an outer and medial view. Lowest reliability was found for cortical thickness in the rostral middle frontal gyrus of the right hemisphere independent of de-identification procedures. The poor reliability in this region was not related to parcellation or segmentation errors and could not be explained by a single outlier.

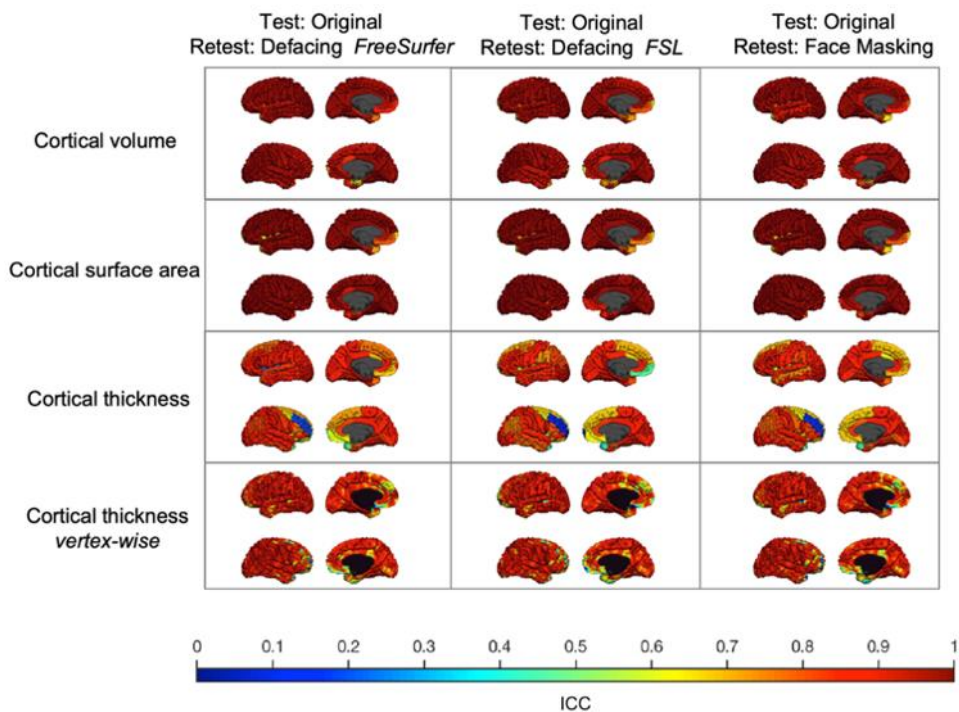


Figure 7. Test-retest reliability in young adults de-identifying only the second scan. The test-retest ICC for each type of brain measure is plotted on the average adult scan. Each column shows the test-retest reliability of the original scan compared to a de-identified second scan using different techniques. Within each square, the left hemisphere (top) and the right hemisphere (bottom) are shown from an outer and medial view.

4. Discussion

In this study we evaluated three de-identification methods for structural MRI scans: Defacing in FreeSurfer, defacing in FSL and the Face Masking toolbox. Our main goal was to assess the effect of these methods on the reliability of global and regional brain measures. In addition, we aimed to determine the utility of these methods in cohort studies investigating development or ageing. To our knowledge this is the first study into the effects of de-identification procedures that includes neuroimaging data from children.

We show that using Face Masking and FSL defacing, voxels in the brain are overwritten, especially in children and especially in proximity to the ears. Face Masking provides the option to decrease the coarseness of the mask. This prevents the blurring of brain tissue and results in similar reliability measures. However, reducing coarseness may come at the expense of the full covering of facial features. The age-dependent effect of these de-identification methods could be related to the distance of the skull to the brain, which increases with ageing due to atrophy. FreeSurfer defacing did not remove voxels in the brain. Furthermore, FreeSurfer defacing and FSL defacing did not succeed in fully removing facial characteristics in all participants.

The obvious visual differences between the methods do not translate to differences in the reliability of brain measure estimates. In general, high correlations were found between brain measures derived from original scans versus de-identified scans but in some regions lower correlations were found independent of the used method. Pearson's correlations were in most cases similar to ICCs. We found some evidence for small systematic biases and significantly different brain measures after application of de-identification methods. These biases were not equally distributed over the age groups, disease-status groups for the ADNI data, and de-identification methods. This suggests age-specific biases. Still, these effects were very small.

The ICCs found in this study are comparable to test-retest ICCs of the original scans in young adults, for a direct comparison of ICCs see Figure 8. Theoretically, the amount of noise added by defacing or Face Masking should be far less than the

amount of noise introduced in the test-retest procedure. Given that brain deviations found in imaging studies are generally small, introducing an error that is similar in size to test-retest differences is undesirable. However, we argue that the amount of added noise is limited based on the finding that ICCs between original first session scans and de-identified second session scans were comparable to test-retest ICCs, suggesting no additional effect of de-identification on reliability. In a previous study we modeled how these test-retest ICCs relate to power and effect of interest (Buimer et al., 2020). Based on the ICCs found in the current study, it can be hypothesized that as long as all scans are uniformly processed, the sample size needed to detect an effect of interest is the same for original and de-identified scans. However, sample sizes may be decreased after the necessary visual quality control which could lead to more exclusions than usual. Scans might need to be removed when privacy-sensitive features are not fully de-identified or when brain tissue is impacted.

Regional test-retest ICCs of unmasked scans had an average of 0.95 for subcortical volume, 0.96 for cortical volume, 0.98 for cortical surface area and 0.84 for cortical thickness. Lower reliability for cortical thickness measures compared to cortical volume or cortical surface area has been reported before (Iscan et al., 2015; Liem et al., 2015; Wonderlick et al., 2009). Brain measures with a high between-subject variability naturally generate higher ICCs. Surface area and subcortical volume have a higher between-subject variability than cortical thickness. This may explain why cortical thickness test-retest ICCs are lower on average, but it does not explain the region-specific lower ICCs when comparing atlas-based results with vertex-wise results. In our study, the lower average ICC for atlas-based cortical thickness was mostly driven by poor reliability in specific frontal regions in the right hemisphere. Vertex-wise cortical thickness in these regions was more reliable. Furthermore, lower test-retest ICCs were found in areas that are known to be unreliable, such as the frontal pole (Desikan et al., 2006).

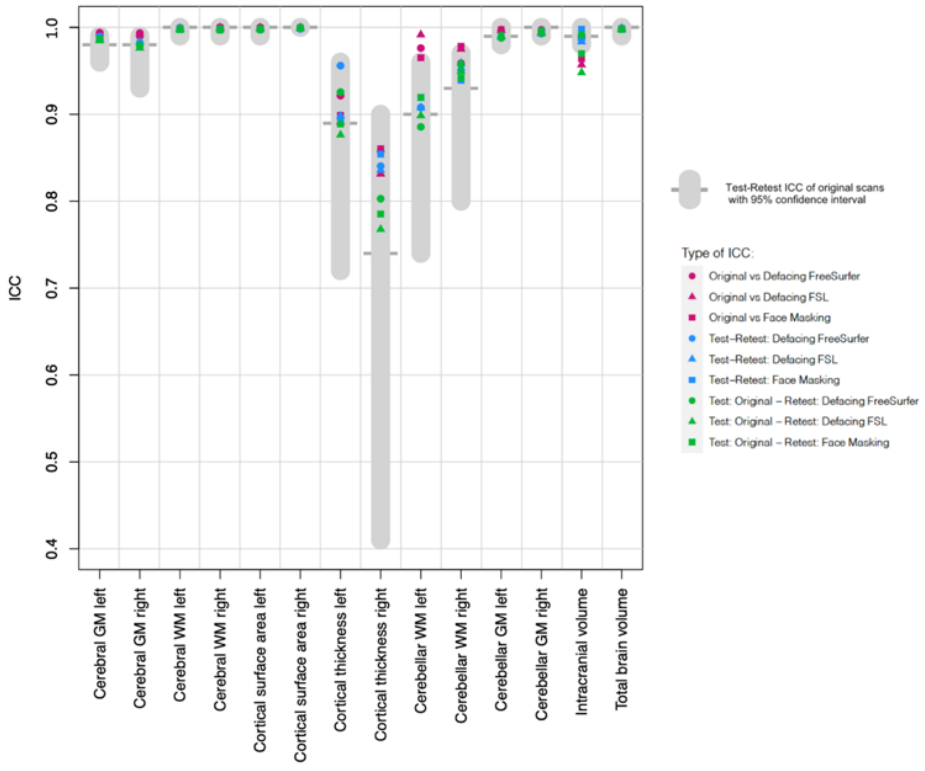


Figure 8. Test-retest ICCs of global brain measures compared to de-identification ICCs. This figure allows for direct comparison of the ICCs for global brain measures in young adults reported in our study. In grey, test-retest ICC and 95% confidence interval of the original scans, i.e. without any de-identification. In purple, ICCs comparing original adult scans to de-identified adult scans. In blue, test-retest ICCs when both scans are de-identified. In green, test-retest ICCs when only the second scan is de-identified. The type of de-identification applied is indicated by the shape of the data point. GM=gray matter; WM=white matter.

Our study complements previous work from de Sitter et al., (2020) and Schwarz et al., (2021). De Sitter et al., (2020) compared three de-identification methods: QuickShear (Schimke & Hale, 2011), Face Masking (Milchenko & Marcus, 2013) and FreeSurfer defacing (Bischoff-Grethe et al., 2007). Brain measures of interest and corresponding processing pipelines in this study were tailored to the patient groups (e.g. BraTumIA for segmentation of glioblastoma) and are therefore hard to compare to our populations, except for an analysis using SIENAX on the ADNI data. Schwarz et al., (2021) compared four different de-identification methods: FreeSurfer defacing (Bischoff-Grethe et al., 2007), FSL defacing (Alfaro-Almagro et al., 2018), pydeface (Gulban et al., 2022) and the newly developed method `mri_reface` (Schwarz et al., 2021). We add to this literature by including neuroimaging data from children to directly compare the effects of defacing on brain images of individuals at different ages and by adding a test-retest dataset of subjects scanned one week apart. Both earlier studies show that all de-identification methods impacted subsequent image processing and highlighted the possible role of altered image registration (de Sitter et al., 2020; Schwarz et al., 2021). Accordingly, we noted that de-identification has a small effect on the Talairach transformation at the start of the FreeSurfer segmentation pipeline, which resulted in altered registration for all de-identification methods used. Small effects on registration can have large effects on brain measures in areas that are difficult to parcellate. In addition, scan-rescan variability can affect registration, resulting in different segmentations and parcellations. Here, we show that in general, the effects of repeated scanning (test-retest) were higher than the effects of de-identification on brain measures. Similar results were found for within-scanning-session test-retest reliability in ADNI data (Schwarz et al., 2021). De Sitter et al., (2020) and Schwarz et al., (2021) both describe a systematic bias towards lower brain volumes. We also found some systematic biases, but these biases were not consistent across all age groups, all methods and all regions. The small systematic effects did not translate in lower ICCs, which suggests that most study results will not be impacted by such a bias because between-subject variation will be similar if all scans are processed using the same de-identification method. These biases may be more important when studying longitudinal trajectories. Whether these biases would influence results and whether within-person effects of de-identification are stable over time remains an open question at this point.

Neuroimaging results are impacted by methodological choices such as scanner make and model, type of acquisition, study sample, processing pipeline and atlas (Vijayakumar et al., 2018). In this study we only used high quality data obtained from T1-weighted scans acquired on a 3T Philips scanner, and all scans were processed with FreeSurfer. Therefore, the reported effects of de-identification could be study-specific. Other scanners or different quality data may generate different results. With regard to pipeline-specific effects, previous studies show the effect of de-identifying ADNI data on brain measures using other processing pipelines (de Sitter et al., 2020; Schwarz et al., 2021). Furthermore, based on this study we cannot draw conclusions about brain measures derived with other software or other types of image acquisitions such as T2-weighted scans. Face Masking can be applied on CT scans and T1- and T2-weighted MRI scans. Defacing with FSL can be applied on T1- and T2-weighted MRI scans. Defacing with FreeSurfer is based on a T1-weighted face mask, but in principle it would be possible to use this method on T2-weighted MRI scans as well. Another limitation of the current study is that only adult test-retest data was included. No ethical approval was granted to collect test-retest data in the YOUth cohort. For the ADNI cohort test-retest data (without repositioning the participant in the scanner) is available. These test-retest data were compared to the effects of de-identification before (Schwarz et al., 2021). Lastly, an important limitation is that we were not able to extensively study whether the de-identification methods were successful, because we did not have access to photographs of the participants. Automated facial recognition methods are probably the best tool to test if de-identification was successful. A recent study using automated facial recognition showed that defacing data with FreeSurfer reduces the probability of identifying a participant from 97% to 10%. Using FSL defacing the probability was reduced even further to 3% (Schwarz et al., 2021). Face Masking was not included in this study because preliminary evidence suggests that this de-identification technique could be reversible (Abramian & Eklund, 2018).

In conclusion, de-identification methods impact recognizable facial characteristics, but the side effect is that brain measures are impacted as well. We show that if de-identification is a necessity, masking or defacing can be considered, as global brain measures can be estimated reliably and in general local brain measures are

minimally affected. We also observe that these methods do not de-identify all participants beyond recognition which may lead to exclusions of scans. Thus, the perfect de-identification method i.e. one that does not impact brain measures and does not result in additional exclusion of scans, does not exist yet. This paper highlights the importance to further develop de-identification methods, especially for neuroimaging data from children.

5. Supplementary materials

Some of the supplementary files were too large to be incorporated here and can be downloaded online via this link <http://doi.org/10.17605/OSF.IO/M5R3U>. For Chapter 3, the online supplement includes one file with the supplementary tables listing all the results.

3

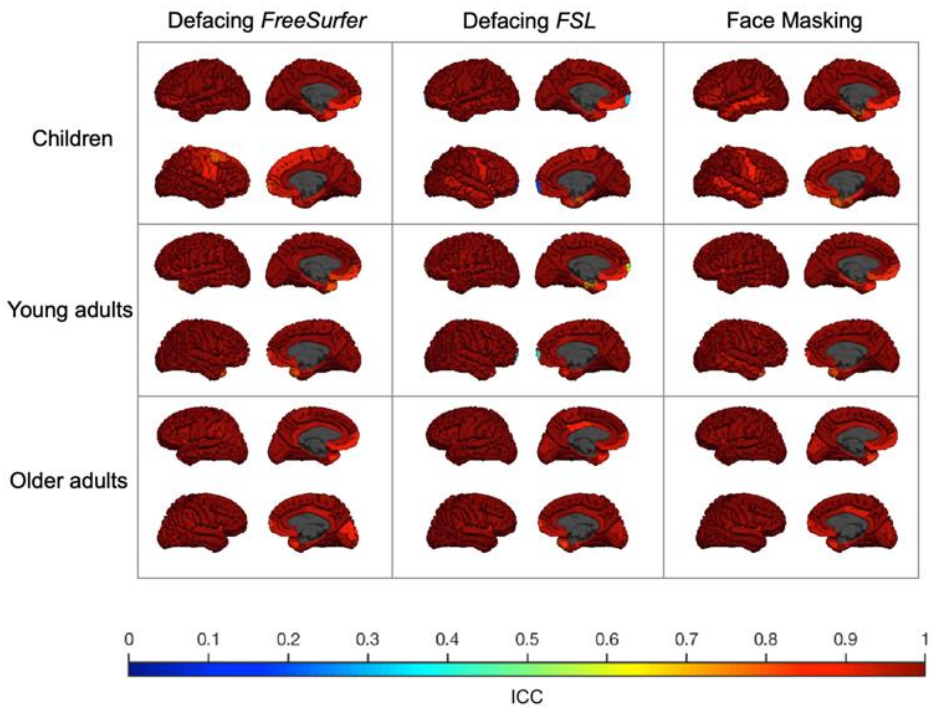


Figure S1. ICC of cortical volume derived from original versus de-identified scans. The ICC for each sample (children, young adults and older adults) is plotted on the corresponding average scan. Each column shows a different de-identification technique. Within each square, the left hemisphere (top) and the right hemisphere (bottom) are shown from an outer and medial view. The lowest ICC (0.21) was found in the right frontal pole in children using FSL defacing.

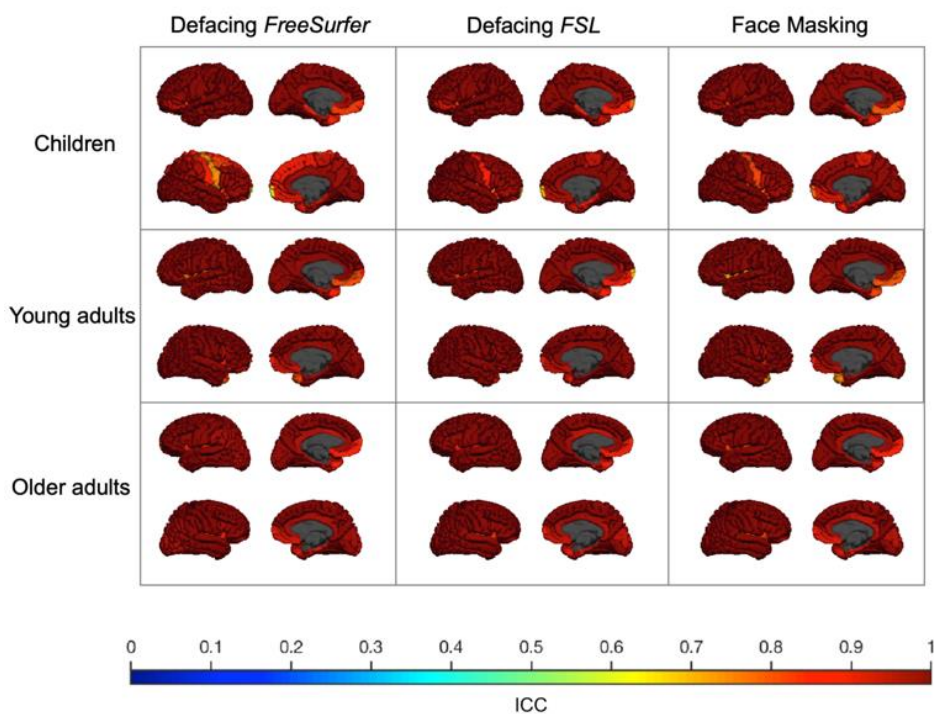


Figure S2. ICC of cortical surface area derived from original versus de-identified scans. The ICC for each sample (children, young adults and older adults) is plotted on the corresponding average scan. Each column shows a different de-identification technique. Within each square, the left hemisphere (top) and the right hemisphere (bottom) are shown from an outer and medial view. The lowest ICC (0.61) was found in the right frontal pole in children using FreeSurfer defacing.

Chapter 4 - Adverse childhood experiences and fronto-subcortical brain structures in YOUth.

Elizabeth E.L. Buimer, Rachel M. Brouwer, René C.W. Mandl, Pascal Pas, Hugo G. Schnack, Hilleke E. Hulshoff Pol

Published in 2022

Frontiers in Psychiatry, 13, 95587145.

<https://doi.org/10.3389/fpsy.2022.955871>

Abstract

The impact of adverse childhood experiences (ACEs) differs between individuals and depends on the type and timing of the ACE. The aim of this study was to assess the relation between various recently occurred ACEs and morphology in the developing brain of children between 8-11 years of age. We measured subcortical volumes, cortical thickness, cortical surface area and fractional anisotropy in regions of interest in brain scans acquired in 1184 children from the YOUth cohort. ACEs were based on parent-reports of recent experiences and included: financial problems; parental mental health problems; physical health problems in the family; substance abuse in the family; trouble with police, justice or child protective services; change in household composition; change in housing; bereavement; divorce or conflict in the family; exposure to violence in the family and bullying victimization. We ran separate linear models for each ACE and each brain measure. Results were adjusted for the false discovery rate across regions of interest. ACEs were reported for 83% of children in the past year. Children were on average exposed to two ACEs. Substance abuse in the household was associated with larger cortical surface area in the left superior frontal gyrus, $t(781) = 3.724$, $p_{FDR} = .0077$, right superior frontal gyrus, $t(781) = 3.409$, $p_{FDR} = .0110$, left pars triangularis, $t(781) = 3.614$, $p_{FDR} = .0077$, left rostral middle frontal gyrus, $t(781) = 3.163$, $p_{FDR} = .0195$ and right caudal anterior cingulate gyrus, $t(781) = 2.918$, $p_{FDR} = .0348$. Household exposure to violence (was associated with lower fractional anisotropy in the left and right cingulum bundle hippocampus region $t(697) = -3.154$, $p_{FDR} = .0101$ and $t(697) = -3.401$, $p_{FDR} = .0085$, respectively. Lower household incomes were more prevalent when parents reported exposure to violence and the mean parental education in years was lower when parents reported substance abuse in the family. No other significant associations with brain structures were found. Longer intervals between adversity and brain measurements and longitudinal measurements may reveal whether more evidence for the impact of adverse childhood experiences on brain development will emerge later in life.

1. Introduction

A history of adverse childhood experiences (ACEs), such as maltreatment, parental divorce, exposure to violence or substance abuse in the family, is a risk factor for developing mental health problems later in life (Green et al., 2010; Kessler et al., 2010; McLaughlin, 2016). ACEs are associated with decreased life expectancy, for example via the effects of ACEs on toxic stress, increased adult health risk behavior, increased suicidality or socioeconomic inequality (Felitti et al., 1998; Hughes et al., 2017; Kalmakis & Chandler, 2015; Merrick et al., 2019). These studies also show that especially individuals exposed to cumulative adversity are at risk for mental or physical health problems later in life.

Individual differences explain how a child is impacted by ACEs. Many individuals exposed to adversity are resilient to negative effects. For example, high psychosocial functioning despite a history of childhood maltreatment can be explained by neurobiological and genetic factors, but the social environment can serve as a protective factor as well (Ioannidis et al., 2020). Even more, volume alterations found in older adults exposed to ACEs vary based on their serotonergic genetic vulnerability (Ancelin et al., 2021).

The impact of ACEs also depends on the timing of the ACE in relation to sensitive periods of brain development (Andersen et al., 2008; Gee & Casey, 2015; Heim & Binder, 2012; Kuhn et al., 2016; Tottenham & Sheridan, 2009). During childhood and adolescence, the brain undergoes considerable developmental changes, including a thinning of the cortex, an increase followed by a decrease in cortical surface area and continued growth of volume of the white matter connections (Frangou et al., 2022; Giedd et al., 1999; Tamnes et al., 2017); and these changes have been related to cognitive functioning (Schnack et al., 2015). Many psychiatric disorders emerge during adolescence (Paus et al., 2008). Studying how ACEs interact with brain development is crucial to better understand mechanisms of latent vulnerability (McCrorry & Viding, 2015) and resilience (Kalisch et al., 2017).

Neuroimaging studies on childhood adversity started with a strong focus on the effects of severe early caregiver adversity, for example in institutionalized children

or children exposed to childhood maltreatment. Furthermore, most studies focused on the fronto-limbic network (frontal cortex, hippocampus and amygdala) because of the well-established role of fronto-limbic regions in the hypothalamic-pituitary-adrenal (HPA) axis functioning in response to stress (Dahmen et al., 2018), although a meta-analysis concluded that there is no evidence for abnormalities in the amygdala and only weak evidence for smaller hippocampal volumes in adults that experienced childhood adversity (Calem et al., 2017). There are only a few whole-brain studies on the association between childhood maltreatment and brain structure. Taking meta-analyses and reviews together, the most consistent findings are in fronto-limbic regions, fronto-striatal regions, fronto-subcortical association fibers and the corpus callosum (Calem et al., 2017; Daniels et al., 2013; Hart & Rubia, 2012; Lim et al., 2020; McCrory et al., 2010; McLaughlin et al., 2019; Paquola et al., 2016). Still, the spatial overlap between studies is weak, because most studies rely on smaller samples or adult samples, as children with experiences of maltreatment are difficult to include in large numbers. Furthermore, there is evidence for differential structural brain correlates depending on the types of ACE (Cassiers et al., 2018). Therefore, studying a variety of ACEs in developmental populations may shed light on the effects of ACEs on brain development.

In the current study, we investigate the effect of ACEs on brain structure in pre-adolescent children, participating in the first wave of the YOUth cohort study, a longitudinal study where each measurement wave covers a narrow period of development. The main question of the current study is: Is there an association between ACEs and brain structure in pre-adolescent children? Regions of interest were selected a priori by integrating studies on structural, functional and neurocognitive correlates of childhood adversity. We focused on subcortical volume, cortical surface area, cortical thickness and fractional anisotropy (FA). The latter was selected as white matter measure because it represents a good measure for integrity of the white matter and has shown good test-retest reliability using our acquisition protocol (Buimer et al., 2020). We expect that our sample size allows for detection of more subtle effects even though the sample is not enriched for children with severe adverse experiences. Within the group of children that experienced adversity, we expect more pronounced effects in children that were exposed to accumulated ACEs,

compared to children exposed to a single ACE. Based on the stress acceleration hypothesis (Callaghan & Tottenham, 2016), we hypothesized that brain development in children exposed to childhood adversity would be ahead of peers. Based on brain development curves in previous studies, we expect that in 8-, 9- and 10-year-olds accelerated development would mean thinner cortices (Teeuw et al., 2019), larger subcortical volumes (Mills et al., 2021), large cortical surface area (LeWinn et al., 2017) and larger FA (Koenis et al., 2015). Age- and sex-effects on global brain measures are included as well, to provide a full description of the YOUth cohort sample for comparison with other cohorts.

2. Materials and methods

2.1 Participants

We included 1184 children that participated in the first wave of the population-based longitudinal YOUth cohort study. The cohort rationale, design and procedures are described in detail elsewhere (Onland-Moret et al., 2020). In short, participants are living in the province of Utrecht (the Netherlands) and its surrounding areas, a densely populated region that combines both urban and rural areas. Compared to the rest of the Netherlands, inhabitants of the province of Utrecht are relatively highly educated. Most children were recruited through their primary school. YOUth includes children and their parents. Parents are considered those with parental authority over the child. Children are excluded if they are not mentally or physically not capable of participating, if they or their parents' language proficiency in Dutch is not sufficient to understand provided information. For the neuroimaging part of the study, we excluded children with metal implants including most braces, following fairly standard MRI procedures. Participating children were between 7.9 and 11.0 years old (56% females). All data was collected prior to the COVID-19 pandemic. The study was approved by the Medical Research Ethics Committee Utrecht. Children participated on a voluntary basis and parents or guardians gave written consent and assent.

2.2 Data on ACEs

ACEs were collected using parent reports on life events that occurred in the household in the past year, available for 1046 children. From the recent life events questionnaire, we selected 9 types of ACEs: financial problems; physical health problems in the family; substance abuse in the family; trouble with police, justice or child protective services (CPS); change in household composition; change in housing; bereavement; divorce or conflict in the family; exposure to violence in the family. In addition, two ACEs were gathered from other questionnaires as they were not covered in the recent life events survey. First, information on bullying behavior towards the child was available for 948 children. From the bullying questionnaire one ACE was created by selecting whether children were exposed to any type of frequent bullying at least one time a week. Second, information on parental psychiatric diagnoses was available for 1056 children. Parental mental health problems were indicated as an ACE if one or more parents or guardians were diagnosed with at least one psychiatric diagnosis.

All ACEs were used as binary variables (yes/no). In most cases, information from different questionnaire items was combined into a single composite variable. For example, change in household composition would be set to yes if at least one new member was added to the household, for example cohabitation of a new partner, cohabitation of a step brother or sister, birth of a new family member etcetera. As these separate items are conceptually very similar, we regard this as a single event rather than multiple independent events. Overlap between the separate items of the recent life events survey was mapped (Supplementary Figure S1), but not decisive when creating the composite variables, because high overlap between items does not necessarily implicate a single underlying environmental factor. Overlap was mapped by taking the subgroup of children that experienced a specific event and then computing the percentage of the children in this subgroup that additionally experienced another event. In the same way, we mapped the overlap between the final 11 ACEs used in this study (Figure 2C).

The prevalence of ACEs was similar in the total group, compared to the subgroups that had MR data available. Because we used various data sources for the ACEs, sample sizes differed between bullying, recent life events and parental psychiatric diagnosis dependent on the overlap of respondents with available MR data.

2.3 Image acquisition

The collection of MRI data is closely monitored in the YOUth cohort study. Patterns in data quality are monitored over time based on human data and weekly collected phantom data. The YOUth MRI protocol, quality control and test-retest reliability are described in detail elsewhere (Buimer et al., 2020). In short, anatomical T1-weighted MRI scans were available for 956 children and diffusion-weighted images (DWI) for 895 children. All MR scans were acquired on the same scanner, a Philips Ingenia 3.0 T CX scanner with a 60 cm bore (Philips Medical Systems, Best, The Netherlands) using a 32-channel SENSE head-coil. A structural T1-weighted 3D gradient echo scan was acquired with the following parameters: TR = 10 ms; TE = 4.6 ms; flip angle = 8°; reconstructed voxel size = 0.75 x 0.75 x 0.80 mm³; parallel imaging factor = 1.70 (AP) and 1.40 (RL). Next, a diffusion-weighted multi-shell multi-band echo planar (EPI) acquisition is obtained including two short DWI scans with a reversed phase encoding readout to correct for susceptibility artefacts. The following parameters were used to acquire the DWI scan: TR = 3500 ms; TE = 99 ms; flip angle = 90°; reconstructed voxel size = 2.0 x 2.0 x 2.0 mm³; multiband acceleration factor = 3; parallel imaging factor = 1.3; b-values = 500 [15], 1000 [30], 2000 [60] and every 10th scan is a diffusion unweighted (b-value = 0) scan. At the start of the study a different DWI protocol was used. Therefore, 13% of the included data was acquired without reversed phase encoding readout and with the following parameters: TR = 6827 ms; TE = 101 ms; flip angle = 90°; reconstructed voxel size = 2.5 x 2.5 x 2.5 mm³; parallel imaging factor = 2.5; b-values = 1000 [15], 2000 [25], 3000 [35] and every 10th scan is a diffusion unweighted scan. We corrected for the different protocols in our analyses (see below for details).

2.4 Image processing

At the time YOUth provided access to the data for this study, defacing or face masking procedures had not yet been implemented. Therefore, T1-weighted scans were not subjected to any defacing or face masking procedures that may have a small effect on outcome measures (Buimer et al., 2021).

FreeSurfer 6.0 was used for automatic brain segmentation and parcellation of the T1-weighted scans (Fischl et al., 2002). Subcortical data was extracted from the output of FreeSurfer's volume-based stream. For the initial registration and nonlinear alignment steps we relied on the default MNI305 atlas. Cortical data was extracted from the output of FreeSurfer's cortical surface-based atlas. For registration of the individual surfaces to an average sphere we used the default fsaverage. The Desikan-Killiany atlas was used for cortical parcellation (Desikan et al., 2006). To be able to study whether our findings were specific to the ROIs rather than a global effect, we also extracted intracranial volume and computed total cortical surface and average cortical thickness.

DWI scans were processed using FSL version 6.0.1 (Jenkinson et al., 2012) in combination with MRtrix 3.0 (Tournier et al., 2019). The processing pipeline consisted of denoising (Cordero-Grande et al., 2019; Veraart et al., 2016), gradient direction corrections (Leemans & Jones, 2009), eddy current corrections (Andersson & Sotiropoulos, 2016), susceptibility corrections (Andersson et al., 2003) and corrections for Gibbs-ringing artefacts (Kellner et al., 2016). FSL's EDDY QC framework was used to get quality control reports for each individual (QUAD) and at the group-level (SQUAD) (Andersson et al., 2003, 2016; Bastiani et al., 2019; Smith et al., 2004). QC parameters in the QUAD output include estimates of absolute motion, relative motion, translations, rotations, eddy current linear terms, susceptibility, B-value outliers, signal-to-noise ratio and contrast-to-noise ratio. Next, FSL's Tract-Based Spatial Statistics (TBSS) was used to skeletonize the fractional anisotropy (FA) maps in standard space (Smith et al., 2004, 2006). For the TBSS registration and transformation to MNI152 space we used the default adult template, FMRIB58_FA. Using the TBSS processing pipeline we also generated a webpage

with, for each individual, slices of the FA maps for visual quality control. Lastly, the intersection between the skeleton and the regions of the JHU-ICBM-DTI-81 atlas (Mori et al., 2008) was used to compute the average FA values for these regions. Furthermore, the mean FA over all atlas regions was computed.

2.5 Quality control

For the T1-weighted scans, an experienced rater visually assessed image quality for each individual based on the original scan and segmentation quality based on the pial reconstruction. From the 956 T1-weighted scans a total of 132 scans were excluded for various reasons: motion artifacts that affected pial surface reconstruction (N=96), inhomogeneity artifacts (N=17), brain anomalies (N=9), FreeSurfer failed (N=5), corrupt DICOM files (N=3), dental artifact (N=1) and incorrect field-of-view (N=1). This resulted in gray matter estimates for 824 children.

For the DWI scans, we started off with 895 scans. We used the presence of artefacts on the T1-weighted scans as a predictor for the quality of the DWI scans, thereby excluding the same children that were excluded in the T1-weighted analysis (N= 123). Furthermore, DWI data was incomplete or missing in some cases (N=10) or failed the processing pipeline (N=17). Next, based on the visual inspection of the FA map snapshots for each individual, we additionally excluded DWI scans that were acquired in a different orientation (N=7) or with an incorrect field-of-view (N=2). The distribution of the SQUAD QC parameters was as expected. Outliers were visually checked once again, but did not lead to exclusions. We performed t-tests for each QC parameter using the individual-based QUAD output, to test if image quality differed between children without any ACE versus children with at least one ACE. The prevalence of differences between the QC parameters in children with versus without an ACE were as would be expected by chance, with no consistent patterns of lower quality for specific ACE subsets. Therefore, no exclusions were made based on these parameters. The results of FSL's TBSS were visually checked as recommended in the user guide (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS/UserGuide>). This QC did not result in additional exclusions.

2.6 Regions of interest

Regions of interest were selected from the gray matter and white matter atlases (Figure 1) based on meta-analyses and reviews described earlier (Calem et al., 2017; Daniels et al., 2013; Hart & Rubia, 2012; Lim et al., 2020; McCrory et al., 2010; McLaughlin et al., 2019; Paquola et al., 2016).

2.7 Statistical analyses

Using R version 4.0.5 (2021-03-31) in R studio version 1.4.1106 each brain measure was regressed on each ACE in a separate linear model. All variables were scaled and centered to create standardized output that is comparable across cohorts and across different brain measures. Apart from each specific ACE, we also tested the relation between brain measures and any ACE and accumulated ACEs (sum of ACEs per child). Any ACE and accumulated ACE variables were only computed if data was available for all ACEs. All analyses were corrected for age and sex. For the DWI analyses we also included a dichotomous variable to correct for the DWI acquisition protocol. Additionally, to assess regional specificity, we repeated the main analysis for the T1-weighted brain measures correcting subcortical volumes for intracranial volume, regional cortical thickness for average cortical thickness and regional cortical surface area for total cortical surface area.

We corrected for multiple comparisons by controlling the false discovery rate (FDR) (Benjamini & Hochberg, 1995). Throughout this manuscript we will report uncorrected p -values (p_{uncorr}) and FDR-adjusted p -values (p_{FDR}). FDR-adjusted p -values were adjusted across T1-weighted or DWI brain measures independently for each ACE separately and thus not across all analyses.

To estimate the robustness of the regression coefficients, we applied non-parametric bootstrapping, drawing random samples with replacement from the residuals of the regression model and added these to the original fitted values to create 5000 new samples. We chose resampling residuals over resampling subjects because of the small groups for some ACEs.

2.8 Data visualization

For (sub)cortical surface data visualization, we used the ENIGMA toolbox (Larivière et al., 2021). For white matter tract visualization, we used Surfice (www.nitrc.org/projects/surface/). Subcortical and white matter structures were overlaid on a reference brain to indicate the orientation of the structures and the approximate location in the brain.

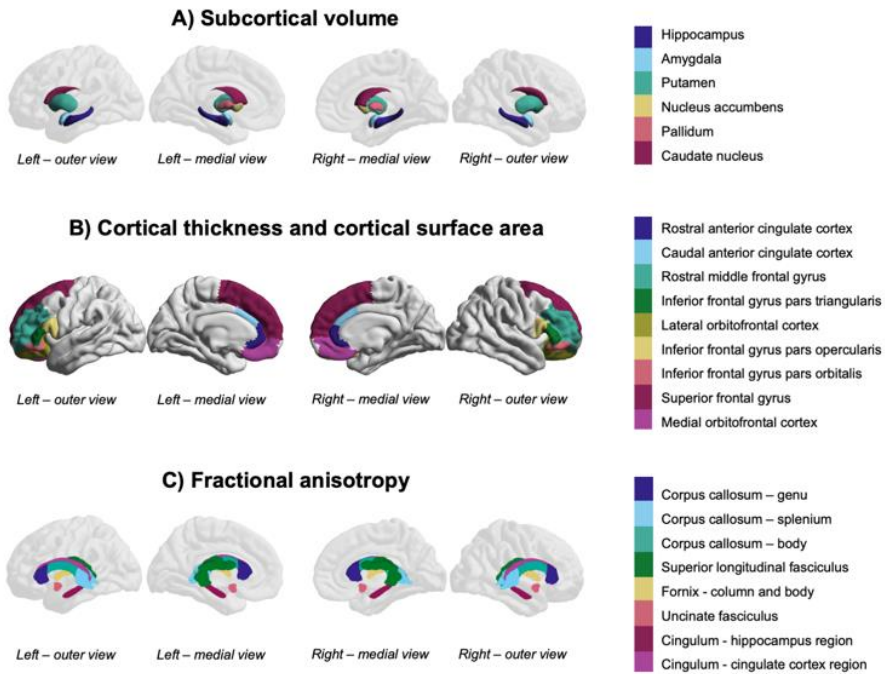


Figure 1. Fronto-limbic and fronto-striatal regions-of-interest.

3. Results

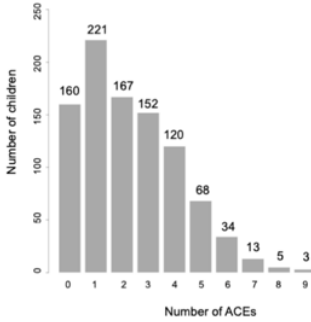
3.1 Exposure to ACEs

On average children were exposed to two ACEs in the last year and up to 9 accumulated ACEs (Figure 2A). The percentage of children with at least one ACE was 83% (Figure 2B). In general, the overlap between different ACEs was as expected based on the prevalence in the group as whole, i.e., for most ACE subgroups the co-occurrence of physical health problems, change in household composition, change in housing, bereavement and divorce or conflict in the family was high. Children growing up in families with financial problems appear to be disproportionately burdened by accumulated ACEs (Figure 2C).

3.2 Age and sex effects for global brain structure

For all brain estimates individual differences were large with overlap between children of different ages and sexes. Still, age effects were found for all global brain estimates and sex effects for all global estimates except average FA (Table 1, Figure 3). ICV was also positively associated with age $t(821) = 3.096$, $p_{\text{uncorr}} = .0020$, $\beta = 14.556$, $CI_{95\%} [5.328, 23.784]$ and larger for boys $t(821) = -17.964$, $p_{\text{uncorr}} < 0.0001$, $\beta = -145.426$, $CI_{95\%} [-161.317, -129.536]$. Total cortical surface area was also positively associated with age $t(821) = -2.076$, $p_{\text{uncorr}} = .0382$, $\beta = 12.025$, $CI_{95\%} [0.656, 23.395]$ and larger for boys $t(821) = -18.109$, $p_{\text{uncorr}} < .0001$, $\beta = -180.631$, $CI_{95\%} [-200.209, -161.052]$. Total average thickness was negatively associated with age $t(821) = -4.580$, $p_{\text{uncorr}} < .0001$, $\beta = -0.015$, $CI_{95\%} [-0.021, -0.0008]$ and was on average lower for boys compared to girls $t(821) = 2.830$, $p_{\text{uncorr}} = .0048$, $\beta = 0.015$, $CI_{95\%} [0.005, 0.026]$. Average FA over the TBSS skeleton was positively associated with age $t(732) = 4.732$, $p_{\text{uncorr}} < .0001$, $\beta = 0.0036$, $CI_{95\%} [0.002, 0.005]$ and not significantly associated with sex $t(732) = 0.101$, $p_{\text{uncorr}} = .9194$, $\beta = 0.004$, $CI_{95\%} [-0.002, 0.002]$. Age effects were modeled in a linear fashion only. Age² did not reach significance in the sample's narrow age range.

A) Number of ACEs per child



B) Prevalence of ACEs

ACEs	ACE (%)	N	N T1w	N DWI
Financial problems	8			
Physical health problems in the family	41			
Substance abuse in the family	2			
Problems with police, law or child protection services	3			
Change in household composition	47	1046	785	702
Change in housing	26			
Bereavement	50			
Divorce or conflict in the family	31			
Exposure to violence	11			
Bullied	6	948	707	631
Parental psychiatric diagnosis	11	1056	791	708
Any ACE	83	943	704	628

C) Overlap between two types of ACEs

Within the group of children that experienced a specific ACE (rows), what is the percentage of children that experienced another specific ACE too (columns)?

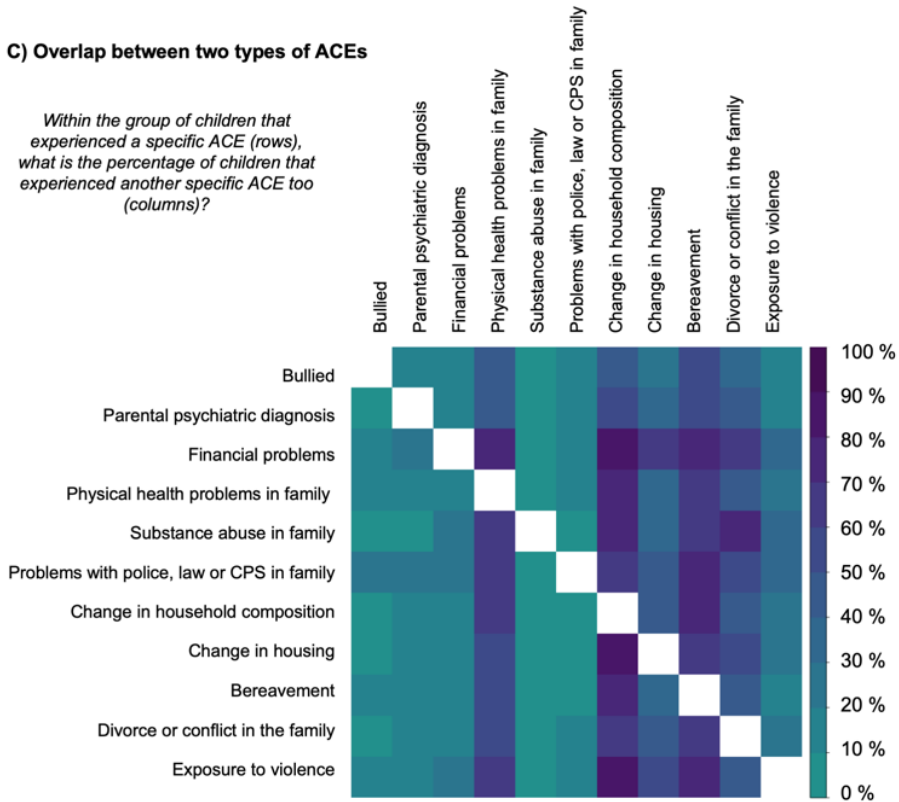


Figure 2. Descriptive statistics for the exposure to ACEs.

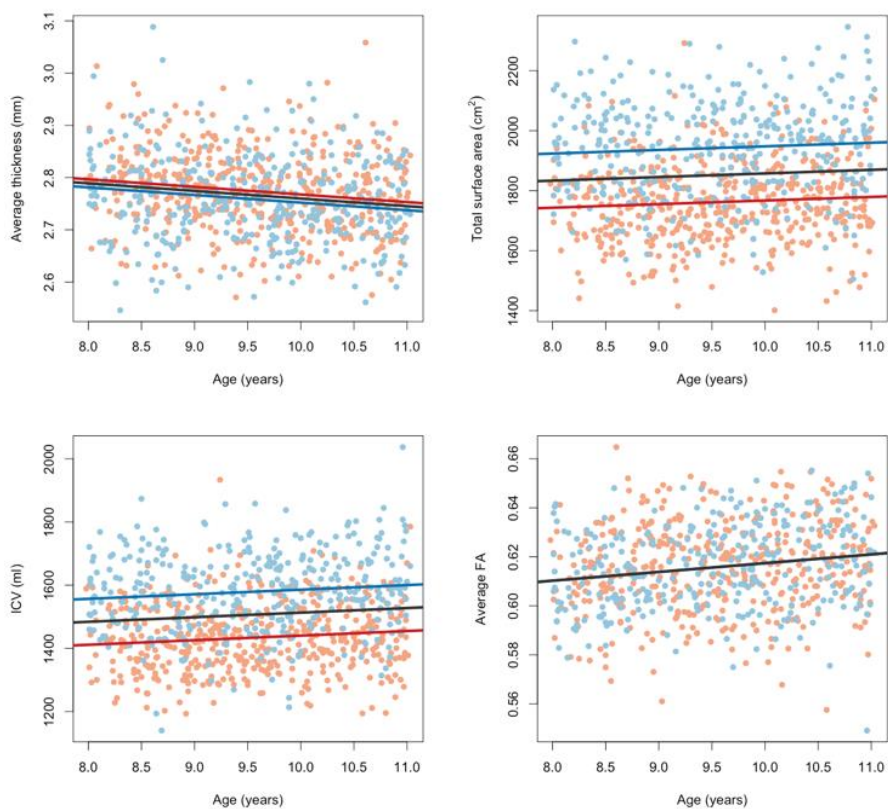


Figure 3. Effects of age and sex on global brain measures. Red dots indicate brain measures in girls and blue dots indicate brain measures in boys. Lines show the relation modeled linearly between brain measures and age (for girls in red, boys in blue and in black for the group as whole). Fractional anisotropy measures were corrected for the different acquisition protocols.

3.3 The association between ACEs and brain structure

When focusing only on the standardized effect sizes (β) and ignoring statistical significance, we observe a pattern of larger subcortical volume and larger cortical surface area in children exposed to ACEs. For fractional anisotropy and cortical thickness, results were more mixed. The results for all analyses, sorted by FDR-adjusted I values and with color-coded effect sizes, can be found in the supplementary tables S1-S5. For most ACEs, none of the ROIs reached significance ($p_{\text{FDR}} < .05$), except for children growing up in a family where substance abuse is an issue and children that grow up in an environment where the parents report exposure to violence.

Substance abuse in the household was associated with larger cortical surface area in the left superior frontal gyrus, $t(781) = 3.724$, $p_{\text{FDR}} = .0077$, $p_{\text{uncorr}} = .0002$, $\beta = -.118$, $\text{CI}_{95\%} [.056, -.181]$, right superior frontal gyrus, $t(781) = 3.409$, $p_{\text{FDR}} = .0110$, $p_{\text{uncorr}} = .0007$, $\beta = -.109$, $\text{CI}_{95\%} [.046, .172]$, left pars triangularis, $t(781) = 3.614$, $p_{\text{FDR}} = .0077$, $p_{\text{uncorr}} = .0003$, $\beta = -.121$, $\text{CI}_{95\%} [.055, .187]$, left rostral middle frontal gyrus, $t(781) = 3.163$, $p_{\text{FDR}} = .0195$, $p_{\text{uncorr}} = .0016$, $\beta = -.101$, $\text{CI}_{95\%} [.038, .164]$, and right caudal anterior cingulate gyrus, $t(781) = 2.918$, $p_{\text{FDR}} = .0348$, $p_{\text{uncorr}} = .0036$, $\beta = -.100$, $\text{CI}_{95\%} [.033, .168]$. After correction for total cortical surface area, effects were attenuated and no longer significant. Effects in the same direction were found in non-significant ROI's. Together, this suggests a more global effect on (frontal) cortical surface area. Figure 4 shows the effect sizes and scatter plots for the association between substance abuse in the household and cortical surface area.

Substance abuse in the family and cortical surface area

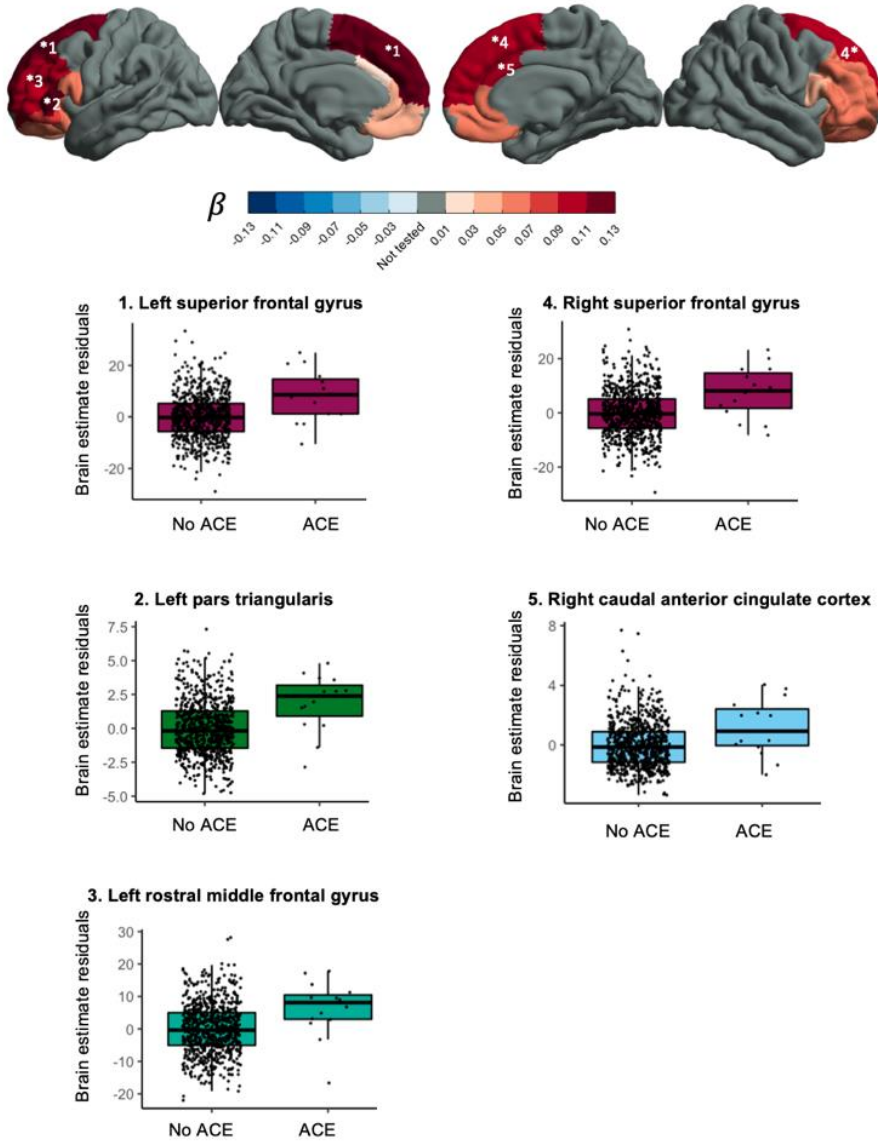


Figure 4. Associations between substance abuse in the household and cortical surface area.

Household exposure to violence was associated with lower fractional anisotropy in the left cingulum bundle hippocampus region, $t(697) = -3.154$, $p_{FDR} = .0101$, $p_{uncorr} = .0017$, $\beta = -.102$, $CI_{95\%} [-.166, -.039]$, and in the right cingulum bundle hippocampus region, $t(697) = -3.401$, $p_{FDR} = .0085$, $p_{uncorr} = .0007$, $\beta = -.121$, $CI_{95\%} [-.191, -.051]$. The direction of effect in other ROIs was mixed. Figure 5 shows the effect sizes and scatter plots for the association between exposure to violence and FA.

The main results were robust as suggested by comparable means and confidence intervals for the bootstrapping of each regression coefficient. Histograms followed a normal distribution (supplementary figures S2-S8 and supplementary table S6-S12). Bootstrapping confidence intervals for all regression coefficients are reported in the supplementary material.

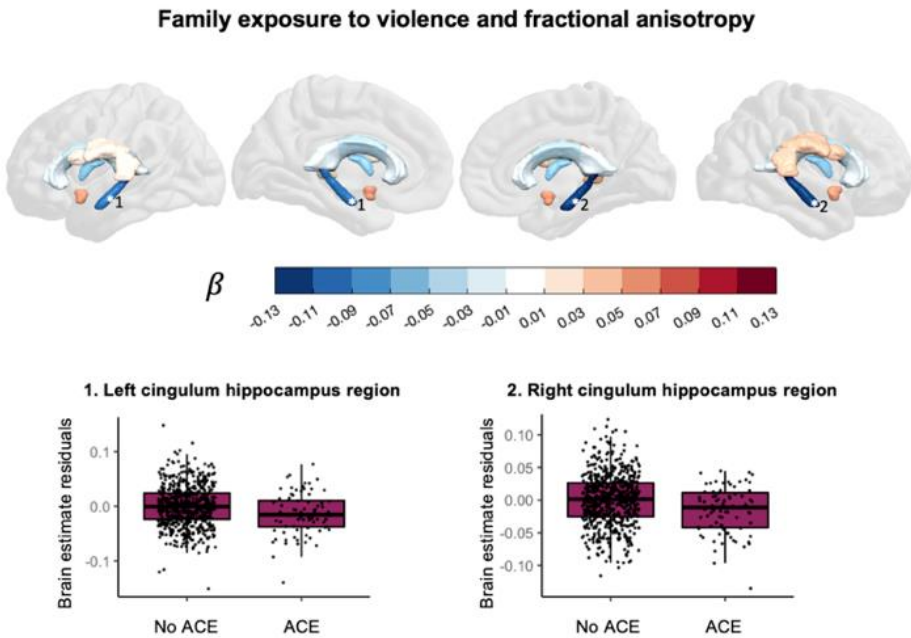


Figure 5. Associations between exposure to violence and fractional anisotropy.

3.4 Post-hoc description of sample subsets

In deviation of our data request (specified in advance of the study) and after the data were seen, we wanted to investigate whether our results could be related to attrition bias and whether our main effects could be explained by other environmental factors. The supplementary material provides the methods related to this section. Household income was analyzed in the same way as in a previous study on socioeconomic status in, among others, the YOUth cohort (Fakkel et al., 2020). Sample characteristics can be found in Table 1. First, children with data on T1-weighted brain measures and recent life events were compared to children with missing data for either the T1-weighted brain measures or the life events survey. Children with missing data scored on average two points higher on the CBCL total problems score (Table S13). The same was found for children with missing DWI data or life events data and, additionally, the percentage of fathers with another self-reported ethnicity than Dutch was higher in this group with missing data (Table S14). Second, subsamples with and without ACEs were compared for the ACEs and brain scans relevant for our main effects. We found that parents' educational attainment in years was shorter for children with versus without exposure to substance abuse in the family (Supplementary tables S15) and lower household incomes were more prevalent when parents reported exposure to violence (Supplementary table S16).

Table 1. Sample characteristics.

All available data	
<i>Sex (% girls)</i>	<i>N=955</i>
	56
<i>Mean age in years (SD)</i>	<i>N=955</i>
	9.54 (0.86)
<i>Mean CBCL total problem score (SD)</i>	<i>N=1055</i>
	22.92 (16.49)
<i>Self-reported ethnicity mother (%)</i>	<i>N=1212</i>
Dutch	91
Dutch and another ethnicity	2
Another ethnicity	7
<i>Self-reported ethnicity father (%)</i>	<i>N=955</i>
Dutch	93
Dutch and another ethnicity	2
Another ethnicity	5
<i>Mean education in years mother (SD)</i>	<i>N=1212</i>
	15.12 (2.00)
<i>Mean education in years father (SD)</i>	<i>N=955</i>
	14.86 (2.50)
<i>Gross monthly household income (%)</i>	<i>N=1133</i>
< €1.250	2
€1.250 - €2.000	6
€2.000 - €3.000	8
€3.000 - €4.000	18
> €4.000	66
<i>Number of children at home (%)</i>	<i>N=1232</i>
0 or 1	12
2	53
3 or more	35

4. Discussion

This study explored the association between various adverse childhood experiences (ACEs) and brain morphometry in selected ROIs during pre-adolescence in a cohort of over 1000 children between 7 and 11 years old. We found an association between substance abuse in the household and larger cortical surface area in frontal regions (Figure 4). Furthermore, we found evidence for an association between exposure to violence and lower fractional anisotropy in the bilateral cingulum bundle in the hippocampus region (Figure 5).

This study contributes to previous work by providing specific ACEs that could be worth further investigation: substance abuse in the family and exposure to violence. Growing up in a family with substance abuse problems was associated with a larger cortical surface area in the bilateral superior frontal gyrus, the left pars triangularis, the left rostral middle frontal gyrus and the right caudal anterior cingulate gyrus. A previous study in pre-adolescent children found an association between growing up in a family with substance abuse problems and larger cortical surface area in frontal regions (Lees et al., 2021). In that same study, thinner cortices were also found and we did not replicate that finding. For white matter we found an association between family exposure to violence and lower fractional anisotropy in the bilateral cingulum bundle hippocampus region. The cingulum is the tract that connects the frontal cortex with the parahippocampal gyrus in the temporal lobe. Exposure to violence has in the past been associated with lower quantitative anisotropy in the hippocampal cingulum (Bell et al., 2021) and lower mean diffusivity but not difference in fractional anisotropy in the hippocampal cingulum (Fani et al., 2021).

In general, we found small effects and no evidence for significant differences in most ACEs. There are several explanations to be considered. One, there is discussion as to what qualifies as an ACE. We rely on parent-report and do not know to what extent children were impacted by the ACEs. Two, only ACEs within one year before the first wave were measured. Thus, effects of ACEs that occurred before that time and may have influenced brain development could not be included in the analysis. Effects of recent ACEs on brain morphology may emerge later in development. Three, the

effects of ACEs on the brain could be too small to detect with our current method. Rather than running separate analyses for each brain measure and each ACE, integrating features from brain measures or ACEs to create latent variables may be a better approach to detect small effects. We will discuss these points below.

There is a lack of consensus which experiences qualify as ACEs. Childhood adversity has been defined as “experiences that are likely to require significant adaptation by an average child and that represent a deviation from the expectable environment” (McLaughlin, 2016). Also, ACEs have been defined as “childhood events, varying in severity and often chronic, occurring in a child’s family or social environment that cause harm or distress, thereby disrupting the child’s physical or psychological health and development” (Kalmakis & Chandler, 2014). An extensive body of research shows that experiences of maltreatment, sexual abuse and neglect impact the brain (Hart & Rubia, 2012; McLaughlin et al., 2019; Teicher & Samson, 2016). However, in our study no data experiences of maltreatment, sexual abuse or neglect were available. Therefore, we focused on other types of ACEs. Our broad definition of ACEs results in only 17% of the children that were not exposed to any type of ACE and thus may not have fully captured the expected complexity and dimensionality of adversity. Because we did not assess the impact or severity of the ACE and did not include measures of functioning, it remains unclear how children experienced these events and whether the experiences are so disruptive that brain development or (future) functioning could be affected.

Another explanation for the small effects is that we study recent experiences, disregarding prenatal early life stress and ACEs that occurred more than a year ago. Effects of recent ACEs on brain structure may emerge later in development or even adulthood (Tottenham & Sheridan, 2009). It is possible that ACEs may impact brain structure over time, but effects on brain functioning may be easier to detect shortly after the experience. There are numerous studies that find a relation between ACEs and brain functioning during childhood or adolescence (Brieant et al., 2021; Cassiers et al., 2018; Hart & Rubia, 2012; Kraaijenvanger et al., 2020; Teicher & Samson, 2016). For the effects that we found, it is plausible that they refer to an adverse family environment throughout childhood rather than a single event that occurred in the

last year. For example, substance abuse problems in the family could have been present throughout childhood or even throughout pregnancy. No information on duration and severity of the ACEs was available and prenatal exposure to alcohol or drugs was not included in this study. In the same way, exposure to violence could be an indication of an adverse family environment in general or neighborhood disadvantage. To get a general idea of the broader social environment, we tested for difference in parental education, parental ethnicity, household income, child's psychopathology and number of children in the household. We found that lower household incomes were more prevalent when parents reported exposure to violence and shorter parental education was related to substance abuse in the household.

A last point to consider is that there are different methodological approaches to deal with the small effect sizes when studying the relationship between childhood adversity and brain structure. In our study we opted for a broad approach to include regions beyond the traditionally studied fronto-limbic structures and to study each ACE separately. Our approach is similar to a recent study in 398 older adults that also included a large number of ROIs and different types of ACEs (10). In this study a similar pattern of subtle effects of ACEs on brain morphology was found. Another approach is to integrate brain measures (atlas-based or voxel-vertex-wise) using multivariate techniques, e.g. non-negative matrix factorization (Anderson et al., 2014), independent component analysis, canonical correlation analysis or partial least squares approaches (Sui et al., 2012). Principal component analysis was used in a study on the effects of brain structure and adverse lifetime experiences in adults (Gheorghe et al., 2021). Another approach would be to improve the way that childhood adversity is measured, for example using extensive interviews to assess different ACEs and their impact (Ansell et al., 2012). However, this approach is often not feasible in population-based cohorts. On the whole, statistical power remains a challenge when studying the effect of many ACEs on many brain morphology estimates in population cohorts.

For the main effects it remains unresolved if they are environmental, genetic or both. Neural effects in response to adversity could be adaptive in the short-term, but in the long-term these adaptations may contribute to risk or resilience. Lockdown

restrictions during the COVID-19 pandemic can be seen as another example of an environmental stressor to some (Achterberg et al., 2021; Creswell et al., 2021; Luijten et al., 2021; Panchal et al., 2021; van der Laan et al., 2021). All children in our study were measured before the onset of the pandemic, but pandemic-related ACEs may impact the follow-up data collected in the YOUh cohort study. Importantly, accelerated or delayed brain development could also be driven by genetic factors (Brouwer et al., 2021, 2022) and the environment that parents can provide for their children is also influenced by genetics (Hart et al., 2021; Kong et al., 2018).

Three limitations of this study were not yet discussed. One, MRI data were processed using adult templates while the use of age appropriate templates could have improved the registrations (Yoon et al., 2009). Two, the group of children with missing MRI data scored slightly higher on the CBCL total problem scale and missing DWI data was more prevalent when children had a father with another self-reported ethnicity than Dutch. Three, the described cohort is homogenous with regard to self-reported ethnicity and on average participants have high socioeconomic status, and thus the sample is not representative of the general Dutch population (Fakkel et al., 2020) and results cannot readily be generalized to other parts of the world.

A better understanding of the impact of adversity on neural development is important given current pressing societal issues (Sheridan & McLaughlin, 2022). Early detection of children that are at risk for negative outcomes later in life can help policy makers, health care professionals, families and schools to break with childhoods characterized by accumulated ACEs. Developmental neuroscience can play a crucial role to inform these interventions with regard to sensitive periods of development.

5. Supplementary materials

Some of the supplementary files were too large to be incorporated here and can be downloaded online via this link <http://doi.org/10.17605/OSF.IO/M5R3U>. For Chapter 4, the online supplement includes one file with the supplementary tables S1 to S5 listing all the results from the main analyses.

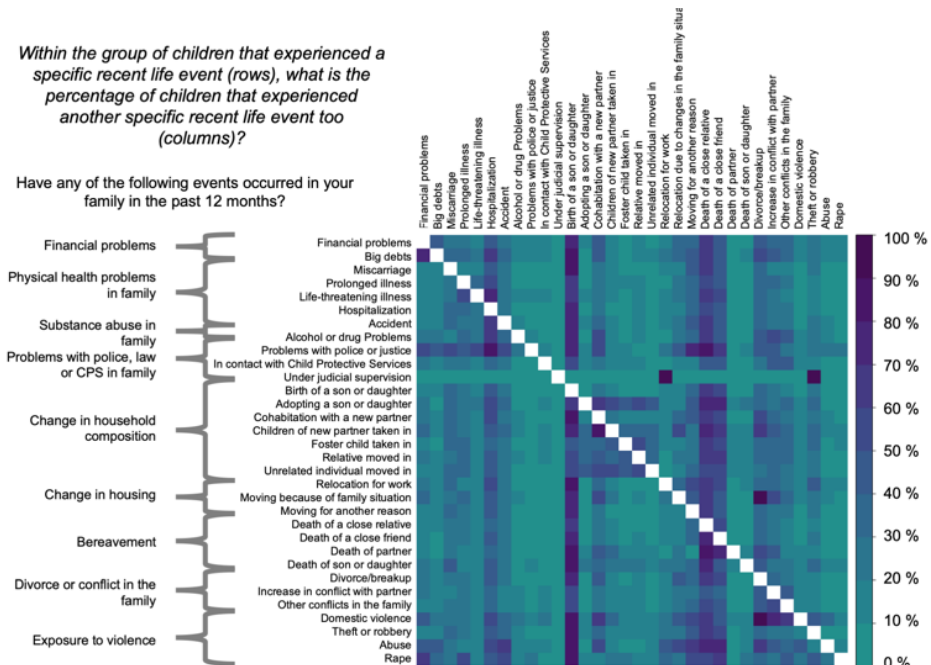


Figure S1. Overlap between scores on items of the recent life events questionnaire.

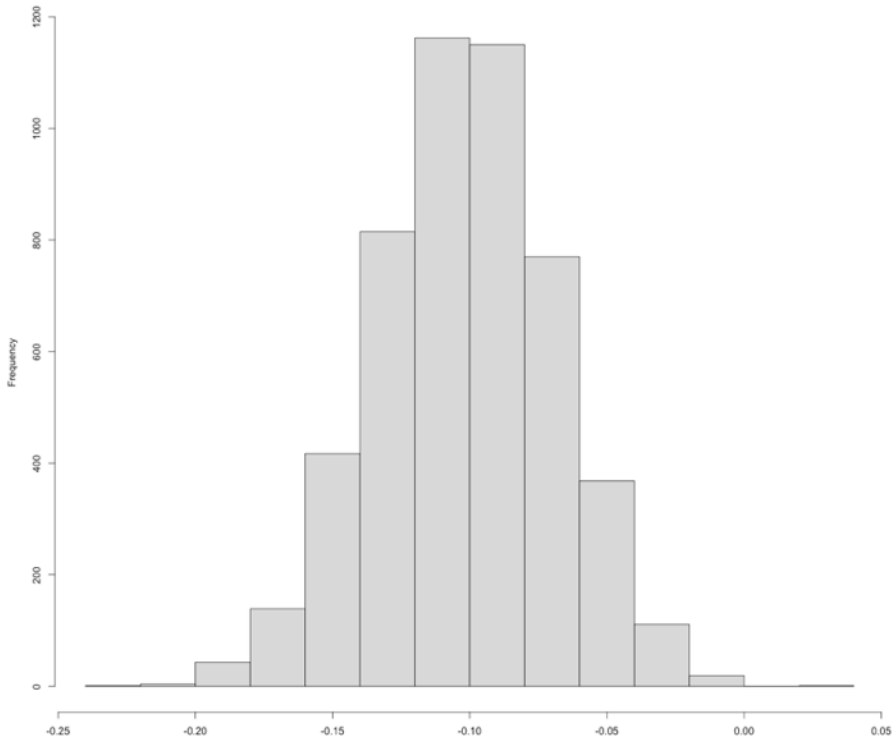


Figure S2. Bootstrap distribution for family exposure to violence and fractional anisotropy in the left cingulum bundle hippocampus region

Table S6 Bootstrap results for family exposure to violence and fractional anisotropy in the left cingulum bundle hippocampus region compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	-.10 [-.17, -.04]	-.10 (0.03) [-.17, -.04]
Age	.05 [-.01, .12]	.05 (0.03) [-.01, .12]
Sex	-.09 [-.16, -.03]	-.09 (0.03) [-.16, -.03]
DWI Acquisition	-.49 [-.56, -.43]	-.49 (0.03) [-.55, -.44]

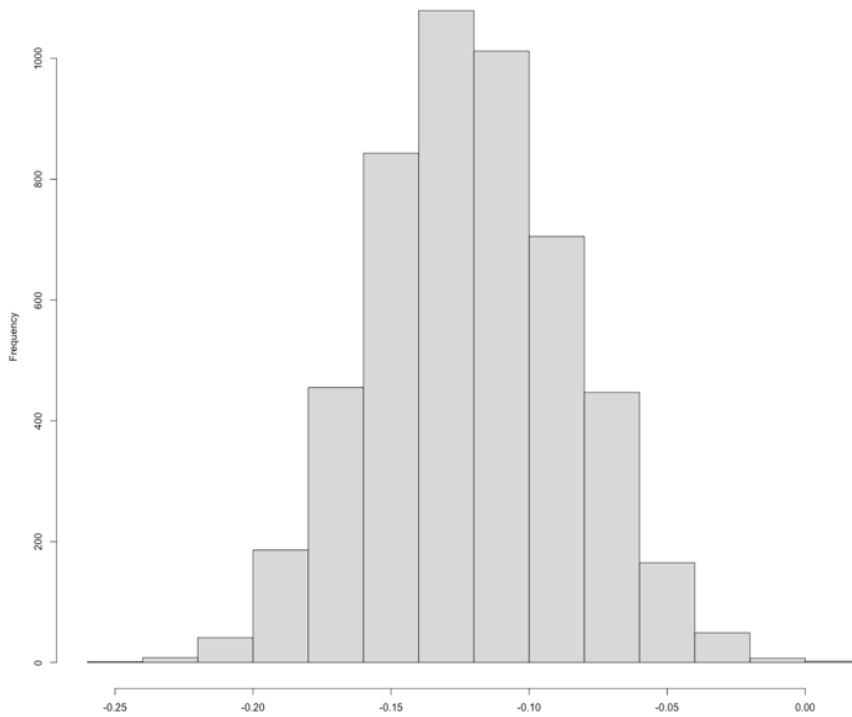


Figure S3. Bootstrap distribution for family exposure to violence and fractional anisotropy in the right cingulum bundle hippocampus region

Table S7 Bootstrap results for family exposure to violence and fractional anisotropy in the right cingulum bundle hippocampus region compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	-0.12 [-0.19, -0.05]	-0.12 (0.03) [-0.19, -0.05]
Age	0.07 [0.00, 0.14]	0.07 (0.04) [0.00, 0.14]
Sex	-0.04 [-0.11, 0.03]	-0.04 (0.04) [-0.10, 0.04]
DWI Acquisition	-0.30 [-0.37, -0.23]	-0.30 (0.03) [-0.37, -0.24]

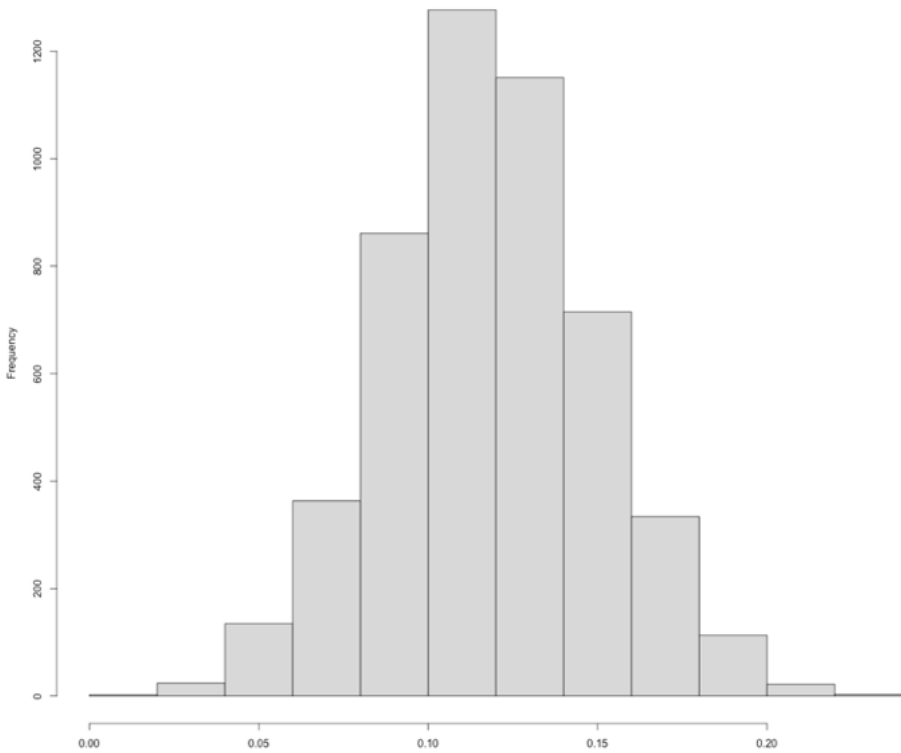


Figure S4. Bootstrap distribution for substance abuse in the household and cortical surface area in the left superior frontal gyrus

Table S8 Bootstrap results for substance abuse in the household and cortical surface area in the left superior frontal gyrus compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	.12 [.06, .18]	.12 (0.03) [.06, .18]
Age	.09 [.02, .15]	.09 (0.03) [.02, .15]
Sex	-.43 [-.50, -.37]	-.44 (0.03) [-.49, -.38]

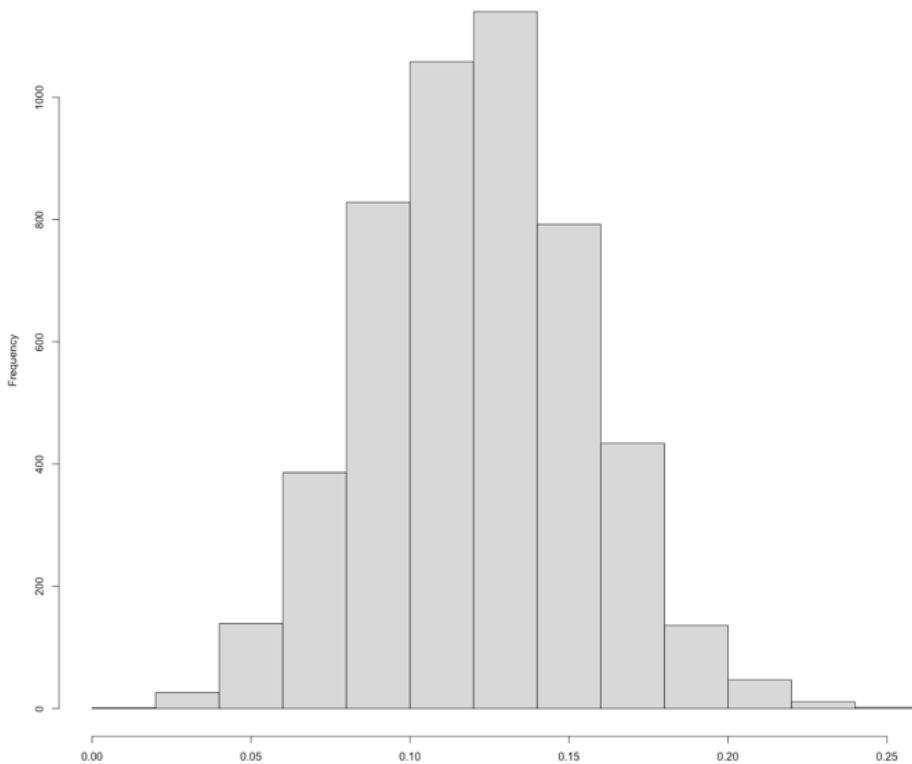


Figure S5. Bootstrap distribution for substance abuse in the household and cortical surface area in the left pars triangularis

Table S9 Bootstrap results for substance abuse in the household and cortical surface area in the left pars triangularis compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	.12 [.06, .19]	.12 (0.03) [.06, .19]
Age	.07 [.01, .14]	.07 (0.03) [.01, .14]
Sex	-.32 [-.39, -.26]	-.32 (0.03) [-.38, -.26]

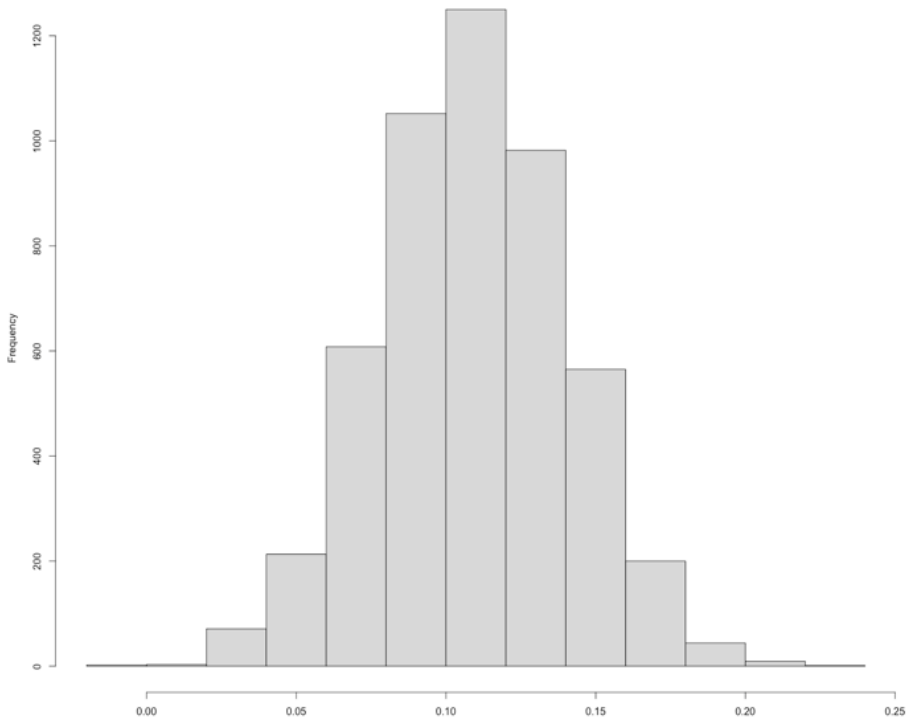


Figure S6. Bootstrap distribution for substance abuse in the household and cortical surface area in the right superior frontal gyrus

Table S10 Bootstrap results for substance abuse in the household and cortical surface area in the right superior frontal gyrus compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	.11 [.05, .17]	.11 (0.03) [.05, .17]
Age	.06 [-.00, .12]	.06 (0.03) [.00, .12]
Sex	-.43 [-.49, -.37]	-.43 (0.03) [-.48, -.37]

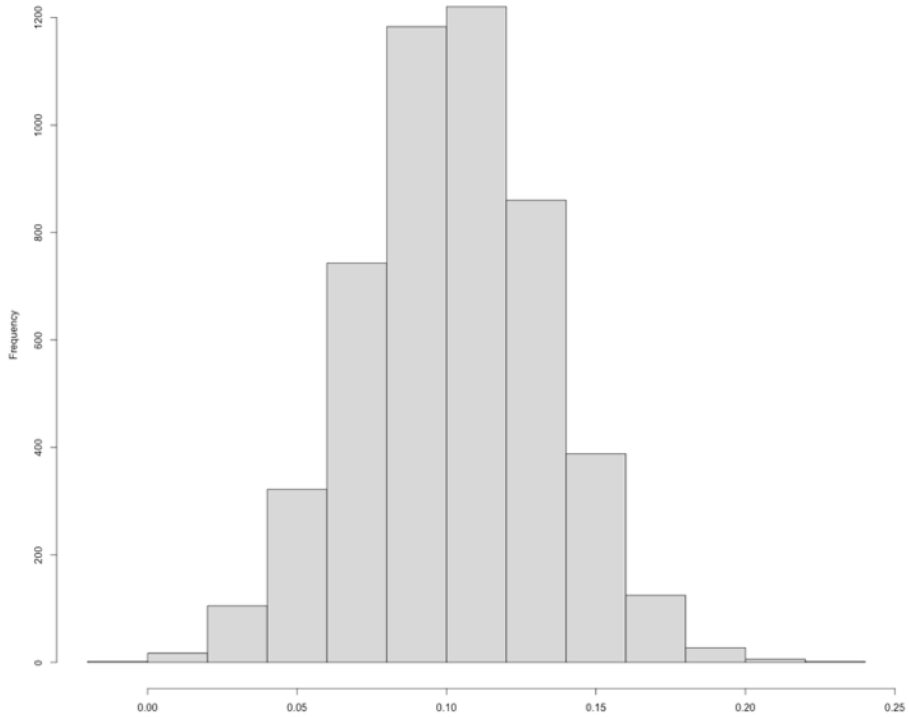


Figure S7. Bootstrap distribution for substance abuse in the household and cortical surface area in the left rostral middle frontal gyrus

Table S11 Bootstrap results for substance abuse in the household and cortical surface area in the left rostral middle frontal gyrus compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	.10 [.04, .16]	.10 (0.03) [.04, .16]
Age	.09 [.03, .15]	.09 (0.03) [.03, .15]
Sex	-.43 [-.49, -.36]	-.43 (0.03) [-.48, -.37]

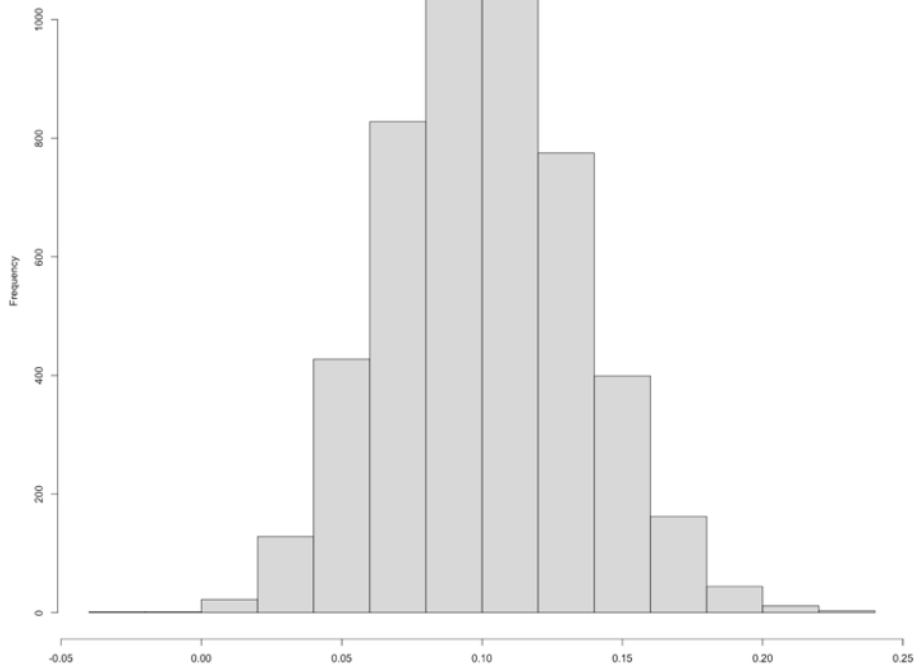


Figure S8. Bootstrap distribution for substance abuse in the household and cortical surface area in the right caudal anterior cingulate gyrus

Table S12 Bootstrap results for substance abuse in the household and cortical surface area in the right caudal anterior cingulate gyrus compared to main results

	Beta [95% CI]	Bootstrapping mean (SD) [95% CI]
ACE	.10 [.03, .17]	.10 (0.03) [.04, .17]
Age	.07 [.00, .14]	.07 (0.03) [.01, .14]
Sex	-.25 [-.32, -.18]	-.25 (0.03) [-.32, -.18]

Supplementary method – Post-hoc analyses

Definitions

We used *mother* to indicate one parent or guardian (mostly female). We used *father* to indicate the second parent or guardian (mostly male). The *primary participating parent* is the parent or guardian that fills in the surveys when the report from only one parent is sufficient. The primary participating parents was the female parent in most cases.

CBCL total problem score

One parent filled out the CBCL to report on problem behavior of their participating child. The 118 specific problem items of the CBCL (0-to-2 scale) were summed for each child. Three additional open-ended questions allowed the respondent to fill out additional problems. We decided not to include these open-ended items in our sum score as most answers were overlapping with the specific problems and selecting valid problems from the answers was not feasible given our large sample size.

Parental self-reported ethnicity

Self-reported ethnicity was based on the question “*What is your ethnic background (multiple answers possible)?*”. Multiple choice options were as followed:

1. *Dutch*
2. *European Union (excluding the Netherlands)*
3. *Turkish*
4. *European but not part of the European Union*
5. *Moroccan*
6. *African (excluding Morocco)*
7. *Surinamese*
8. *Antillean (from the former Netherlands Antilles, including Aruba, Bonaire, Curaçao, Saba, Sint Eustatius and Sint Maarten)*
9. *Latin American (excluding Suriname and the former Netherlands Antilles)*
10. *Indonesian*
11. *Asian (excluding Indonesia and Japan)*
12. *Other countries outside Europe (United States, Canada, Japan, Oceania)*

The multiple-choice options in our study were based on the former categorization by the Statistics Netherlands (CBS) based on countries that share specific migration history with the Netherlands and a distinction between ‘western’ and ‘non-western’ migration. From 2022 CBS uses a new classification system replacing the ‘western’ and ‘non-western’ main categories (<https://www.cbs.nl/en->

[gb/longread/statistische-trends/2022/new-classification-of-population-by-origin](#)). Given the small prevalence of parents with a migration background in our cohort and to not be bound by the categorization system above, we decided to indicate if parents had at least one other self-reported ethnicity in addition to or instead of Dutch.

Parental education level

Educational attainment was calculated by conversing the highest completed education to education in years to facilitate international comparison:

- *Primary education (BAO) – 6 years*
- *Special primary education (SBAO) – 6 years*
- *Education abroad comparable to primary education in the Dutch school system – 6 years*
- *(Secondary) Special education ((V)SO) – 10 years*
- *Practical education (PRO) – 10 years*
- *Secondary vocational education 'basis/kader' (VMBO-BK) – 10 years*
- *Secondary vocational education 'theoretische leergang' (VMBO-TL) – 10 years*
- *Higher general secondary education (HAVO) – 11 years*
- *Higher secondary education (VWO) – 12 years*
- *Secondary vocational education (MBO) – 14 years*
- *Education abroad comparable to practical or secondary vocational education in the Dutch school system – 12 years*
- *Applied university (HBO) – 15 years*
- *Education abroad comparable to applied university in the Dutch school system – 15 years*
- *University education (WO) – 17 years*
- *Education abroad comparable to higher vocational education or university in the Dutch school system – 17 years*

Monthly household income

Similar to the study by Fakkkel et al. (2020), monthly household income was based on the answer of the primary participating parent on the question: “Can you make an estimation of [if single] your gross income or [if not single] the total gross household income?”. Options “prefer not to say” and “can’t estimate” were not included in the tables and percentages.

Number of children at home

Both parents were asked about the composition of their household. We assumed that the answers of the primary participating parent were most indicative of the number

of children that the participating child grows up around. Therefore, we created a sum score based on the answers of the participating parent.

Who do you live with? (multiple answers possible):

- *Indicate the number of biological children*
- *Indicate the number of foster children*
- *Indicate the number of adopted children*
- *Indicate the number of stepchildren*

Next, the sum scores were classified to 0 or 1; 2; over 3 children at the home of the participating parent.

Statistics

To compare the main sample versus the subsets (Table S13 and S14) and the children with versus without ACEs (Table S15 and Table S16), we used Chi-squared tests for the categorical variables and the Mann-Whitney U test (non-parametric because of skewness) for the CBCL and the education in years. We used an alpha level of .05 to test for significance.

Table S13 Responders versus non-responders for T1 and life events data

	T1 and life events survey available	
	Yes	No - missing
<i>Sex (% girls)</i>	<i>N=785</i>	<i>N=170</i>
	56	56
<i>Mean age in years (SD)</i>	<i>N=785</i>	<i>N=170</i>
	9.56 (0.85)	9.44 (0.89)
<i>Mean CBCL total problems (SD)*</i>	<i>N=642</i>	<i>N=413</i>
	22.17 (16.42)	24.10 (16.54)
<i>Self-reported ethnicity mother (%)</i>	<i>N=740</i>	<i>N=472</i>
Dutch	92	90
Dutch and another ethnicity	2	3
Another ethnicity	6	7
<i>Self-reported ethnicity father (%)</i>	<i>N=592</i>	<i>N=363</i>
Dutch	94	91
Dutch and another ethnicity	2	2
Another ethnicity	3	6
<i>Mean education in years mother (SD)</i>	<i>N=740</i>	<i>N=472</i>
	15.21 (1.92)	14.97 (2.11)
<i>Mean education in years father (SD)</i>	<i>N=592</i>	<i>N=363</i>
	14.91 (2.47)	14.76 (2.54)
<i>Gross monthly household income (%)</i>	<i>N=690</i>	<i>N=443</i>
< €1.200	2	1
€1.200 - €2.000	5	7
€2.000 - €3.200	9	7
€3.200 - €4.000	17	20
> €4.200	68	64
<i>Number of children at home (%)</i>	<i>N=750</i>	<i>N=472</i>
0 or 1	11	13
2	54	51
3 or more	35	35

* significant Mann-Whitney U test.

Table S14 Responders versus non-responders for DWI and life events data

	DWI and life events survey available	
	Yes	No - missing
<i>Sex (% girls)</i>	<i>N=702</i>	<i>N=253</i>
	57	55
<i>Mean age in years (SD)</i>	<i>N=702</i>	<i>N=253</i>
	9.55 (0.85)	9.49 (0.88)
<i>Mean CBCL total problems (SD)*</i>	<i>N=578</i>	<i>N=477</i>
	22.17 (16.70)	23.84 (16.20)
<i>Self-reported ethnicity mother (%)</i>	<i>N=660</i>	<i>N=552</i>
Dutch	91	90
Dutch and another ethnicity	2	3
Another ethnicity	7	7
<i>Self-reported ethnicity father (%)**</i>	<i>N=531</i>	<i>N=424</i>
Dutch	95	91
Dutch and another ethnicity	2	2
Another ethnicity	3	6
<i>Mean education in years mother (SD)</i>	<i>N=660</i>	<i>N=552</i>
	15.21 (1.93)	15.00 (2.08)
<i>Mean education in years father (SD)</i>	<i>N=531</i>	<i>N=424</i>
	14.98 (2.41)	14.71 (2.60)
<i>Gross monthly household income (%)</i> <i>N=617</i>	<i>N=548</i>	
< €1.200	1	2
€1.200 - €2.000	5	7
€2.000 - €3.200	8	8
€3.200 - €4.000	17	20
> €4.200	69	64
<i>Number of children at home (%)</i>	<i>N=669</i>	<i>N=563</i>
0 or 1	11	13
2	54	52
3 or more	35	34

* significant Mann-Whitney U test. ** significant χ^2 test.

Table S15 Demographics for children with and without experiences of substance abuse in the household

	Substance abuse in the household	
	Yes	No
<i>Sex (% girls)</i>	<i>N=15</i>	<i>N=770</i>
	53	56
<i>Mean age in years (SD)</i>	<i>N=15</i>	<i>N=770</i>
	9.52 (0.84)	9.56 (0.85)
<i>Mean CBCL total problems (SD)</i>	<i>N=10</i>	<i>N=632</i>
	17.20 (7.19)	22.24 (16.51)
<i>Self-reported ethnicity mother (%)</i>	<i>N=14</i>	<i>N=726</i>
Dutch	86	92
Dutch and another ethnicity	7	2
Another ethnicity	7	6
<i>Self-reported ethnicity father (%)</i>	<i>N=13</i>	<i>N=579</i>
Dutch	100	94
Dutch and another ethnicity	0	2
Another ethnicity	0	3
<i>Mean education in years mother (SD)*</i>	<i>N=14</i>	<i>N=726</i>
	14.07 (2.23)	15.23 (1.91)
<i>Mean education in years father (SD)*</i>	<i>N=13</i>	<i>N=579</i>
	13.38 (2.96)	14.95 (2.45)
<i>Gross monthly household income (%)</i>	<i>N=1</i>	<i>N=679</i>
< €1.200	9	1
€1.200 - €2.000	0	5
€2.000 - €3.200	9	9
€3.200 - €4.000	27	17
> €4.200	55	68
<i>Number of children at home (%)</i>	<i>N=14</i>	<i>N=736</i>
0 or 1	14	11
2	64	54
3 or more	21	35

* significant Mann-Whitney U test.

Table S16 Demographics for children with and without family exposure to violence

	Exposure to violence	
	Yes	No
Sex (% girls)	<i>N=80</i>	<i>N=622</i>
	66	56
Mean age in years (SD)	<i>N=80</i>	<i>N=622</i>
	9.54 (0.86)	9.55 (0.85)
Mean CBCL total problems (SD)	<i>N=68</i>	<i>N=510</i>
	24.06 (20.98)	21.91 (16.05)
Self-reported ethnicity mother (%)	<i>N=76</i>	<i>N=584</i>
Dutch	86	92
Dutch and another ethnicity	4	2
Another ethnicity	11	6
Self-reported ethnicity father (%)	<i>N=67</i>	<i>N=464</i>
Dutch	93	95
Dutch and another ethnicity	4	2
Another ethnicity	3	3
Mean education in years mother (SD)	<i>N=76</i>	<i>N=584</i>
	15.28 (2.03)	15.20 (1.92)
Mean education in years father (SD)	<i>N=67</i>	<i>N=464</i>
	14.75 (2.78)	15.01 (2.36)
Gross monthly household income (%) * <i>N=69</i>	<i>N=548</i>	
< €1.200	6	1
€1.200 - €2.000	6	5
€2.000 - €3.200	4	9
€3.200 - €4.000	16	17
> €4.200	68	69
Number of children at home (%)	<i>N=76</i>	<i>N=593</i>
0 or 1	18	10
2	58	53
3 or more	24	37

* significant χ^2 test.

Chapter 5 - Inter-individual differences in facial emotion processing in pre-adolescence

Elizabeth E.L. Buimer, Pascal Pas, Carlijn van den Boomen, Mathijs Raemaekers,
Rachel M. Brouwer, Hilleke E. Hulshoff Pol

5

In preparation

Abstract

Identification of facial expressions is important to navigate social interactions. Deficits in behavioral emotion labeling are associated with less favorable developmental outcomes. Furthermore, functional MRI studies show that aberrant neural processing of facial emotions is a risk factor for psychiatric outcomes later in life. While it is presumed that behavioral emotion labeling, neural processing of emotional faces and social competence are related, this has rarely been studied. Here, we investigate interrelations between these three aspects, as well as associations with age and sex, in a cohort of 1055 children between 8 and 11 years old participating in the Dutch YOUth study, using a multistep linear modelling statistical approach. Behavioral emotion labeling was measured using labeling of basic emotions on pictures of facial expressions (accuracy and median response time during correct trials); neural processing of emotional faces was based on a passive-watching task with alternating blocks of images of houses or faces (happy, fearful, and neutral expressions) during fMRI; social competence included questions on perspective taking, empathic concern, peer problems and prosocial behavior. Parents reported higher social competence for girls on all subscales. For the subscales prosocial behavior and perspective taking older children scored higher. We replicate recent studies that an advantage in emotion labeling accuracy and reaction time is present in older as compared to younger children and in girls compared to boys. These age- and sex-differences were not reflected in the fMRI signal when comparing emotional versus neutral faces. During fMRI, emotional faces elicited more activity than neutral faces, and happy faces elicited more activity than fearful faces in boys and girls. However, we did not find evidence for the hypothesized links between social competence and behavioral emotion labeling, and with neural activity. To conclude, age- and sex-related variation in emotion labeling and social competence exists in pre-adolescence, but neural activity, behavioral emotion labeling and social competence may reflect separate constructs to navigating social interactions in a typically developing population.

1. Introduction

Social competence can be defined as the ability to engage in meaningful interactions with others (Junge et al., 2020). Emotion reasoning helps to navigate these social interactions. Emotion reasoning is the ability to make reasonably accurate inferences and predictions about the emotion states of other people (Ruba & Pollak, 2020). Emotion labelling is considered one of the building blocks of emotion reasoning (Ruba & Pollak, 2020). In this manuscript we will use the term *emotion labeling* rather than the traditional term *emotion recognition* that is now considered less favorable (Barrett et al., 2019; Hoemann et al., 2020; Ruba & Pollak, 2020). Emotion labeling is correlated with higher social competence and less behavioral problems in childhood and adolescence (Trentacosta & Fine, 2010). Individuals differ in their ability to accurately label emotions on facial expressions and in the speed at which emotion labeling occurs. Accuracy and speed of emotion labeling are partly heritable (Swagerman et al., 2016), but also influenced by environmental factors (Assed et al., 2020; Bérubé et al., 2023).

During development, the age and sex of a child is an important contributor to inter-individual variation in emotion labeling skills. Developmental trajectories of emotion labeling start early in life. Newborns are not able to differentially process emotional faces, but this ability rapidly develops in the first year of life and continues to improve during childhood (Bayet & Nelson, 2019). The ability to accurately label happy faces is thought to develop first, while for example the identification of fearful emotions has a more protracted developmental trajectory (Bayet & Nelson, 2019; Durand et al., 2007). Older children are faster and more accurate in emotion labeling (Herba & Phillips, 2004). The effect of age decreases after the first year of life, but age effects on emotion labeling accuracy and reaction time continue to exist throughout pre-adolescence (Gur et al., 2012; Verpaalen et al., 2019). Regarding sex effects, most studies report a female advantage in emotion labeling skills (McClure, 2000).

Age- and sex-related variation in emotion labeling skills (accuracy and reaction time) may also reflect underlying developmental processes in the brain. Functional MRI (fMRI) can be used to study the neural processing of emotional faces. The neural basis

of facial-emotion processing requires different levels of specialization, as faces convey a range of hierarchically embedded information (Adolphs, 2002; Bayet & Nelson, 2019). Developmental studies suggest that separate processes underlie the perception of emotional faces and the processing of other facial information such as identity, even though these processes can affect each other (Bayet & Nelson, 2019). Facial emotion processing engages visual, emotional, and higher-cognitive processing pathways. A meta-analysis including 105 fMRI studies on the processing of emotional faces showed that emotional faces elicited activity in several visual, limbic, temporoparietal and prefrontal areas; the putamen; and the cerebellum (Fusar-Poli et al., 2009). Neural activity in the visual cortex and cerebellum was observed independent of emotional valence. Happy, fearful, and sad faces specifically activated the amygdala. Disgusted and angry faces specifically activated the insula (Fusar-Poli et al., 2009).

Emotion labeling abnormalities impact developmental outcomes: emotion-specific differences in labeling accuracy were associated with internalizing and externalizing problems in preschoolers participating in the Generation R cohort (Székely et al., 2014). Even more, reviews show that emotion labeling abnormalities are associated with a wide variety of neuropsychiatric conditions, such as autism spectrum disorder (Harms et al., 2010), mood disorders, anxiety disorders or attention deficit hyperactivity disorder (Collin et al., 2013). The relation between emotion labeling abnormalities and psychiatric vulnerability co-exists with atypical neural processing of emotions. Aberrant neural processing of emotional faces is one of the most consistent neuroimaging findings in the childhood maltreatment literature (Hein & Monk, 2017) and is related to various psychiatric conditions (Delvecchio et al., 2013; Etkin & Wager, 2007; Harms et al., 2010; Mitchell et al., 2014; Stuhmann et al., 2011). So far, behavioral emotion labeling, neural processing of emotional faces and social competence are mostly studied in isolation and the interrelations between these three aspects requires further investigation. Investigating interrelations between these three factors may help further the understanding of mechanisms underlying social behavior.

In this study we aim to investigate age- and sex-related differences in the behavioral emotion labeling and neural processing of facial expressions of emotions in pre-adolescence. Furthermore, we are interested in the link between inter-individual differences in emotion labeling accuracy and reaction time, neural facial-emotion processing, and social competence. We hypothesize 1) that older children are faster and more accurate when labeling emotions and that there is a female advantage; 2) that shorter emotion labeling reaction time and higher accuracy is related to higher social competence; 3) that variation in neural processing of emotional faces can be partly explained by age, sex and emotion labeling skills, with older children, girls and children with superior emotion labeling skills showing different activation patterns; 4) that social competence correlates with brain activation patterns in response to emotional versus neutral faces.

2. Materials and methods

2.1 Participants

The YOUTH cohort study is a longitudinal population-based study on brain development with a specific focus on social competence and self-regulation (Onland-Moret et al., 2020). In the current study we included data from the first wave of YOUTH: Child & Adolescent, in which 1332 children between 7.9 and 11.0 years old participated (57% female). The YOUTH cohort study was conducted in Utrecht, a province of the Netherlands, with on average highly educated inhabitants with high incomes (Fakkel et al., 2020; Buimer et al., 2022). All data included here was collected prior to the COVID-19 pandemic. This study was approved by the Medical Research Ethics Committee Utrecht. Children participated on a voluntary basis and parents/guardians gave written consent and assent. Figure 1 shows the available data for the domains relevant to the current study.

2.2 Social competence data

Social competence was defined using the subscales *perspective taking* and *empathic concern* of the Interpersonal Reactivity Index (IRI, self-report, Davis, 1983) and the subscales *prosocial behaviour* and *peer problems* of the Strengths and Difficulties Questionnaire (SDQ, parent-report, Goodman, 1997, 2001). Together the subscales of the two questionnaires tap different aspects of social competence (Junge et al., 2020). Each of the four subscales contains 5 items, which were summed to get total subscale scores for each child.

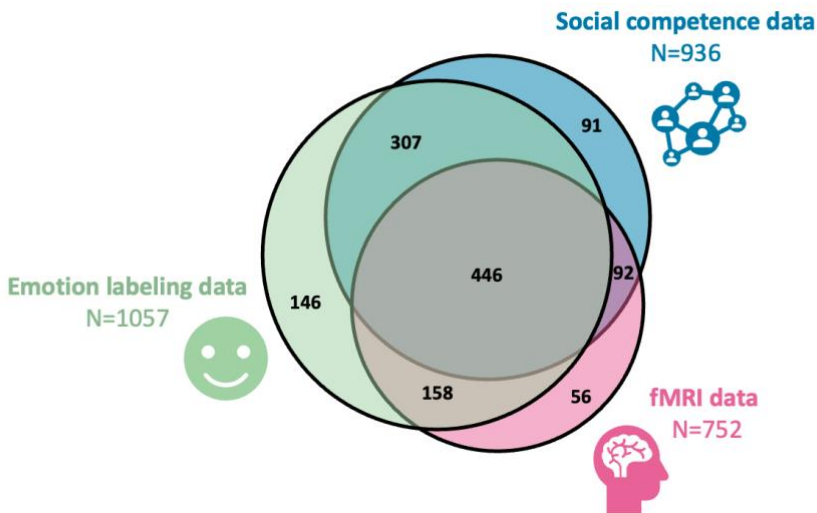


Figure 1 Venn diagram of the available data. Labels specify the data domains of interest for this study and the total number of participants with data for the domain. Colors of the labels correspond to the colors of the three circles. Area of the circles and the overlapping spheres are proportional, and numbers indicate absolute numbers of children. Figure adapted from web application DeepVenn (Hulsen, 2020).

2.3 Behavioral emotion labeling

The Penn Computerized Neurobehavioral Battery (CNB) is developed by the University of Pennsylvania to capture specific cognitive domains that link to brain function (Gur et al., 2010). Within YOUth: Child & Adolescent, a subset of the web-based CNB was collected, including the 40-item Emotion Recognition Test (ER-40). In the ER-40, the child labels the emotion on presented images of facial expressions in a multiple-choice format: happy, sad, anger, fear or neutral. The multiple-choice options were presented in the children's native language, Dutch. From the ER-40, we used accuracy (the number of correct responses) and reaction time (the median response times computed over the trials with correct identifications). We did not observe irregularities in the data due to non-compliance (for example, continuously picking the same answer). One participant had a response time for fearful facial expressions of 11.6 seconds. The participant with this extreme outlier was removed from the dataset because the median response time for fearful faces was based on only this one correct trial resulting in a response time 13 standard deviations from the mean of 2.5 seconds. Other outliers were not as extreme or based on more than one trial. As we were interested in inter-individual variation, we did not remove these other outliers. Boxplots of reaction times in relation to correct responses with and without the outlier can be found in Supplementary figure S1.

2.4 Neuroimaging data

2.4.1 Stimuli presentation

The face/house task is a passive viewing task in which children are presented with pictures of faces in a semi-random order (happy, fearful or neutral expression) and pictures of houses. Stimuli are presented in blocks of 18 seconds, with four blocks of houses and four blocks of faces. Within each block, stimuli are presented for 1 second followed by a 1 second fixation cross. Between blocks there is a period of rest. To ensure that the children remain focused, they are instructed to press a button when a red circle appears in the center of the screen, after each rest period. No other behavioral data is collected during the scan. We used stimuli from the Radboud Faces

Database (Langer et al., 2010). The stimuli were presented on an MRI-compatible 23-inch LCD screen with a resolution of 1080 by 1920 pixels (BOLDscreen, Cambridge Research Systems).

2.4.2 Neuroimaging acquisition

In the YOUth cohort study, the collection of MRI data is monitored closely over time based on human data and weekly collected phantom data. The YOUth MRI protocol, quality control and test-retest reliability are described in detail elsewhere (Buimer et al., 2020). All data was acquired on the same Philips Ingenia CX 3.0 T MRI scanner. Whole-brain, T2*-weighted echo planar images were acquired with the following parameters: TR = 1000 ms; TE = 25 ms; flip angle 65°; 2.5 mm × 2.5 mm in-plane resolution; 2.5 mm slice thickness; 51 slices per volume; SENSE factor 1.8 (anterior–posterior); multiband factor 3. Data was acquired in a single run of 389 dynamic scans. For anatomical reference a structural T1-weighted 3D gradient echo scan was acquired with the following parameters: TR = 10 ms; TE = 4.6 ms; flip angle = 80°; voxel size = 0.75 mm x 0.75 mm x 0.80 mm; parallel imaging factor = 1.70 (AP) and 1.40 (RL).

2.4.3 Preprocessing

Preprocessing and subsequent processing of fMRI scans were done using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>) in MATLAB 2020b (The MathWorks Inc., Massachusetts, United States). The steps described here are identical to those used in previous studies that included YOUth fMRI data (Buimer et al., 2020; Pas et al., 2021). In short, preprocessing consisted of realignment, slice timing correction, spatial normalization to MNI-152 space, and smoothing (8 mm full width at half maximum) to correct for inter-individual differences in functional anatomy.

2.4.4 Individual analyses

Task activity was estimated using a general linear model (GLM) including factors for happy faces, fearful faces, neutral faces, and houses. The six realignment parameters were added to the design matrix to model residual effect of head motion. All data were

high-pass filtered with a cut-off of 128 seconds to remove low-frequency drifts. We used a global signal threshold of 80% to avoid including brain areas with low signal. Participants exhibiting significant signal drops within the brain mask, leading to holes in the mask, were excluded from the analysis (Pas et al., 2021). After the GLM, we defined four contrasts: 1) faces > houses; 2) happy faces > neutral faces; 3) fearful faces > neutral faces; 4) fearful faces > happy faces. The first-level analyses produced four contrast maps for each participant.

2.4.5 Group analyses

In the second-level analyses, task activation maps were thresholded at $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z -value of 3.1 (based on Eklund et al., 2016). The threshold for significance was converted into a voxel size threshold (k) based on the SPM file of each contrast using the SPM Cluster Size Threshold Estimation tool:

https://github.com/CyclotronResearchCentre/SPM_ClusterSizeThreshold

This resulted in a cluster size threshold of 21 voxels for contrast 1 (faces > houses); 27 voxels for contrast 2 and 4 (happy faces > neutral faces and fearful faces > happy faces); 28 voxels for contrast 3 (fearful faces > neutral faces). Because we found widespread and very large clusters (even with our stringent thresholds), we included a watershed procedure to subdivide clusters based on local minima and maxima. The peaks and local minima were used to define borders and to split the cluster into separate segments (Figure 2). Individual contrast maps were masked with the different segments and the average of the β -values for the voxels within a mask were extracted for subsequent analyses.

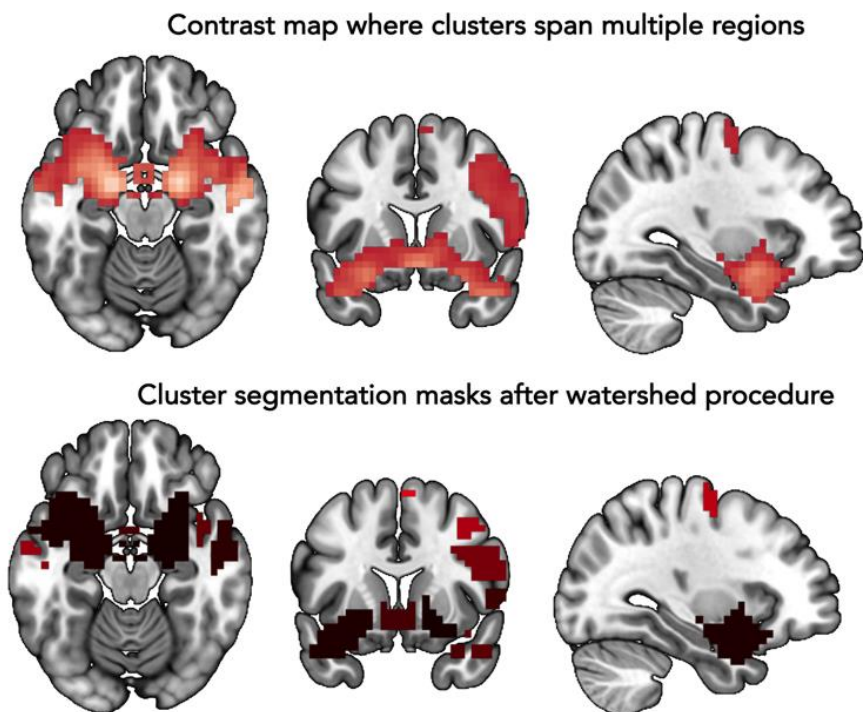


Figure 2 Example of the watershed procedure. As input, we use a contrast map (in this example Faces > Houses) thresholded with $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z-value of 3.1. The watershed procedure then finds local peaks and minima and creates binary masks of the segmentations. The segmentation masks can be used as regions-of-interest in subsequent analyses.

2.5 Statistical analyses

All statistical analyses were conducted in R version 4.0.5 (2021-03-31). The effect of age and sex on social competence was tested with separate linear models for each social competence subscale. The threshold for significance was adjusted to $p < .0125$ based on Bonferroni correction for the four subscales.

For the emotion labeling data, we started with an analysis of variance (ANOVA) to test if the median response time on correct trials was on average different for different types of emotions. Tukey's test was used as post-hoc analysis to test for

differences in group means. Next, we investigated the effect of age and sex on ER skills (accuracy and response time) using separate linear models for the different emotions. The threshold for significance was adjusted to $p < .005$ based on a Bonferroni correction for the 10 analyses (5 contrasts for both accuracy and speed). Additionally, we reran these analyses after adding the 4 social competence subscales using the same Bonferroni corrected threshold for significance.

For the fMRI analyses, the average β -values in individual segments were the dependent variables in linear regression models. As independent variables we started off with only age and sex. Next, we added emotion labeling skills: response time and accuracy for happy emotions as predictors for active subclusters in the happy versus neutral contrast; response time and accuracy for fearful emotions as predictors for active subclusters in the fearful versus neutral contrast; response time and accuracy independent of emotional valence as predictors for the faces versus houses contrast. We controlled the number of false positives by adjusting the p -values over the different subclusters for the false discovery rate (FDR) within each contrast and using a threshold of $p_{FDR} < .05$ (Benjamini & Hochberg, 1995). Lastly, we ran these analyses again with the four social competence subscales as predictors in addition to age and sex, instead of emotion labeling skills, again using FDR -adjusted p -values for determining significance.

2.6 Addressing non-normality with residual-based permutations

In the case when dependent variables were not normally distributed, we ran the linear models as usual to get effect sizes, standard errors and the t-statistic of the variable of interest but determined significance by computing p -values through residual-based permutations (Buzkova, 2016). For each variable of interest, a separate model was fitted leaving this variable out of the equation, which acted as the null model in this analysis. Next, residuals of this null model were used to create new observations with the same sample size as the original sample. First, we computed the fitted values for each observation and added permuted residuals. The effect of the variable of interest (left out in the null-model) was tested in the permuted sample by fitting the full linear model including the variable of interest, providing a

t-statistic for this variable. This procedure was repeated 10.000 times resulting in a distribution of t_i -statistics for the variable of interest. Finally, the p -value was calculated by assessing the probability of the t -statistic of the original model (t_{orig}) given the t_i distribution:

$$p_{permutated}=(1+\text{sum}(\text{abs}(t_i)\geq\text{abs}(t_{orig})))/(10000+1)$$

This procedure is repeated for each variable of interest. Illustrative examples of this procedure can be found in Figure S3 and Figure S4.

3. Results

3.1 Social competence

3.1.1 Variation

For the Strengths and Difficulties Questionnaire (SDQ) scores ranged from 1 to 10 for *prosocial behavior* and 0 to 9 for *peer problems* (theoretical range for both subscales is from 0 to 10). As can be expected in a typically developing sample, the SDQ subscales were skewed towards typical socio-emotional behavior with a mean and standard deviation of 8.49 (1.69) for *prosocial behavior* and 1.07 (1.56) for *peer problems*. For the Interpersonal Reactivity Index (IRI) both subscales (*perspective taking* and *emphatic concern*) ranged from 0 to 28 covering the full range of possible scores. The data was normally distributed with a mean and standard deviation of 14.37 (4.96) for *perspective taking* and 18.44 (4.40) for *emphatic concern*.

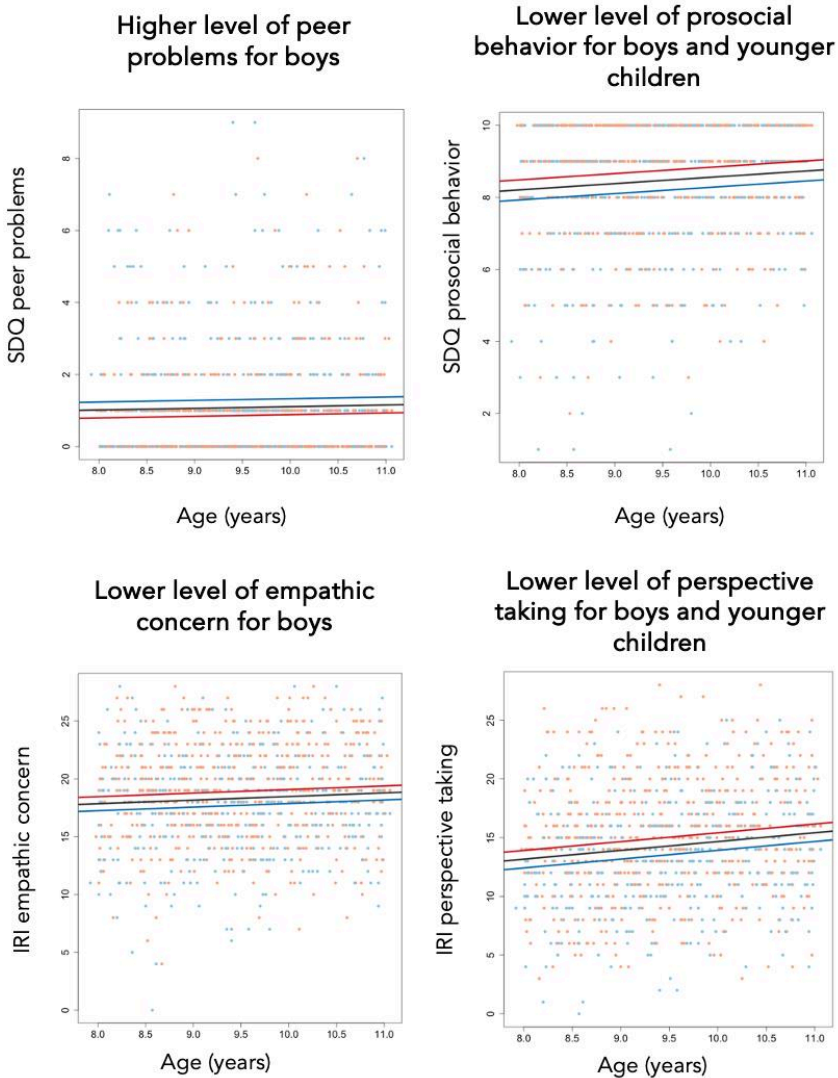


Figure 3. Effects of age and sex on social competence subscales. Red dots indicate social competence scores for girls and blue dots indicate scores for boys. Lines show the relation modeled linearly between social competence score and age (for girls in red, boys in blue and in black for the group as whole). Peer problems and prosocial behavior are subscales from the for the Strengths and Difficulties Questionnaire (SDQ). Empathic concern and perspective taking are subscales from the Interpersonal Reactivity Index (IRI).

3.1.2 The effects of age and sex

Better social competence was found for girls compared to boys for all subscales and age effects were found for two subscales (Figure 3). From the SDQ, *prosocial behavior* increased with age $t(933) = 2.731$, $\beta = 0.175$, standard error = 0.064, $p_{\text{permutated}} = .0066$ and was higher for girls $t(933) = 5.096$, $\beta = 0.557$, standard error = 0.109, $p_{\text{permutated}} = .0001$. No age effects were found for *peer problems* and higher scores were reported for boys compared to girls $t(933) = -4.314$, $\beta = -0.443$, standard error = 0.103, $p_{\text{permutated}} = .0002$. From the IRI, *perspective taking* increased with age $t(933) = 3.983$, $\beta = 0.751$, standard error = 0.188, $p_{\text{permutated}} = .0001$ and was higher for girls $t(933) = 4.645$, $\beta = 1.488$, standard error = 0.320, $p_{\text{permutated}} = .0001$. No age effects were found for *empathic concern* and lower scores were reported for boys compared to girls $t(933) = 4.252$, $\beta = 1.211$, standard error = 0.285, $p_{\text{permutated}} = .0001$. Reported results survived Bonferroni correction based on the four subscales ($p < .0125$).

3.2 Emotion labeling skills

3.2.1 Accuracy and response time

Children were highly accurate when labeling happy, fearful, and neutral facial expressions in all trials, while some angry and sad faces proved more difficult on average (Supplementary figure S1). There was a significant difference in median response time on correct trials between the different emotions, $F(4)=66.93$, $p < .001$. The Tukey post-hoc test revealed that children were faster on correct trials for happy faces compared to each of the other emotions (all $p < .001$), with no statistical differences between the other emotions (Figure 4). The analysis was repeated on log-transformed data for a better approximation of normality, and this did not alter the results. The relation between the median response time on correct trials and the number of correct trials for each child can be found in Supplementary figure S2.

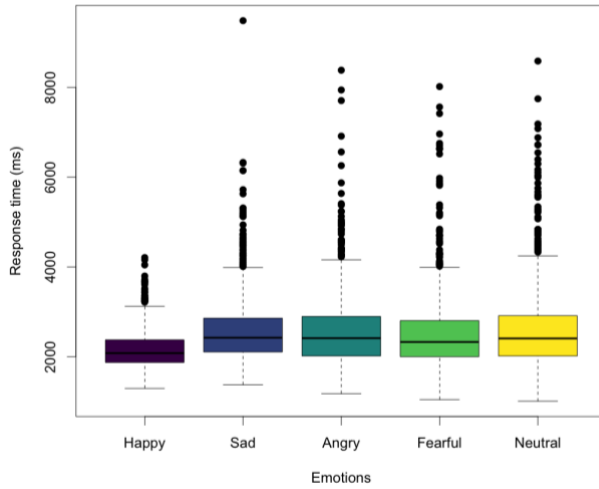


Figure 4. Response time on correct trials for the different emotions. Boxplot of the quartiles of the median response times on correct trials for different facial expressions of emotion. Each color represents a different facial expression stimuli. Black dots are individual data points outside the interquartile range.

3.2.2 The effects of age and sex

As the distribution of the emotion labeling data violated assumptions of normality, we tested for significance with residuals-based permutations (Buzkova, 2016). For more information on this procedure, see the supplementary materials and Figure S3.

In general, older children and girls had an advantage and were significantly more accurate and faster on accurate trials when labeling most of the emotions (Table 1). The accuracy when labeling sad, angry, fearful, and neutral emotions was significantly higher in older children. For happy emotions, we found a trend-level effect of age on accuracy in the same direction, but the effect did not survive Bonferroni correction ($p = .0404$). There was a statistically significant advantage for older children in the reaction time for correctly labeling all emotions. Furthermore, girls were more accurate when labeling happy, sad, and fearful facial expressions, and faster in the correct trials for happy and angry facial expressions. See supplementary table S1 for all the statistics.

Table 1. The effect of age and sex on emotion labeling accuracy and speed. The results of linear models with age and sex as independent variables and number of correct responses or median reaction time in milliseconds based on correct trials only as dependent variables. Each row shows the results of a separate linear model for a specific emotion. The subscript orig indicates that the statistics are computed from the original linear model (β_{orig} , SE_{orig} , t_{orig}) while the subscript permuted indicates that the p -values are computed from the residual-based permutations (p_{permuted}).

	Model	Age				Sex			
	<i>Df</i>	β_{orig}	SE_{orig}	t_{orig}	p_{permuted}	β_{orig}	SE_{orig}	t_{orig}	p_{permuted}
Accuracy									
<i>Happy</i>	1054	0.05	0.02	2.05	.0404	0.12	0.04	3.06	.0023*
<i>Sad</i>	1054	0.18	0.06	3.23	.001*	0.37	0.10	3.80	<.0001*
<i>Angry</i>	1054	0.28	0.05	5.64	<.001*	0.15	0.09	1.73	.0839
<i>Fearful</i>	1054	0.29	0.06	5.12	<.001*	0.32	0.10	3.33	.0011*
<i>Neutral</i>	1054	0.23	0.06	3.98	<.001*	-0.10	0.10	-1.01	.3219
Reaction time									
<i>Happy</i>	1054	-150	14	10.77	<.0001*	-96	24	-4.00	.0002*
<i>Sad</i>	1050	-145	26	-5.62	<.0001*	-97	45	-2.17	.0296
<i>Angry</i>	1051	-181	28	-6.41	<.0001*	-156	49	-3.20	.0022*
<i>Fearful</i>	1054	-213	28	-7.51	<.0001*	-67	49	-1.37	.1684
<i>Neutral</i>	1043	-253	31	-8.11	<.0001*	18	54	0.33	.7465

* = survives Bonferroni correction for the total number of analyses in the table, i.e., $p < .005$.

3.2.3 The relation with social competence

Again, residuals-based permutations were used to determine significance (Buzkova, 2016). See supplementary methods and Figure S4. None of the four subscales of social competence were significant predictors of emotion labeling accuracy or speed for any of the emotions in linear models corrected for age and sex (see Supplementary table S2).

3.3 The neural processing of emotional faces

3.3.1 Task activation

Whole brain analyses showed wide-spread task activation (Figure 5). We found more activation during faces versus houses in the bilateral middle temporal gyrus, bilateral amygdala, left supramarginal gyrus, bilateral precuneus, and left precentral gyrus. Higher activity in houses versus faces was found in the bilateral fusiform gyrus and right superior occipital gyrus. See Table 2 for an overview of activation clusters. Higher activity in happy faces versus neutral faces was found in the bilateral occipital pole, right inferior occipital gyrus, left posterior orbital gyrus, left amygdala, left anterior insula, left middle cingulate gyrus, left middle frontal gyrus, and left superior frontal gyrus (Table 3). Higher activity in fearful faces versus neutral faces was found in the bilateral occipital fusiform gyrus, left inferior occipital gyrus, bilateral entorhinal area, right temporal pole, and bilateral thalamus (Table 4). Happy faces elicited more activation than fearful faces in the bilateral occipital pole, right inferior occipital gyrus, left anterior insula, left putamen, and left middle frontal gyrus (Table 5).

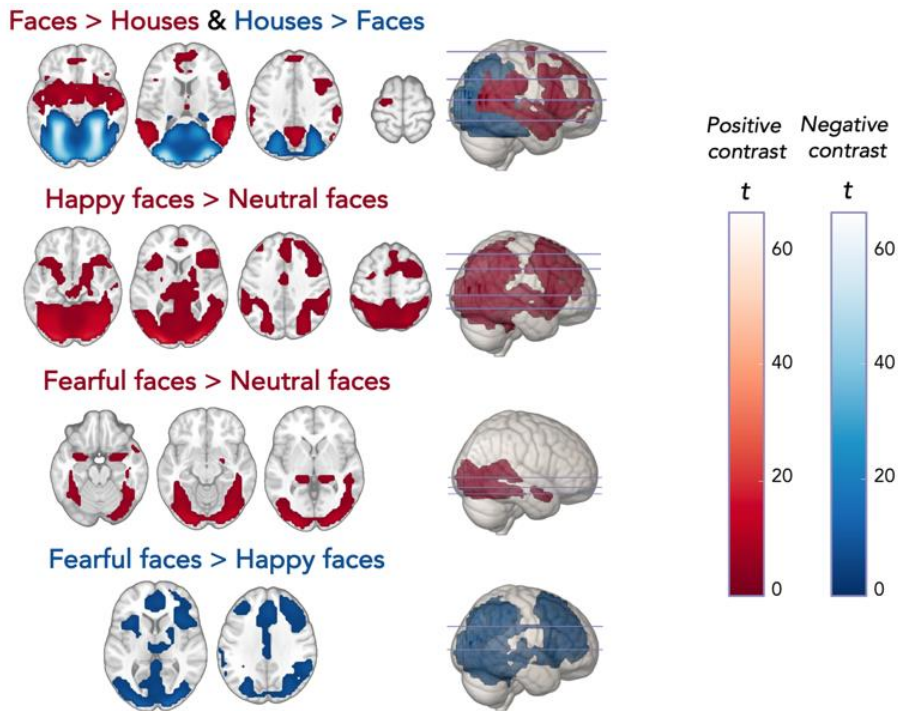


Figure 5. Task activation. Axial slices of the task activity (left hemisphere on the left side). The render on the right shows a transparent overview of the activity in both hemispheres and the location of the axial slices. The activity for each contrast is thresholded at $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z-value of 3.1.

Table 2. Task activation: Faces versus houses. Main clusters (in bold) and subclusters of the Faces > Houses (positive and negative). The X-, Y-, Z-coordinates depict the location of the maximum in MNI space. Cluster size reflects the number of voxels that are part of the main cluster. The activity is thresholded at pFWE < .05 and a cluster extent threshold based on $p < .001$ which corresponds to a z -value of 3.1.

Regions per contrast	Side	Cluster size	X	Y	Z	t_{\max}	Direction
Middle temporal gyrus	R	2422	60	-60	12	20.89	Positive
<i>Amygdala</i>	L		-16	-4	-16	15.95	Positive
<i>Amygdala</i>	R		20	-4	-16	16.71	Positive
Middle temporal gyrus	L	311	-60	-60	12	16.97	Positive
<i>Supramarginal gyrus</i>	L		-64	-24	28	8.49	Positive
Precuneus	L/R	193	4	-56	36	13.16	Positive
Precentral gyrus	L	48	-24	-12	68	6.42	Positive
Fusiform gyrus	R	4679	28	-48	-12	66.34	Negative
<i>Fusiform gyrus</i>	L		-28	-52	-12	62.75	Negative
<i>Superior occipital gyrus</i>	R		32	-88	16	57.80	Negative

Table 3. Task activation: Happy versus Neutral faces. Main clusters (in bold) and subclusters of the Happy > Neutral (positive) contrast. The X-, Y-, Z-coordinates depict the location of the maximum in MNI space. Cluster size reflects the number of voxels that are part of the main cluster. The activity is thresholded at $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z -value of 3.1.

Regions per contrast	Side	Cluster size	X	Y	Z	t_{max}	Direction
<i>Occipital pole</i>	R	7094	20	-96	4	25.18	Positive
<i>Inferior occipital gyrus</i>	R		28	-92	4	24.93	Positive
<i>Occipital pole</i>	L		24	-96	4	23.95	Positive
<i>Posterior orbital gyrus</i>	L	298	-32	20	-12	8.01	Positive
<i>Amygdala</i>	L		-20	-4	-16	7.63	Positive
<i>Anterior insula</i>	L		-32	16	8	7.56	Positive
<i>Middle cingulate gyrus</i>	L	29	0	-8	36	5.45	Positive
<i>Middle frontal gyrus</i>	L	38	-24	0	52	5.25	Positive
<i>Superior frontal gyrus</i>	L		-20	8	68	4.82	Positive
<i>Middle frontal gyrus</i>	L	28	-32	48	24	5.00	Positive

Table 4. Task activation: Fearful versus Neutral faces. Main clusters (in bold) and subclusters of the Fearful > Neutral (positive) contrast. The X-, Y-, Z-coordinates depict the location of the maximum in MNI space. Cluster size reflects the number of voxels that are part of the main cluster. The activity is thresholded at $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z -value of 3.1.

Regions per contrast	Side	Cluster size	X	Y	Z	t_{max}	Direction
<i>Occipital fusiform gyrus</i>	L	1294	-24	-88	-12	15.25	Positive
<i>Inferior occipital gyrus</i>	L		-24	-96	-4	15.16	Positive
<i>Occipital fusiform gyrus</i>	R		24	-88	-12	14.96	Positive
<i>Entorhinal area</i>	R	62	20	0	-20	6.75	Positive
<i>Temporal pole</i>	R		52	12	-24	5.72	Positive
<i>Entorhinal area</i>	L	37	-28	0	-20	6.53	Positive
<i>Thalamus proper</i>	L	37	-12	-32	0	6.28	Positive
<i>Thalamus proper</i>	R	42	20	-32	0	6.26	Positive

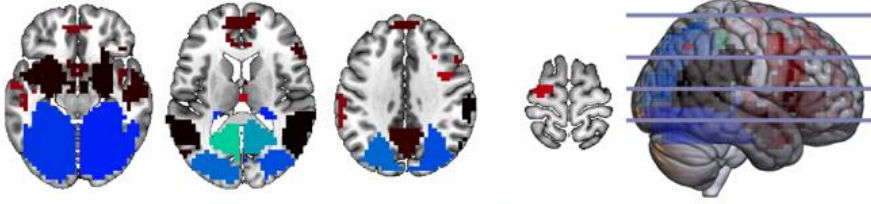
Table 5. Task activation: Fearful versus Happy faces. Main clusters (in bold) and subclusters of the Fearful > Happy (negative). The X-, Y-, Z-coordinates depict the location of the maximum in MNI space. Cluster size reflects the number of voxels that are part of the main cluster. The activity is thresholded at $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z-value of 3.1.

Regions per contrast	Side	Cluster size	X	Y	Z	t_{max}	Direction
<i>Occipital pole</i>	R	6404	20	-96	4	16.16	Negative
<i>Inferior occipital gyrus</i>	R		28	-92	4	14.17	Negative
<i>Occipital pole</i>	L		16	-100	0	13.44	Negative
<i>Anterior insula</i>	L	190	-32	16	4	8.35	Negative
<i>Putamen</i>	L		-24	0	0	5.21	Negative
<i>Middle frontal gyrus</i>	L	152	-40	36	32	7.17	Negative

Inspired by Miller et al., 2016, we additionally visualized the percentage of children passing simple voxel-wise activation thresholding ($t > 1.96$) for each contrast. Despite the widespread and strong activation patterns for all contrasts, only the activation in the bilateral fusiform gyrus extending to the superior occipital gyrus was robust and this cluster was significantly activated in over 50% of the participants for the faces > houses (negative) contrast (Supplementary Figure S5). This suggests that this contrast elicits the most robust brain activation across individuals in our study.

Next, the large clusters were split up in subclusters based on the local peaks of the whole brain activation using a watershed procedure (Figure 6). We ended up with 34 subclusters for faces > houses; 10 subclusters for houses > faces; 49 subclusters for happy faces > neutral faces; 15 subclusters for fearful faces > neutral faces; 50 subclusters for happy faces > fearful faces. These subclusters were then used for subsequent analyses.

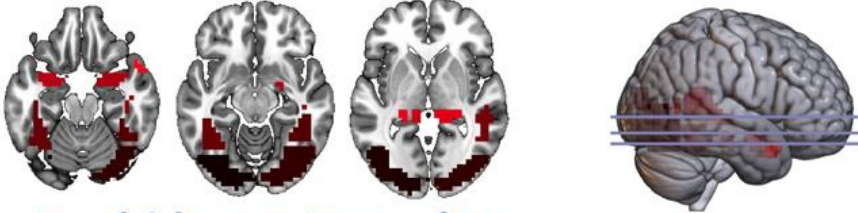
Faces > Houses & Houses > Faces



Happy faces > Neutral faces



Fearful faces > Neutral faces



Fearful faces > Happy faces

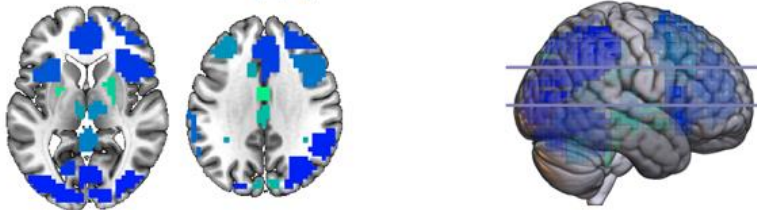


Figure 6. Cluster segmentations after the watershed procedure. Axial slices of the cluster segmentation masks (left hemisphere on the left side) based on the brain activity in each contrast thresholded at $p_{FWE} < .05$ and a cluster extent threshold based on $p < .001$ which corresponds to a z -value of 3.1. The render on the right shows a transparent overview of the segmentations in both hemispheres and the location of the axial slices. Blue masks are based on negative contrast maps and red masks on positive contrast maps.

3.3.2 The effects of age and sex

We did not find age or sex effects to explain individual differences in activation patterns in the contrasts happy versus neutral (positive), fearful versus neutral (positive) and happy versus fearful (negative) (Supplementary tables S3 to S5). We did find a positive correlation between age and brain activity in the faces > houses contrast (i.e., larger contrast in older children; Figure 7, Supplementary table S6). This effect was significant for 4 subclusters extracted from faces > houses (positive) in the left superior temporal gyrus, $t(749) = 3.251$, $p_{FDR} = .0193$, $p_{uncorr} = .0012$, $\beta = 0.038$ (SE = 0.012), the left medial frontal gyrus, $t(749) = 3.415$, $p_{FDR} = .0193$, $p_{uncorr} = .0007$, $\beta = 0.078$ (SE = 0.023), the left planum polare, $t(749) = 3.225$, $p_{FDR} = .0193$, $p_{uncorr} = .0013$, $\beta = 0.051$ (SE = 0.016), and the left superior frontal gyrus (medial segment), $t(749) = 2.823$, $p_{FDR} = .0494$, $p_{uncorr} = .0049$, $\beta = 0.072$ (SE = 0.026). Furthermore, we also found a positive correlation between age and brain activity in 1 subcluster extracted from faces > houses (negative) indicating less deactivation (i.e., higher activation) in older children for faces compared to houses in the bilateral posterior cingulate gyrus in older children, $t(749) = 2.777$, $p_{FDR} = 0.0494$, $p_{uncorr} = .0056$, $\beta = 0.041$ (SE = 0.015). Additionally, we found a larger contrast for faces > houses (positive) for girls in the right supplementary motor cortex, $t(749) = 3.309$, $p_{FDR} = .0431$, $p_{uncorr} = .0010$, $\beta = 0.088$ (SE = 0.026). No age effect was found in this subcluster.

3.3.3 The relation with social competence

None of the four subscales of social competence were significant predictors of brain activity in any of the contrasts in linear models corrected for age and sex (Supplementary tables S10-S12).

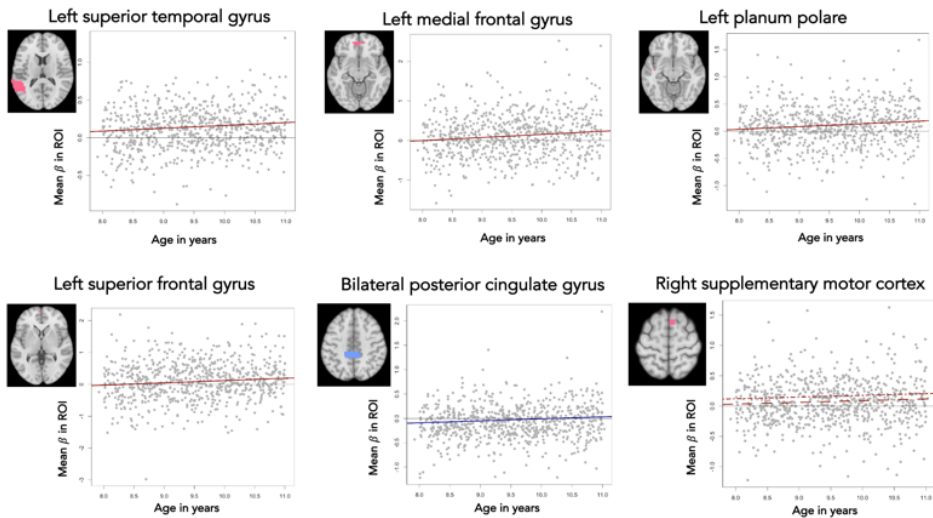


Figure 7. Age and sex effects on activity during faces versus houses. An axial slice shows the location of the subcluster (left hemisphere on the left side). In the plot individual β -weights from the second-level analysis, averaged over the ROI, are plotted against age in years. The solid red lines indicate that the contrast between faces > houses (positive) increases with age as group-average higher activity in faces than houses. The solid blue line indicates a large contrast in younger children with as group-average higher activity in the faces > houses (negative) contrast. The rightmost plot in the bottom row shows the significant effect of sex with a larger contrast for faces > houses (positive) in girls and no age effect in this subcluster. The dot-dash line is based on the girls only and the dashed line is based on the boys only.

3.4 The behavioral labeling and the neural processing of emotional faces

The ability to accurately label happy faces and the response time when correctly labeling happy faces did not significantly relate to brain activity in the happy versus neutral faces contrast (Supplementary table S7). Similarly, emotion labeling skills for fearful faces did not relate to brain activity during the processing of fearful versus neutral faces (Supplementary table S8). As a post-hoc analysis, we wondered if the most robust contrast (faces versus houses), would relate to response time during correct trials and accuracy in the emotion labeling task (independent of emotional valence). Again, we did not find significant associations between brain activity during

faces versus houses and emotion labeling skills at a behavioral level (Supplementary table S9).

4. Discussion

In this study, we tested whether inter-individual variation in behavioral facial emotion labeling and neural facial-emotion processing in pre-adolescence could be explained by age, sex, and social competence. To this end we used data from 1054 children between 8- and 11-years-old participating in the YOUth cohort study. We found effects of age and sex on social competence, ER and neural face versus house processing, but no effects on neural differentiation between emotional expression nor interrelations between the different factors.

4.1 Social competence

We used four subscales to assess social competence (Junge et al., 2020). From the Interpersonal Reactivity Index (Davis, 1983) we used the subscales perspective taking (IRI-pt) and empathic concern (IRI-ec) and from the Strengths and Difficulties Questionnaire (Goodman, 1997, 2001) we used the subscales peer problems (SDQ-pp) and prosocial behavior (SDQ-ps). The subscales were distributed as can be expected in a population-based study. Social competence was higher for girls than boys, consisted with literature (Maurice-Stam et al., 2018; Murris et al., 2003; Overgaauw et al., 2017). Furthermore, IRI-pt and SDQ-ps increased with age. Previous work shows age group differences between early adolescence and late adolescence for the IRI-pt but not the IRI-ec (Hawk et al., 2013). This may suggest that perspective taking (grouped under cognitive empathy) may have a protracted developmental trajectory compared to empathic concern (grouped under affective empathy). For prosocial behavior measures with the SDQ increases with age have also been reported before (Marzocchi et al., 2004).

4.2 Behavioral emotion labeling

We assessed ER using the Penn CNB, a neurocognitive test battery with good validity and reliability (Swagerman et al., 2016). Accuracy reached a ceiling effect in pre-adolescence, especially for happy faces, neutral and fearful faces, leaving less room for inter-individual variation. Still, we show that in pre-adolescence there is an advantage in emotion labeling accuracy and response time for older children compared to younger children. Furthermore, girls were more accurate than boys in labeling happy, sad, and fearful expressions, and faster than boys in correct trials for happy and angry expressions. These associations are consistent with previous work using the Penn ER task a population with a wider age range 8 to 21 (Gur et al., 2012). Furthermore, our findings are in line with age and sex effects found for ER accuracy, reaction time was not included, in 8- to 12-year-old children using the Radboud Faces Database (Verpaalen et al., 2019). It remains unclear if the age- and sex-effects on emotion labeling speed are domain-specific or reflect improvements in general cognitive ability or processing speed (Swagerman et al., 2016).

We did not find an association between ER accuracy or speed and any of the social competence subscales in models corrected for age and sex, contradicting previous studies. Previous work showed negative associations between ER accuracy and the SDQ-pp or total problem scores and positive associations between ER accuracy and SDQ-ps in young children (Burley et al., 2022), children with attention deficit hyperactivity disorder (Staff et al., 2022), children with neurodevelopmental disorders (Löytömäki et al., 2022), children with disruptive behavior (Hunnikin et al., 2020) and adopted children (Paine et al., 2023). For the IRI subscales, previous studies are less consistent. In a study on healthy adults, IRI-ec was negatively associated with accuracy in the Penn ER task (Beals et al., 2022), while in two other studies on adults IRI-ec was positively associated with ER accuracy (Israelashvili et al., 2020). Taken together, previous studies show that in young children, vulnerable populations and atypical developing children ER skills may be predictive of social competence. Here, we show that in typical development in pre-adolescence ER skills do not predict social competence.

4.3 Neural facial emotion processing

The fMRI task resulted in wide-spread activation for the faces > houses (positive and negative) contrast, the happy faces > neutral faces (positive) contrast, the fearful faces > neutral faces (positive) contrast and the fearful faces > happy faces (negative) contrast. Emotion faces elicited more activity than neutral faces and happy faces elicited more activity than fearful faces. When comparing our results to results in a meta-analysis including 105 fMRI studies on the neural processing of emotional faces (Fusar-Poli et al., 2009), we found some differences. The activation in the current study is much more widespread compared to the activation maps in the meta-analysis. There are three possible explanations for this widespread activity. One, the meta-analysis pools data from 1600 participants, but our sample size of 753 participants with fMRI data is roughly 20 times larger than the studies included in the meta-analysis. Large sample sizes can result in large clusters spanning multiple regions (Woo et al., 2014), something we controlled for with stringent statistical thresholding. Two, it is suggested that in this age range brain activity is more diffuse while maturation results in more focal activation patterns (Durston et al., 2006), even though this interpretation has also been criticized (Brown et al., 2006; Poldrack, 2010). Three, we found large inter-individual variation in the first-level contrast maps which may result in widespread second-level activation patterns.

Apart from more widespread activation in our study compared to the meta-analysis, the regions with peak voxels differed between our study and the meta-analysis. For the faces > houses (positive) contrast the activation patterns were most consistent with previous literature. Here, we found peak activation in the bilateral middle temporal gyrus, bilateral amygdala, left supramarginal gyrus, bilateral precuneus and left precentral gyrus. This resembles the activation pattern found when faces independent of emotional valence were contrasted with a fixation cross (Fusar-Poli et al., 2009). Importantly, in this meta-analysis fusiform gyrus activity was found for the same contrast, but in our study bilateral fusiform gyrus activity was higher in response to houses as compared to faces. The different directionality in the fusiform gyrus may be explained by the inclusion of different age categories. Older participants showed greater neural response when processing emotional faces than

younger participants in a meta-analysis (Fusar-Poli et al., 2009). Alternatively, the use of different task stimuli could explain the different results: pictures of faces were contrasted to pictures of houses in the YOUth study (Langer et al., 2010) and pictures of faces were contrasted to a fixation of a crosshair on the screen (Fusar-Poli et al., 2009).

For subsequent analyses we segmented the large clusters in subclusters with the aim of detecting the true signal seeds. Then, we investigated if age, sex, or social competence subscales were associated with the neural processing of emotional faces. Contrary to our expectations, no effects of age and sex were found for emotional versus neutral faces. There are no previous studies of this magnitude in 8-, 9- and 10-year-olds. Potentially, our brain-wide approach prevented us to pick up subtle effects (Marek et al., 2022), especially as fMRI data in general and for this task are only moderately reliable due to the state-dependent nature of brain function and other sources of variations such as noise (Buimer et al., 2020). Another explanation could be that the task design (passive-watching) did not elicit sufficient region-specific brain activity. Still, activity in four subclusters, that were more active during faces compared to houses, was positively associated with age with an increased contrast in older children in the left superior temporal gyrus, the left medial frontal gyrus, the left planum polare, and the medial segment of the left superior frontal gyrus. The activity in one subcluster that was more active during houses compared to faces was positively associated with age with a decreased contrast in older children in the bilateral posterior cingulate gyrus. Sex effects were found in one subcluster faces > houses (positive) with an increased contrast for girls in the right supplementary motor cortex.

We did not find an association between neural processing and any of the social competence subscales. Previous studies did find associations between neural processing and the social competence subscales IRI-ec and IRI-pt, although not always. In adolescents, the IRI-pt was associated with seed-based functional connectivity with a negative association for most regions (Tremblay et al., 2022). Within the default mode network connectivity was positively associated with IRI-ec and IRI-pt in adolescence (Winters et al., 2021). In adults, activity in the bilateral

superior medial frontal cortex (a node within the DMN) was positively associated with the IRI-pt and negatively with the IRI-ec (Oliveira-Silva et al., 2018). A study in adults using a false-belief task found positive associations between the IRI-pt and medial prefrontal cortex activity (False-Belief > False-Photograph), but no effect for the IRI-ec (Dodell-Feder et al., 2014). In young adults, functional brain activity in response to familiar versus unfamiliar faces was not related to the empathic concern subscale of the Interpersonal Reactivity Index (IRI-ec) (Heckendorf et al., 2016). No associations between neural activity during prosocial choices for friends and the IRI-ec or the IRI-pt were found in a study during mid-adolescence (Schreuders et al., 2019). Overall, in one of the largest studies to date in older children between 8 and 11 years of age we show no relation between neural processing of emotional faces contradicting smaller studies with different designs (resting-state functional connectivity or task-based fMRI with different tasks).

4.4 Association between behavioral emotion labeling and neural facial emotion processing

The Penn CNB was developed to capture different neurocognitive domains and its tests were previously validated with functional neuroimaging (Gur et al., 2010, 2012; Roalf et al., 2016). Still, in this study we did not find an association between behavioral ER skills and the neural processing of emotional faces. One possible explanation is that variation in performance on the Penn task does rely on more than emotional processing brain networks, as motor speed, processing speed and cognitive ability play a role as well (Swagerman et al., 2016). In the same way, the passive watching task elicited widespread activity and may not have been able to selectively target the facial emotion processing network. Alternatively, labeling of facial emotions and processing of facial emotions have distinct neural underpinnings. In the same way, the social competence subscales included in this study may tap on different aspects of social behavior that are unrelated to ER or neural facial emotion processing.

4.5 Conclusion

We tested for interrelations between three factors thought to predict social functioning: social competence, facial emotion labeling accuracy and response time and neural processing of emotional faces. In large cohort of typically developing pre-adolescents, we show an advantage for girls and older children consistent with literature for social competence and facial emotion labeling, but no relation between the two factors. Furthermore, we show strong and widespread brain activity in response to faces (happy faces > fearful faces > neutral faces) and houses, but no association between the task contrasts and social competence or behavioral emotion labeling. To conclude, age- and sex-related variation in emotion labeling skills and social competence exists in pre-adolescence, but neural activity in response to faces, behavioral emotion labeling and social competence may reflect separate constructs to navigating social interactions in a typically developing population.

5. Supplementary materials

Some of the supplementary files were too large to be incorporated here and can be downloaded online via this link <http://doi.org/10.17605/OSF.IO/M5R3U>. For Chapter 5, the online supplement includes one file with the supplementary tables S1 to S12 listing all the results from the main analyses.

Reaction times on correct trials versus number of correct response

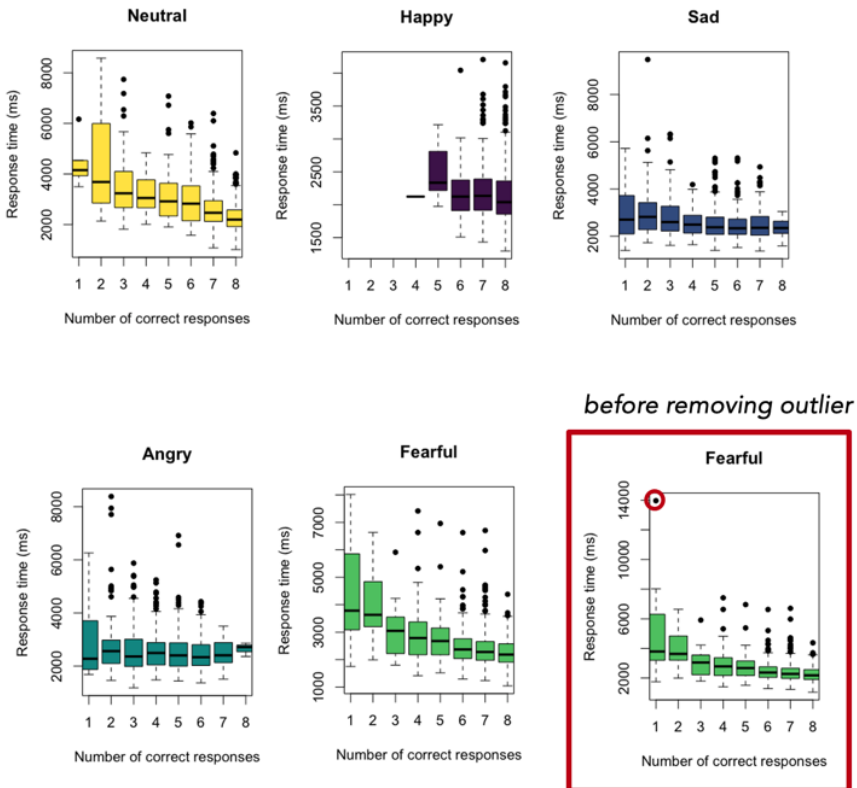


Figure S1. Reaction for correct trials versus number of correct trials. We removed one participant from the final dataset because of the slow response time (14 seconds, i.e., 13 standard deviations from the mean) based on only one correct trial.

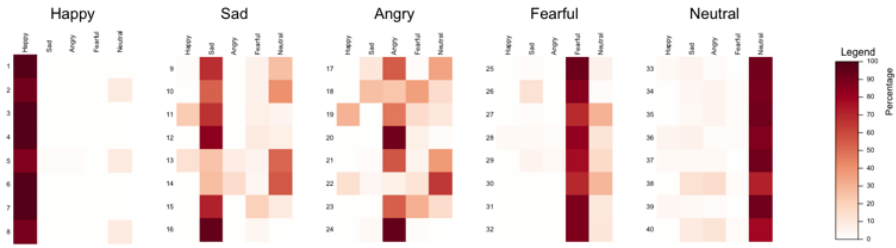


Figure S2. Accuracy for emotion labeling for each trial

Example 1

Significant effects of age and sex on accuracy when labelling fearful faces

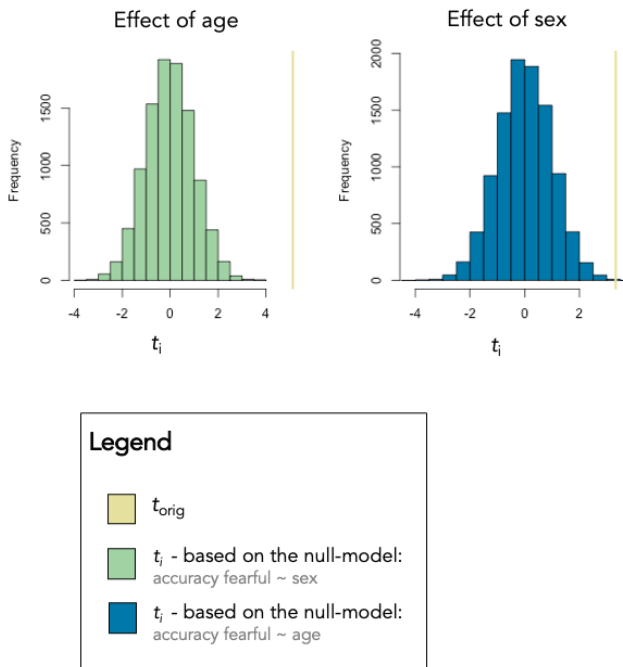


Figure S3. Results of residual-based permutations for the effects of age and sex on accuracy when labelling fearful faces.

Example 2

Non-significant effects of social competence on accuracy when labelling fearful faces

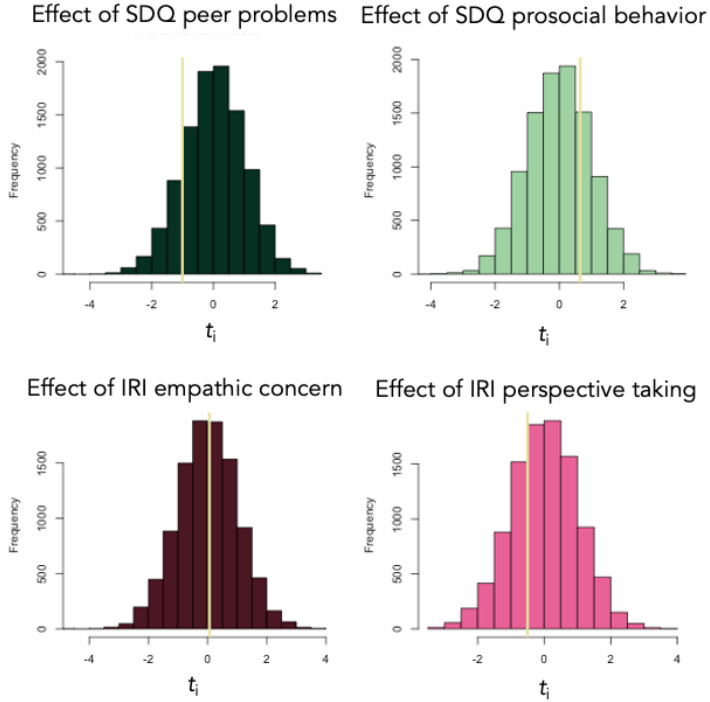


Figure S4. Results of residual-based permutations for the effects of social competence subscales on accuracy when labelling fearful faces.

Percentage of subjects passing
a voxel-wise activation threshold of $t > 1.96$

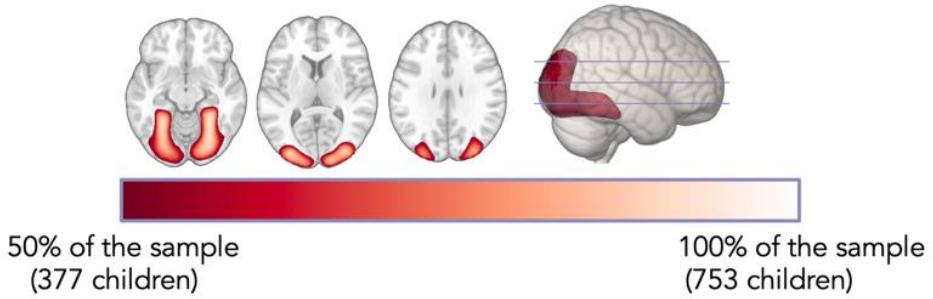


Figure S5. Overlap in first-level activation patterns across individuals for faces > houses (negative). The percentage of children passing simple voxel-wise activation thresholding ($t > 1.96$). Faces > houses (negative) was the only contrast where we found voxels that were significantly activated in at least half of the sample.

Chapter 6 - Summary and conclusion

Summary

The aim of this thesis was to study sources of inter-individual variation in brain structure and function in pre-adolescence. To do so, I used data from 1000 children (8-, 9- and 10-year-olds) participating in the first wave of the YOUth: Child & Adolescent cohort. The first article contains a full description of the YOUth MRI protocol including test-retest reliability of brain measures derived with this protocol (**Chapter 2**). In the second article different de-identification (defacing) procedures for anatomical MRI data are compared (**Chapter 3**). In the third article, I investigated associations between anatomical brain measures and adverse childhood experiences (ACEs) (**Chapter 4**). For the last article, I used functional neuroimaging data to investigate interrelations between different factors related to social cognition: behavioral emotion labeling, neural processing of emotional faces and social competence (**Chapter 5**).

In **Chapter 2** I showed the YOUth MRI protocol and the efforts that were made to collect good quality MR data. The information shared in this chapter, such as the state-of-the-art acquisition parameters, extensive quality control procedures and reliability measures, can be useful to the neuroimaging community by aiding to increase the reproducibility and harmonization across developmental neuroimaging studies. The test-retest study using the YOUth MRI protocol in young adults, indeed showed that test-retest reliability was comparable to literature for brain measures derived from all MRI modalities: structural T1-weighted imaging, diffusion-weighted imaging (DWI), resting-state functional MRI and task-based functional MRI. Global brain measures derived from structural T1-weighted and DWI scans were highly reliable. Resting-state functional connectivity and task-based functional brain measures for both the inhibition task (stop versus go) and the emotion task (face versus house) were moderately reliable.

Chapter 3 answers the question to what extent de-identification methods can introduce noise. De-identification methods remove or blur identifiable facial characteristics on anatomical MRI scans. Implementing de-identification methods enables researchers to share data while protecting participants privacy. I show that de-identification methods introduce some variability in outcome measures, but the reproducibility was comparable to test-retest reliability and no large systematic effects were found. Furthermore, I show that some de-identification methods do not work well on child brain scans. Thus, de-identification methods can be implemented, but some methods require visual quality assurance especially in developmental cohorts.

In **Chapter 4** I show the association between anatomical brain measures and adverse childhood experiences (ACEs). I used a broad definition of adversity resulting in ACEs being highly prevalent: 83% of the children experienced at least one ACE. I show aberrant brain structure in regions of interest for 2 out of 11 ACEs: Substance abuse in the household was associated with larger cortical surface area in frontal regions and exposure to violence was associated with lower fractional anisotropy in the bilateral cingulum bundle in the hippocampus region. Exposure to violence could be an indicator of neighborhood disadvantage as it is made up of parental reports on *theft* and *robbery* (high prevalence) and *rape*, *abuse*, and *domestic violence* (low prevalence). Figure 1 shows a summary of the main findings in this chapter.

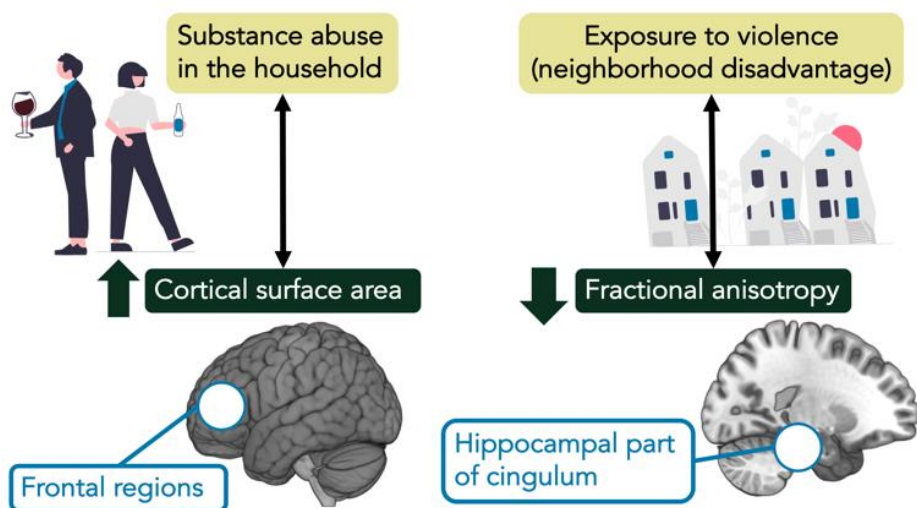


Figure 1. Summary of main findings in Chapter 4.

In **Chapter 5** I tested whether three proxies of social functioning were interrelated: social competence, behavioral emotion labeling and neural facial emotion processing. Surprisingly, I did not find evidence for interrelations between these three factors. Still, interesting effects of age and sex were found. Emotion labeling skills were better in girls than in boys and in older than in younger children. Furthermore, higher social competence on all subscales was reported for girls compared to boys and higher prosocial behavior and perspective taking was reported in older children. Neural processing of facial emotions resulted in strong and widespread activity. The strongest brain activity was found in response to happy faces (more than the brain response to fearful or neutral faces).

General discussion

Reliability of brain measures

From 2014 to 2016 a healthy researcher was scanned over 100 times (Poldrack et al., 2015). With this sensational experiment he showed that even in the same subject,

you will never get the same brain measures. Even though, researchers have reported test-retest reliability of brain-derived measures for a long time, it was never so clear how dynamic brain function is. Apart from variation related to the scanner or movement, for functional connectivity we now differentiate between “State” and “Trait” components (Geerligs et al., 2015). In **Chapter 2**, I show that functional brain measures have lower test-retest reliability than structural brain measures (Buimer et al., 2020). This is explained by the dynamic nature of fMRI rather than the quality of the acquisition, as the reliability reported in this study is comparable or even on the higher end of reliability reported in previous studies (Bennett & Miller, 2010). Transparency about test-retest reliability and related difficulties with cross-study replication is important as society more and more asks for individual predictions from neuroimaging data. It is suggested that thousands individuals are needed to detect a reproducible association between the brain and cognitive or mental health phenotypes (Marek et al., 2022). In the supplementary materials of **Chapter 2**, I did my own power analysis and showed that an effect with an effect size of 0.1 can be reliably detected with regional T1-weighted brain measures in 1000 individuals, while fMRI-based ROI-measures need a much larger sample size to detect such an effect (Buimer et al., 2020). When studying small effects, there are two paths to reliability (Gratton et al., 2022): ensure that high quality data is collected or increase the sample size. The latter is facilitated by data sharing and harmonizing data collection across the globe. In my opinion, the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium is a perfect example of the power of world-wide multicenter collaboration to unravel the genetic and brain correlates of various phenotypes through data pooling (Thompson et al., 2014).

Open science

I strongly believe that complex scientific questions can only be answered by sharing data, publications, methods etc. *Open science* is an umbrella term for changes in technological architecture, access to knowledge, accessibility of knowledge creation, impact measurement and collaborative research (Fecher & Friesike, 2014). Ultimately, open science as movement aims to improve integrity, equity, collaboration and impact (Figure 2). In **Chapter 2**, I describe the YOUTH MRI

protocol and test-retest reliability of brain measures derived with this protocol (Buimer et al., 2020). The protocol and quality control measures are described in detail, so the methods can be replicated by others. Furthermore, as YOUth encourages and facilitates extensive and appropriate use of its data by bona fide researchers, this chapter increases transparency for researchers wanting to request data. Different aspects should be weighted when sharing data (Veldkamp & Kemner, 2023). It is crucial to safeguard the connection between science and society and the trust volunteers have in researchers. As facial identification from anatomical MRI data is possible nowadays (Schwarz et al., 2019), de-identification procedures are necessary. Luckily, de-identification methods make data sharing safer (Schwarz et al., 2021). In **Chapter 3**, I show that de-identification methods minimally affect subsequent processing (Buimer et al., 2021) in line with another study (Schwarz et al., 2021). The different interacting factors that play a role in development ask for interdisciplinary collaboration and *team science* (Zanolie et al., 2022). Still, I show that this does not have to come at the cost of the privacy of participants involved in the research.

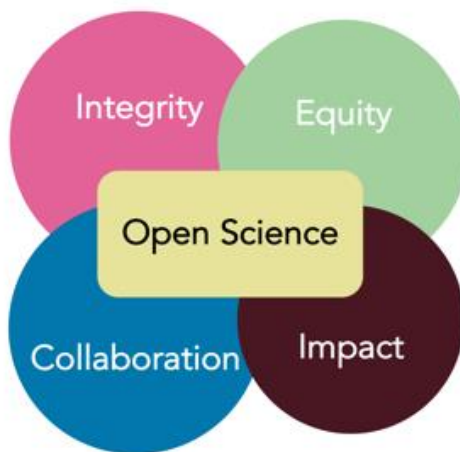


Figure 2 Adapted from an image by *WIM ontwerpers* for the TU Delft.

Late childhood

Currently, policy makers prioritize the first 1000 days of life to promote positive development. The emphasize on this period of development stems from the rapid brain growth during this time (Dubois et al., 2021) and the long-lasting effects of severe adverse factors, such as malnutrition, in this period (Schwarzenberg et al., 2018). Still, developmental neuroscience studies in the past decade underlined the importance to look past this period (Graf et al., 2021). The brain undergoes significant maturation during (pre-)adolescence (Mills et al., 2021; Teeuw et al., 2019; Frangou et al., 2022; Giedd et al., 1999; Koenis et al., 2015; Tamnes et al., 2017). Region-specific periods of heightened plasticity, provide opportunities as well as vulnerabilities (Fuhrmann et al., 2015). As an example, the effect of household dysfunction on outcomes later in life is much higher during mid-childhood and early adolescence compared to children exposed to this factor earlier in life (Andersen, 2021). The heightened neuroplasticity during adolescence co-exists with the emergence of mental health problems (Paus et al., 2008), again underlining the importance of studying brain development in late childhood and adolescence.

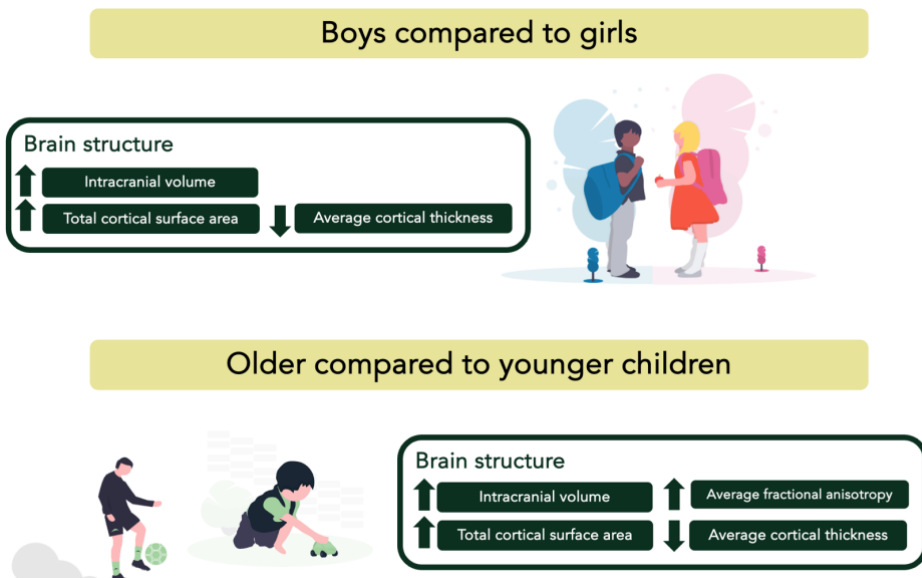


Figure 3. Summary of age- and sex-effects in Chapter 4.

The transition from childhood to adulthood may be especially important for social development (Blakemore, 2012). The children in our cohort are on the verge of big changes. They are preparing for high school, while their levels of reproductive hormones are rapidly increasing (Koenis et al., 2013) and some of them (mostly girls) are starting puberty (Pas et al., 2021). A previous study in the same cohort showed that despite the narrow age-range (8-11 years old), there is a positive association between behavioral inhibition and age (Pas et al., 2021). In my PhD thesis, I also find robust age effects. I find age- and sex-effects for emotion labeling and social competence subscales. Furthermore, the reported age- and sex-effects on brain structure underline the importance of this period of development. The age- and sex-effects found in this study (summarized in Figure 3 and 4) provide fundamental information on the stage of (brain) development in pre-adolescence and complements previous research in larger age ranges or with smaller sample sizes.

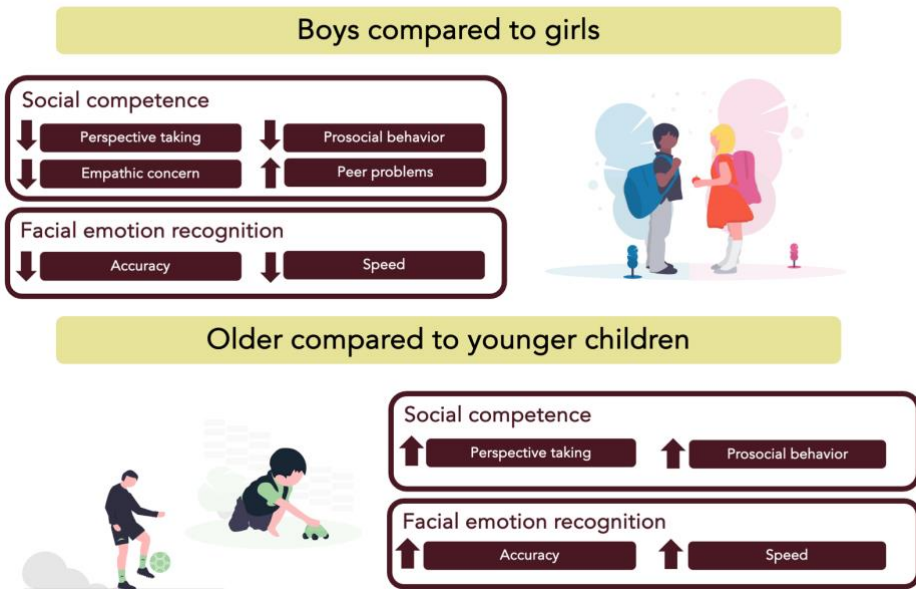
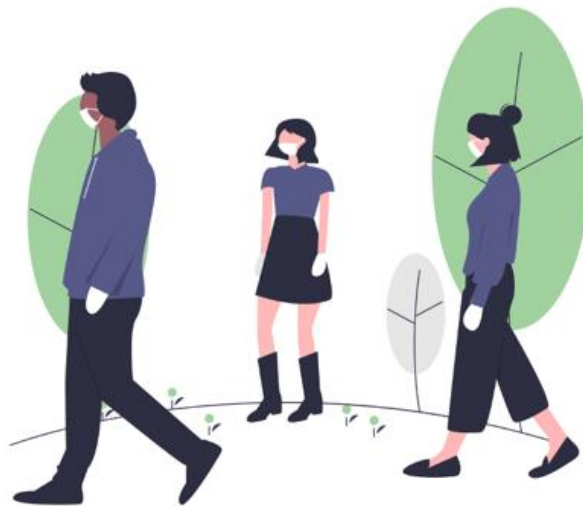


Figure 4. Summary of age- and sex-effects in Chapter 5.

While the YOUth cohort study collected data throughout the pandemic and thereafter, all data included in this thesis was collected prior to the COVID-19 pandemic. Therefore, the potential effect of the COVID-19 pandemic on social development in young people (Crone & Achterberg, 2022) was not assessed in this thesis. For an overview report of the effects of the pandemic and related restrictions on young people, see the report by The Netherlands Youth Institute in collaboration with a large group of scientists, stakeholders, organizations, and youth (van den Berg et al., 2023). In short, the report suggests that on average quality of life decreased during the pandemic and emotional problems increased. Friendship support diminished the negative effects. School closure impacted motivation to learn and resulted in feelings of loneliness as children missed their classmates. Furthermore, tension at home increased. Some individuals may be disproportionately affected by the pandemic and related restrictions (Pierce et al., 2020), such as children from racial/ethnic minorities (Morales et al., 2020), children from families dealing with socioeconomic hardship (Green et al., 2023), children with pre-existing mental health vulnerability (Zijlmans et al., 2022) and children with ACEs (Stinson et al., 2021).



Adverse childhood experiences

Almost half of the Dutch children between 9 and 13 years of age report adverse childhood experiences and these experiences are associated with lower quality of life in childhood (Vink et al., 2019). In adulthood, adverse childhood experiences are associated with mental health and other problems (Felitti et al., 1998; Hughes et al., 2017; Kalmakis & Chandler, 2015; Merrick et al., 2019). I show aberrant brain structure for two adverse childhood experiences. Follow-up data is needed to establish long-term effects and compare whether the impact of ACEs differs across windows of development. The effect of ACEs may differ depending on the developmental timing of the experience. Importantly, the timing (opening and closing) of sensitive periods of brain development can also change in response to environmental factors (Gabard-Durnam et al., 2020).

Facial emotion processing

Aberrant facial emotion processing is associated with a wide range of mental health problems (Collin et al., 2013) and more prevalent in children exposed to childhood adversity (Assed et al., 2020; Bérubé et al., 2023). Therefore, facial emotion processing is an interesting concept for research on the relation between childhood adversity and mental health problems. Furthermore, facial emotion processing is important for meaningful social connections. In **Chapter 5** we show that seemingly overlapping concepts are not interrelated as expected. Emotion labeling, facial emotion processing and social competence were not associated with each other. This is a noteworthy finding as theoretical frameworks often rely on these concepts being naturally linked together. Still, we show that emotion labeling and social competence both improve with age. These findings in typically developing children provide starting points for future research in children with externalizing or internalizing symptoms or children exposed to ACEs.

Methodological considerations and future directions

What are the benefits of developmental cohort studies?

Cohort studies are a powerful research design. One, cohort data often have large sample sizes, which is important as effects of interest have small effect sizes in developmental cognitive neuroscience (Dick et al., 2021). Two, the longitudinal design in cohort studies allows for investigation of both inter-individual and intra-individual variation, which is particularly important when studying mechanisms underlying dynamic processes such as brain development (Crone & Elzinga, 2015). Three, cohort studies provide a wealth of data that better capture individual complexity.

Why do we need representative samples?

In my opinion, the largest challenge for cohort studies is including a representative group. Behavioral studies (Henrich et al., 2010) and neuroscience studies (Chiao & Cheon, 2010) are dominated by western, educated, industrialized, rich and democratic (WEIRD) people. Furthermore, historically marginalized groups (i.e., Black, Indigenous, and People of Color (BIPOC), individuals with bi-or multi-cultural origin, individuals from low socioeconomic status, individuals from the LHBQTQIA+ community, and individuals with disabilities or functional impairment) are underrepresented due to exclusion bias, sampling bias or attrition bias (Green et al., 2022). This is problematic for multiple reasons. Results from research funded with public money apply only to a selective group. During the COVID-19 pandemic, policies world-wide were partly informed by mental health surveys that were not representative and thus may have failed to capture the severity of the crisis (Pierce et al., 2020). Furthermore, my own work in **Chapter 4** shows that adverse childhood experiences can be less prevalent in cohorts than in the society, thereby reducing the power to detect an effect.

This raises the question whether study designs in which characteristics of interest (experiencing adversity, experiencing racism, psychopathology, atypical behavior etc.) are specifically targeted are more fitting when investigating mechanisms of risk

and resilience. Representative sampling may enhance our understanding of fundamental neural processes, as exemplified by earlier brain maturation, and more reliable developmental trajectories in a sample weighted for socioeconomic status, race/ethnicity, and sex compared to the initial sample (LeWinn et al., 2017). In the YOUth cohort study, as in other cohorts, an effort has been put in to recruit a representative sample by specifically targeting less represented children. Still, children with a high socio-economic status are overrepresented (Fakkell et al., 2020). Scientists need to make an ongoing effort to establish a sustainable connection with society (Green et al., 2022). Science communication, co-creation, transparency, and diversity are key in making science more inclusive.

What is normative?

Within developmental neuroimaging and throughout my dissertation, you may stumble across words like *typical development*, *atypical development*, *optimal development*, *normative development*, *normative environments*, and *adverse environments*. Some of these words may hold a value judgment, for example the use of *normative development* is criticized (Poldrack, 2010). Poldrack (2010) argues that adult-like brain activity or structure is taken as the norm while developmental neuroimaging should be descriptive rather than normative. Furthermore, although brain charts of typical development are very useful when investigating psychopathology (Bethlehem et al., 2022; Brouwer et al., 2022; Tamnes et al., 2017), it is important to emphasize that there is a large heterogeneity in individual developmental trajectories (Becht et al., 2020, 2021; Drakulich et al., 2021; Foulkes & Blakemore, 2018; Mills et al., 2021). Lastly, what is defined as a normative or adverse environment depends on socio-cultural norms, although there are also universal agreements on basic needs during childhood (Kolhatkar & Berkowitz, 2014; Raman & Hodes, 2012).

Can we study environmental factors without genetics?

This thesis focuses on associations between the brain and environmental factors. Still, these associations may as well be driven by genetic factors or gene-environment

interactions. Genetic factors are known to impact all outcome measures in this dissertation, thus anatomical brain measures (Brouwer et al., 2010, 2012; den Braber et al., 2013; Hulshoff Pol et al., 2012; Koenis et al., 2015; Lenroot et al., 2009; Lenroot & Giedd, 2008; Panizzon et al., 2009; Peper et al., 2009; Schmitt et al., 2014; van Soelen et al., 2011), task-based functional MRI measures (Achterberg et al., 2018; Blokland et al., 2008) and emotion labeling skills (Swagerman et al., 2016). In the same way, the predictors used in this dissertation are also influenced by genetic factors and therefore gene-environment interactions are at play. The environment that parents can provide for their children is influenced by genetics (Hart et al., 2021; Kong et al., 2018). Even more, childhood adversity is associated with genetic predispositions (Warrier et al., 2021). Lastly, social competence is partly heritable (St Pourcain et al., 2015; Knafo-Noam et al., 2015; Warrier et al., 2018; van der Meulen et al., 2020). Adding genetic information and longitudinal data in future research designs helps to interpret the main effects in terms of mechanisms or even directionality of the effects. Still, gene-environment interactions are difficult to map.

Capturing complexity

The more we learn about the human brain, the more complex it turns out to be. Traditional approaches to study the brain focused on the idea that individual brain regions support cognition and behavior, while the brain may be best represented as a multi-scale network (e.g., Betzel & Bassett, 2017; Friston, 2002; McIntosh, 2000). This shift towards more complex methods and models is not unique to developmental cognitive neuroscience. For example, research in biological psychiatry was long focused on discovering single genes responsible for psychiatric disorders, while the pathogenesis of psychiatric disorders is much more complex and orchestrated by many genetic and environmental factors. The phenotype side of the equation also turned out to be more complex than previously thought. For example, the symptomatic overlap between mental health diagnoses is large.

Regarding conceptualizing childhood adversity, I show that different experiences have a differential effect on brain structure and that simply summing them does not suffice. Traditionally, ACEs have been lumped together while more recently

dimensional models, such as the dimensional model of adversity and psychopathology (McLaughlin & Sheridan, 2016), are used to study differential effects of specific ACEs or dimensions of ACEs on outcomes late in life. Others point out problems with these so-called specificity models and suggest a topological approach focused on factors that influence the way events are experienced by the child rather than the type or category of the event (Smith & Pollak 2021).

To be able to detect the weak associations between brain and phenotype, neuroimaging studies benefit from increasing sample sizes and signal-to-noise ratios as described earlier. Still, the variance explained remains small. One approach to battle this is to integrate brain measures (atlas-based or voxel-/vertex-wise) using multivariate techniques, e.g. non-negative matrix factorization (Anderson et al., 2014), independent component analysis, canonical correlation analysis or partial least squares approaches (Sui et al., 2012). Principal component analysis was used in a study on the effects of brain structure and adverse lifetime experiences in adults (Gheorghe et al., 2021). Therefore, an important future direction will be to use advanced computational modelling to integrate individual data from different timepoints and from different modalities (structural neuroimaging, functional neuroimaging, health information, surveys, genetics). Machine learning approaches are a promising avenue to make individual predictions and to overcome the observed heterogeneity in psychiatry (Schnack, 2019) as well as developmental cognitive neuroscience (Fair et al., 2021; Scheinost et al., 2023). Integrating multimodal neuroimaging data can help to increase the predictive value of machine learning algorithms (Tulay et al., 2018).

Shift to resilience

In recent years a lot of progress has been made in the field of resilience research. While resilience was long defined as a trait, current resilience frameworks focus on resilience as a dynamic and partly malleable construct (Ioannidis et al., 2020; Kalisch et al., 2019). Multisystem models of resilience argue that resilience arises from complex interrelations across cultural, social, psychological and neurobiological systems (Masten & Cichetti, 2010; Masten et al., 2021). New studies focus on the

neurobiology of resilient functioning (Ioannidis et al., 2020; Moreno-López et al., 2020; González-García et al., 2023). Furthermore, more studies now confirm the importance of social support as a buffer between the effects of childhood adversity on mental health later in life (van Harmelen et al., 2017, 2021). This work will in the future provide clear guidance to better support children growing up in adverse environments and adolescents suffering from mental health problems related to childhood adversity (Cicchetti & Toth, 2015).

Concluding remarks

The aim of this thesis was to study sources of inter-individual variation in brain structure and function in pre-adolescence. This thesis focused on several factors that contribute to this variation, including test-retest variability, biological factors such as age and sex, socioemotional skills, and childhood experiences. In 8-, 9- and 10-year-olds participating in the YOUTH cohort study (N>1000), I showed age- and sex-effects for social competence, emotion labeling and global anatomical brain measures. Furthermore, I found evidence for associations between two adverse childhood experiences and local brain structure. Overall, associations between inter-individual differences in brain characteristics and behavioral or environmental phenotypes were low, despite the relatively large sample size. The inherently complex associations between the social environment and the developing brain ask for interdisciplinary collaborations, embracing diversity in all aspects, and pooling data to enhance statistical power. Understanding how childhood experiences convey risk for mental health problems later in life can help to identify malleable factors that could ultimately support children growing up.

Chapter 7 - References

A

- Abramian, D., & Eklund, A. (2018). Refacing: Reconstructing Anonymized Facial Features Using Gans [Preprint]. *Neuroscience*. <https://doi.org/10.1101/447102>
- Achterberg, M., Dobbelaar, S., Boer, O. D., & Crone, E. A. (2021). Perceived stress as mediator for longitudinal effects of the COVID-19 lockdown on wellbeing of parents and children. *Scientific Reports*, 11(1), 2971. <https://doi.org/10.1038/s41598-021-81720-8>
- Achterberg, M., & van der Meulen, M. (2019). Genetic and environmental influences on MRI scan quantity and quality. *Developmental Cognitive Neuroscience*, 38, 100667. <https://doi.org/10.1016/j.dcn.2019.100667>
- Achterberg, M., van Duijvenvoorde, A. C. K., van der Meulen, M., Bakermans-Kranenburg, M. J., & Crone, E. A. (2018). Heritability of aggression following social evaluation in middle childhood: An fMRI study. *Human Brain Mapping*, 39(7), 2828–2841. <https://doi.org/10.1002/hbm.24043>
- Adhikari, B. M., Jahanshad, N., Shukla, D., Glahn, D. C., Blangero, J., Fox, P. T., Reynolds, R. C., Cox, R. W., Fieremans, E., Veraart, J., Novikov, D. S., Nichols, T. E., Hong, L. E., Thompson, P. M., & Kochunov, P. (2018). Comparison of heritability estimates on resting state fMRI connectivity phenotypes using the ENIGMA analysis pipeline. *Human Brain Mapping*, 39(12), 4893–4902. <https://doi.org/10.1002/hbm.24331>
- Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, 12(2), 169–177. [https://doi.org/10.1016/S0959-4388\(02\)00301-X](https://doi.org/10.1016/S0959-4388(02)00301-X)
- 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: Motion Bias in Analyses of Structural MRI. *Human Brain Mapping*, 37(7), 2385–2397. <https://doi.org/10.1002/hbm.23180>
- Alfaro-Almagro, F., Jenkinson, M., Bangerter, N. K., Andersson, J. L. R., Griffanti, L., Douaud, G., Sotiropoulos, S. N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D. C., Zhang, H., Dragonu, I., Matthews, P. M., ... Smith, S. M. (2018). Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*, 166, 400–424. <https://doi.org/10.1016/j.neuroimage.2017.10.034>
- Ancelin, M.-L., Carrière, I., Artero, S., Maller, J. J., Meslin, C., Dupuy, A.-M., Ritchie, K., Ryan, J., & Chaudieu, I. (2021). Structural brain alterations in older adults exposed to early-life adversity. *Psychoneuroendocrinology*, 129, 105272. <https://doi.org/10.1016/j.psyneuen.2021.105272>

- Andersen, S. L., Tomada, A., Vincow, E. S., Valente, E., Polcari, A., & Teicher, M. H. (2008). Preliminary Evidence for Sensitive Periods in the Effect of Childhood Sexual Abuse on Regional Brain Development. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 20(3), 292–301. <https://doi.org/10.1176/jnp.2008.20.3.292>
- Andersen, S. H. (2021). Association of youth age at exposure to household dysfunction with outcomes in early adulthood. *JAMA Network Open*, 4(1), e2032769-e2032769.
- Anderson, A., Douglas, P. K., Kerr, W. T., Haynes, V. S., Yuille, A. L., Xie, J., Wu, Y. N., Brown, J. A., & Cohen, M. S. (2014). Non-negative matrix factorization of multimodal MRI, fMRI and phenotypic data reveals differential changes in default mode subnetworks in ADHD. *NeuroImage*, 102, 207–219. <https://doi.org/10.1016/j.neuroimage.2013.12.015>
- Andersson, J. L. R., Graham, M. S., Zsoldos, E., & Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *NeuroImage*, 141, 556–572. <https://doi.org/10.1016/j.neuroimage.2016.06.058>
- Andersson, J. L. R., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: Application to diffusion tensor imaging. *NeuroImage*, 20(2), 870–888. [https://doi.org/10.1016/S1053-8119\(03\)00336-7](https://doi.org/10.1016/S1053-8119(03)00336-7)
- Andersson, J. L. R., & Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage*, 125, 1063–1078. <https://doi.org/10.1016/j.neuroimage.2015.10.019>
- Ansell, E. B., Rando, K., Tuit, K., Guarnaccia, J., & Sinha, R. (2012). Cumulative Adversity and Smaller Gray Matter Volume in Medial Prefrontal, Anterior Cingulate, and Insula Regions. *Biological Psychiatry*, 72(1), 57–64. <https://doi.org/10.1016/j.biopsych.2011.11.022>
- Assed, M. M., Khafif, T. C., Belizario, G. O., Fatorelli, R., Rocca, C. C. de A., & de Pádua Serafim, A. (2020). Facial Emotion Recognition in Maltreated Children: A Systematic Review. *Journal of Child and Family Studies*, 29(5), 1493–1509. <https://doi.org/10.1007/s10826-019-01636-w>

B

- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1), 1-68.
- Bartko, J. J., & Carpenter, W. T. (1976). On the methods and theory of reliability. *The Journal of Nervous and Mental Disease*, 163(5), 307–317. <https://doi.org/10.1097/00005053-197611000-00003>
- Bastiani, M., Cottaar, M., Fitzgibbon, S. P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S. N., Jbabdi, S., & Andersson, J. L. R. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *NeuroImage*, 184, 801–812. <https://doi.org/10.1016/j.neuroimage.2018.09.073>

- Bayet, L., & Nelson, C. A. (2019). The Perception of Facial Emotion in Typical and Atypical Development. In V. LoBue, K. Pérez-Edgar, & K. A. Buss (Eds.), *Handbook of Emotional Development* (pp. 105–138). Springer International Publishing. https://doi.org/10.1007/978-3-030-17332-6_6
- Beals, K., Sperry, S. H., & Sheffield, J. M. (2022). Empathy, Emotion Recognition, and Paranoia in the General Population. *Frontiers in Psychology*, 13, 804178. <https://doi.org/10.3389/fpsyg.2022.804178>
- Becht, A. I., Klapwijk, E. T., Wierenga, L. M., van der Cruijssen, R., Spaans, J., van der Aar, L., Peters, S., Branje, S., Meeus, W., & Crone, E. A. (2020). Longitudinal associations between structural prefrontal cortex and nucleus accumbens development and daily identity formation processes across adolescence. *Developmental Cognitive Neuroscience*, 46, 100880. <https://doi.org/10.1016/j.dcn.2020.100880>
- Becht, A. I., Wierenga, L. M., Mills, K. L., Meuwese, R., van Duijvenvoorde, A., Blakemore, S.-J., Güroğlu, B., & Crone, E. A. (2021). Beyond the average brain: Individual differences in social brain development are associated with friendship quality. *Social Cognitive and Affective Neuroscience*, 16(3), 292–301. <https://doi.org/10.1093/scan/nsaa166>
- Behzadi, Y., Restom, K., Liao, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>
- Bell, K. L., Purcell, J. B., Harnett, N. G., Goodman, A. M., Mrug, S., Schuster, M. A., Elliott, M. N., Emery, S. T., & Knight, D. C. (2021). White Matter Microstructure in the Young Adult Brain Varies with Neighborhood Disadvantage in Adolescence. *Neuroscience*, 466, 162–172. <https://doi.org/10.1016/j.neuroscience.2021.05.012>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences*, 1191(1), 133–155. <https://doi.org/10.1111/j.1749-6632.2010.05446.x>
- Bérubé, A., Turgeon, J., Blais, C., & Fiset, D. (2023). Emotion Recognition in Adults With a History of Childhood Maltreatment: A Systematic Review. *Trauma, Violence, & Abuse*, 24(1), 278–294. <https://doi.org/10.1177/15248380211029403>
- 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*, 604(7906), 525–533. <https://doi.org/10.1038/s41586-022-04554-y>
- Betzal, R. F., & Bassett, D. S. (2017). Multi-scale brain networks. *Neuroimage*, 160, 73–83.

- Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., & Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, 83, 550–558. <https://doi.org/10.1016/j.neuroimage.2013.05.099>
- Bischoff-Grethe, A., Ozyurt, I. B., Busa, E., Quinn, B. T., Fennema-Notestine, C., Clark, C. P., Morris, S., Bondi, M. W., Jernigan, T. L., Dale, A. M., Brown, G. G., & Fischl, B. (2007). A technique for the deidentification of structural brain MR images. *Human Brain Mapping*, 28(9), 892–903. <https://doi.org/10.1002/hbm.20312>
- Bjork, J. M., Straub, L. K., Provost, R. G., & Neale, M. C. (2017). The ABCD Study of Neurodevelopment: Identifying Neurocircuit Targets for Prevention and Treatment of Adolescent Substance Abuse. *Current Treatment Options in Psychiatry*, 4(2), 196–209. <https://doi.org/10.1007/s40501-017-0108-y>
- Blakemore, S.-J., Dahl, R. E., Frith, U., & Pine, D. S. (2011). Developmental cognitive neuroscience. *Developmental Cognitive Neuroscience*, 1(1), 3–6. <https://doi.org/10.1016/j.dcn.2010.08.003>
- Blakemore, S. J. (2012). Development of the social brain in adolescence. *Journal of the Royal Society of Medicine*, 105(3), 111–116.
- Blokland, G. A. M., McMahon, K. L., Hoffman, J., Zhu, G., Meredith, M., Martin, N. G., Thompson, P. M., de Zubicaray, G. I., & Wright, M. J. (2008). Quantifying the heritability of task-related brain activation and performance during the N-back working memory task: A twin fMRI study. *Biological Psychology*, 79(1), 70–79. <https://doi.org/10.1016/j.biopsycho.2008.03.006>
- Braams, B. R., van Duijvenvoorde, A. C. K., Peper, J. S., & Crone, E. A. (2015). Longitudinal Changes in Adolescent Risk-Taking: A Comprehensive Study of Neural Responses to Rewards, Pubertal Development, and Risk-Taking Behavior. *Journal of Neuroscience*, 35(18), 7226–7238. <https://doi.org/10.1523/JNEUROSCI.4764-14.2015>
- Briant, A. E., Sisk, L. M., & Gee, D. G. (2021). Associations among negative life events, changes in cortico-limbic connectivity, and psychopathology in the ABCD Study. *Developmental Cognitive Neuroscience*, 52, 101022. <https://doi.org/10.1016/j.dcn.2021.101022>
- Brouwer, R. M., Klein, M., Grasby, K. L., Schnack, H. G., Jahanshad, N., Teeuw, J., Medland, S. E., Franke, B., Thompson, P., Pol, H. E. H., & ENIGMA Plasticity Working Group. (2020). Genetic markers for brain plasticity: Neuroimaging: Imaging genetics. *Alzheimer's & Dementia*, 16(S5). <https://doi.org/10.1002/alz.042812>
- Brouwer, R. M., Klein, M., Grasby, K. L., Schnack, H. G., Jahanshad, N., Teeuw, J., Thomopoulos, S. I., Sprooten, E., Franz, C. E., Gogtay, N., Kremen, W. S., Panizzon, M. S., Olde Loohuis, L. M., Whelan, C. D., Aghajani, M., Alloza, C., Alnæs, D., Artiges, E., Ayesa-Arriola, R., ... Hulshoff Pol, H. E. (2022). Genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nature Neuroscience*, 25(4), 421–432. <https://doi.org/10.1038/s41593-022-01042-4>

- Brouwer, R. M., Mandl, R. C. W., Peper, J. S., van Baal, G. C. M., Kahn, R. S., Boomsma, D. I., & Hulshoff Pol, H. E. (2010). Heritability of DTI and MTR in nine-year-old children. *NeuroImage*, 53(3), 1085–1092. <https://doi.org/10.1016/j.neuroimage.2010.03.017>
- Brouwer, R. M., Mandl, R. C. W., Schnack, H. G., van Soelen, I. L. C., van Baal, G. C., Peper, J. S., Kahn, R. S., Boomsma, D. I., & Pol, H. E. H. (2012). White Matter Development in Early Puberty: A Longitudinal Volumetric and Diffusion Tensor Imaging Twin Study. *PLoS ONE*, 7(4), e32316. <https://doi.org/10.1371/journal.pone.0032316>
- Brouwer, R. M., Panizzon, M. S., Glahn, D. C., Hibar, D. P., Hua, X., Jahanshad, N., Abramovic, L., de Zubicaray, G. I., Franz, C. E., Hansell, N. K., Hickie, I. B., Koenis, M. M. G., Martin, N. G., Mather, K. A., McMahon, K. L., Schnack, H. G., Strike, L. T., Swagerman, S. C., Thalamuthu, A., ... Hulshoff Pol, H. E. (2017). Genetic influences on individual differences in longitudinal changes in global and subcortical brain volumes: Results of the ENIGMA plasticity working group: Heritability Estimates of Brain Changes. *Human Brain Mapping*, 38(9), 4444–4458. <https://doi.org/10.1002/hbm.23672>
- Brouwer, R. M., Schutte, J., Janssen, R., Boomsma, D. I., Hulshoff Pol, H. E., & Schnack, H. G. (2021). The Speed of Development of Adolescent Brain Age Depends on Sex and Is Genetically Determined. *Cerebral Cortex*, 31(2), 1296–1306. <https://doi.org/10.1093/cercor/bhaa296>
- Brown, S. A., Brumback, T., Tomlinson, K., Cummins, K., Thompson, W. K., Nagel, B. J., De Bellis, M. D., Hooper, S. R., Clark, D. B., Chung, T., Hasler, B. P., Colrain, I. M., Baker, F. C., Prouty, D., Pfefferbaum, A., Sullivan, E. V., Pohl, K. M., Rohlfing, T., Nichols, B. N., ... Tapert, S. F. (2015). The National Consortium on Alcohol and NeuroDevelopment in Adolescence (NCANDA): A Multisite Study of Adolescent Development and Substance Use. *Journal of Studies on Alcohol and Drugs*, 76(6), 895–908. <https://doi.org/10.15288/jsad.2015.76.895>
- Brown, T. T., Petersen, S. E., & Schlaggar, B. L. (2006). Does human functional brain organization shift from diffuse to focal with development? *Developmental Science*, 9(1), 9–11. <https://doi.org/10.1111/j.1467-7687.2005.00455.x>
- Brunson, K. L., Chen, Y., Avishai-Eliner, S., & Baram, T. Z. (2003). Stress and the Developing Hippocampus: A Double-Edged Sword? *Molecular Neurobiology*, 27(2), 121–136. <https://doi.org/10.1385/MN:27:2:121>
- Buimer, E. E. L., Brouwer, R. M., Mandl, R. C. W., Pas, P., Schnack, H. G., & Hulshoff Pol, H. E. (2022). Adverse childhood experiences and fronto-subcortical structures in the developing brain. *Frontiers in Psychiatry*, 13. <https://www.frontiersin.org/articles/10.3389/fpsy.2022.955871>
- Buimer, E. E. L., Pas, P., Brouwer, R. M., Froeling, M., Hoogduin, H., Leemans, A., Luijten, P., van Nierop, B. J., Raemaekers, M., Schnack, H. G., Teeuw, J., Vink, M., Visser, F., Hulshoff Pol, H. E., & Mandl, R. C. W. (2020). The YOUth cohort study: MRI protocol and test-retest reliability in adults. *Developmental Cognitive Neuroscience*, 45, 100816. <https://doi.org/10.1016/j.dcn.2020.100816>

- Buimer, E. E. L., Schnack, H. G., Caspi, Y., Haren, N. E. M., Milchenko, M., Pas, P., Alzheimer's Disease Neuroimaging Initiative, Hulshoff Pol, H. E., & Brouwer, R. M. (2021). De-identification procedures for magnetic resonance images and the impact on structural brain measures at different ages. *Human Brain Mapping, 42*(11), 3643–3655. <https://doi.org/10.1002/hbm.25459>
- Burley, D. T., Hobson, C. W., Adegboye, D., Shelton, K. H., & van Goozen, S. H. M. (2022). Negative parental emotional environment increases the association between childhood behavioral problems and impaired recognition of negative facial expressions. *Development and Psychopathology, 34*(3), 936–945. <https://doi.org/10.1017/S0954579420002072>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Buzkova, P. (2016). Interaction testing: residuals-based permutations and parametric bootstrap in continuous, count, and binary data. *Epidemiologic Methods, 5*(1), 119–128.

C

- Caballero-Gaudes, C., & Reynolds, R. C. (2017). Methods for cleaning the BOLD fMRI signal. *NeuroImage, 154*, 128–149. <https://doi.org/10.1016/j.neuroimage.2016.12.018>
- Calem, M., Bromis, K., McGuire, P., Morgan, C., & Kempton, M. J. (2017). Meta-analysis of associations between childhood adversity and hippocampus and amygdala volume in non-clinical and general population samples. *NeuroImage: Clinical, 14*, 471–479. <https://doi.org/10.1016/j.nicl.2017.02.016>
- Callaghan, B. L., & Tottenham, N. (2016). The Stress Acceleration Hypothesis: Effects of early-life adversity on emotion circuits and behavior. *Current Opinion in Behavioral Sciences, 7*, 76–81. <https://doi.org/10.1016/j.cobeha.2015.11.018>
- Caspi, Y., Brouwer, R. M., Schnack, H. G., van de Nieuwenhuijzen, M. E., Cahn, W., Kahn, R. S., Niessen, W. J., van der Lugt, A., & Pol, H. H. (2020). Changes in the intracranial volume from early adulthood to the sixth decade of life: A longitudinal study. *NeuroImage, 220*, 116842. <https://doi.org/10.1016/j.neuroimage.2020.116842>
- Cassiers, L. L. M., Sabbe, B. G. C., Schmaal, L., Veltman, D. J., Penninx, B. W. J. H., & Van Den Eede, F. (2018). Structural and Functional Brain Abnormalities Associated With Exposure to Different Childhood Trauma Subtypes: A Systematic Review of Neuroimaging Findings. *Frontiers in Psychiatry, 9*, 329. <https://doi.org/10.3389/fpsy.2018.00329>
- Chai, X. J., Castañón, A. N., Öngür, D., & Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *NeuroImage, 59*(2), 1420–1428. <https://doi.org/10.1016/j.neuroimage.2011.08.048>

- Cheng, H., Wang, Y., Sheng, J., Kronenberger, W. G., Mathews, V. P., Hummer, T. A., & Saykin, A. J. (2012). Characteristics and variability of structural networks derived from diffusion tensor imaging. *NeuroImage*, 61(4), 1153–1164. <https://doi.org/10.1016/j.neuroimage.2012.03.036>
- Chiao, J. Y., & Cheon, B. K. (2010). The weirdest brains in the world. *The Behavioral and Brain Sciences*, 33(2–3), 88–90. <https://doi.org/10.1017/S0140525X10000282>
- Collin, L., Bindra, J., Raju, M., Gillberg, C., & Minnis, H. (2013). Facial emotion recognition in child psychiatry: A systematic review. *Research in Developmental Disabilities*, 34(5), 1505–1520. <https://doi.org/10.1016/j.ridd.2013.01.008>
- Cicchetti, D., & Toth, S. L. (2015). Multilevel developmental perspectives on child maltreatment. *Development and psychopathology*, 27(4pt2), 1385–1386.
- Collins, D. L., Holmes, C. J., Peters, T. M., & Evans, A. C. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Human Brain Mapping*, 3(3), 190–208. <https://doi.org/10.1002/hbm.460030304>
- Cordero-Grande, L., Christiaens, D., Hutter, J., Price, A. N., & Hajnal, J. V. (2019). Complex diffusion-weighted image estimation via matrix recovery under general noise models. *NeuroImage*, 200, 391–404. <https://doi.org/10.1016/j.neuroimage.2019.06.039>
- Creswell, C., Shum, A., Pearcey, S., Skripkauskaitė, S., Patalay, P., & Waite, P. (2021). Young people's mental health during the COVID-19 pandemic. *The Lancet Child & Adolescent Health*, 5(8), 535–537. [https://doi.org/10.1016/S2352-4642\(21\)00177-2](https://doi.org/10.1016/S2352-4642(21)00177-2)
- Crone, E. A., & Achterberg, M. (2022). Prosocial development in adolescence. *Current opinion in psychology*, 44, 220–225.
- Crone, E. A., & Elzinga, B. M. (2015). Changing brains: How longitudinal functional magnetic resonance imaging studies can inform us about cognitive and social-affective growth trajectories. *WIREs Cognitive Science*, 6(1), 53–63. <https://doi.org/10.1002/wcs.1327>

D

- Dahmen, B., Puetz, V. B., Scharke, W., von Polier, G. G., Herpertz-Dahlmann, B., & Konrad, K. (2018). Effects of Early-Life Adversity on Hippocampal Structures and Associated HPA Axis Functions. *Developmental Neuroscience*, 40(1), 13–22. <https://doi.org/10.1159/000484238>
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1), 113.
- Daniels, J. K., Lamke, J.-P., Gaebler, M., Walter, H., & Scheel, M. (2013). White matter integrity and its relationship to PTSD and childhood trauma - A systematic review and meta-analysis: White Matter Integrity. *Depression and Anxiety*, 30(3), 207–216. <https://doi.org/10.1002/da.22044>

- de Reus, M. A., & van den Heuvel, M. P. (2013). Estimating false positives and negatives in brain networks. *NeuroImage*, 70, 402–409. <https://doi.org/10.1016/j.neuroimage.2012.12.066>
- de Sitter, A., Visser, M., Brouwer, I., Cover, K. S., van Schijndel, R. A., Eijgelaar, R. S., Müller, D. M. J., Ropele, S., Kappos, L., Rovira, Á., Filippi, M., Enzinger, C., Frederiksen, J., Ciccarelli, O., Guttman, C. R. G., Wattjes, M. P., Witte, M. G., de Witt Hamer, P. C., Barkhof, F., & Vrenken, H. (2020). Facing privacy in neuroimaging: Removing facial features degrades performance of image analysis methods. *European Radiology*, 30(2), 1062–1074. <https://doi.org/10.1007/s00330-019-06459-3>
- Delvecchio, G., Sugranyes, G., & Frangou, S. (2013). Evidence of diagnostic specificity in the neural correlates of facial affect processing in bipolar disorder and schizophrenia: A meta-analysis of functional imaging studies. *Psychological Medicine*, 43(3), 553–569. <https://doi.org/10.1017/S0033291712001432>
- den Braber, A., Bohlken, M. M., Brouwer, R. M., van 't Ent, D., Kanai, R., Kahn, R. S., de Geus, E. J. C., Hulshoff Pol, H. E., & Boomsma, D. I. (2013). Heritability of subcortical brain measures: A perspective for future genome-wide association studies. *NeuroImage*, 83, 98–102. <https://doi.org/10.1016/j.neuroimage.2013.06.027>
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Dick, A. S., Lopez, D. A., Watts, A. L., Heeringa, S., Reuter, C., Bartsch, H., Fan, C. C., Kennedy, D. N., Palmer, C., Marshall, A., Haist, F., Hawes, S., Nichols, T. E., Barch, D. M., Jernigan, T. L., Garavan, H., Grant, S., Pariyadath, V., Hoffman, E., ... Thompson, W. K. (2021). Meaningful associations in the adolescent brain cognitive development study. *NeuroImage*, 239, 118262. <https://doi.org/10.1016/j.neuroimage.2021.118262>
- Dimitriadis, S. I., Drakesmith, M., Bells, S., Parker, G. D., Linden, D. E., & Jones, D. K. (2017). Improving the Reliability of Network Metrics in Structural Brain Networks by Integrating Different Network Weighting Strategies into a Single Graph. *Frontiers in Neuroscience*, 11, 694. <https://doi.org/10.3389/fnins.2017.00694>
- Dodell-Feder, D., Tully, L. M., Lincoln, S. H., & Hooker, C. I. (2014). The neural basis of theory of mind and its relationship to social functioning and social anhedonia in individuals with schizophrenia. *NeuroImage: Clinical*, 4, 154–163. <https://doi.org/10.1016/j.nicl.2013.11.006>
- Drakulich, S., Thiffault, A.-C., Olafson, E., Parent, O., Labbe, A., Albaugh, M. D., Khundrakpam, B., Ducharme, S., Evans, A., Chakravarty, M. M., & Karama, S. (2021). Maturational trajectories of pericortical contrast in typical brain development. *NeuroImage*, 235, 117974. <https://doi.org/10.1016/j.neuroimage.2021.117974>
- Duan, F., Zhao, T., He, Y., & Shu, N. (2015). Test-retest reliability of diffusion measures in cerebral white matter: A multiband diffusion MRI study: Test-Retest Reliability of Diffusion

Measures. *Journal of Magnetic Resonance Imaging*, 42(4), 1106–1116.
<https://doi.org/10.1002/jmri.24859>

Dubois, J., Alison, M., Counsell, S. J., Hertz-Pannier, L., Hüppi, P. S., & Benders, M. J. N. L. (2021). MRI of the Neonatal Brain: A Review of Methodological Challenges and Neuroscientific Advances. *Journal of Magnetic Resonance Imaging*, 53(5), 1318–1343.
<https://doi.org/10.1002/jmri.27192>

Duning, T., Kloska, S., Steinstrater, O., Kugel, H., Heindel, W., & Knecht, S. (2005). Dehydration confounds the assessment of brain atrophy. *Neurology*, 64(3), 548–550.
<https://doi.org/10.1212/01.WNL.0000150542.16969.CC>

Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J.-Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology*, 97(1), 14–27. <https://doi.org/10.1016/j.jecp.2006.12.001>

Durston, S., Davidson, M. C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J. A., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development. *Developmental Science*, 9(1), 1–8. <https://doi.org/10.1111/j.1467-7687.2005.00454.x>

Durston, S., Nederveen, H., van Dijk, S., van Belle, J., de Zeeuw, P., Langen, M., & van Dijk, A. (2009). Magnetic Resonance Simulation is Effective in Reducing Anxiety Related to Magnetic Resonance Scanning in Children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(2), 206–207. <https://doi.org/10.1097/CHI.0b013e3181930673>

E

Etkin, A., & Wager, T. D. (2007). Functional Neuroimaging of Anxiety: A Meta-Analysis of Emotional Processing in PTSD, Social Anxiety Disorder, and Specific Phobia. *American Journal of Psychiatry*, 164(10), 1476–1488. <https://doi.org/10.1176/appi.ajp.2007.07030504>

Evans, A. C. (2006). The NIH MRI study of normal brain development. *NeuroImage*, 30(1), 184–202.
<https://doi.org/10.1016/j.neuroimage.2005.09.068>

F

Fair, D. A., Dosenbach, N. U., Moore, A. H., Satterthwaite, T. D., & Milham, M. P. (2021). Developmental cognitive neuroscience in the era of networks and big data: Strengths, weaknesses, opportunities, and threats. *Annual Review of Developmental Psychology*, 3, 249–275.

Fakkel, M., Peeters, M., Lugtig, P., Zondervan-Zwijnenburg, M. A. J., Blok, E., White, T., van der Meulen, M., Kevenaar, S. T., Willemsen, G., Bartels, M., Boomsma, D. I., Schmengler, H., Branje, S., & Vollebbergh, W. A. M. (2020). Testing sampling bias in estimates of adolescent social competence and behavioral control. *Developmental Cognitive Neuroscience*, 46, 100872. <https://doi.org/10.1016/j.dcn.2020.100872>

- Fani, N., Stenson, A. F., Rooij, S. J. H., La Barrie, D. L., & Jovanovic, T. (2021). White matter microstructure in trauma-exposed children: Associations with pubertal stage. *Developmental Science*, 24(6). <https://doi.org/10.1111/desc.13120>
- Farrell, J. A. D., Landman, B. A., Jones, C. K., Smith, S. A., Prince, J. L., van Zijl, P. C. M., & Mori, S. (2007). Effects of signal-to-noise ratio on the accuracy and reproducibility of diffusion tensor imaging–derived fractional anisotropy, mean diffusivity, and principal eigenvector measurements at 1.5T. *Journal of Magnetic Resonance Imaging*, 26(3), 756–767. <https://doi.org/10.1002/jmri.21053>
- Fecher, B., & Friesike, S. (2014). Open science: one term, five schools of thought. *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*, 17-47.
- Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., Koss, M. P., & Marks, J. S. (1998). Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults. *American Journal of Preventive Medicine*, 14(4), 245–258. [https://doi.org/10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8)
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole Brain Segmentation. *Neuron*, 33(3), 341–355. [https://doi.org/10.1016/S0896-6273\(02\)00569-X](https://doi.org/10.1016/S0896-6273(02)00569-X)
- Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., & Collins, D. L. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54(1), 313–327. <https://doi.org/10.1016/j.neuroimage.2010.07.033>
- Foulkes, L., & Blakemore, S.-J. (2018). Studying individual differences in human adolescent brain development. *Nature Neuroscience*, 21(3), 315–323. <https://doi.org/10.1038/s41593-018-0078-4>
- Frangou, S., Modabbernia, A., Williams, S. C. R., Papachristou, E., Doucet, G. E., Agartz, I., Aghajani, M., Akudjedu, T. N., Albajes-Eizagirre, A., Alnæs, D., Alpert, K. I., Andersson, M., Andreasen, N. C., Andreassen, O. A., Asherson, P., Banaschewski, T., Bargallo, N., Baumeister, S., Baur-Streubel, R., ... Dima, D. (2022). Cortical thickness across the lifespan: Data from 17,075 healthy individuals aged 3–90 years. *Human Brain Mapping*, 43(1), 431–451. <https://doi.org/10.1002/hbm.25364>
- Friedman, L., & Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *Journal of Magnetic Resonance Imaging*, 23(6), 827–839. <https://doi.org/10.1002/jmri.20583>
- Friedman, L., Glover, G. H., & The FBIRN Consortium. (2006). Reducing interscanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. *NeuroImage*, 33(2), 471–481. <https://doi.org/10.1016/j.neuroimage.2006.07.012>

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series: Movement Artifacts in fMRI. *Magnetic Resonance in Medicine*, 35(3), 346–355. <https://doi.org/10.1002/mrm.1910350312>

Friston, K. (2002). Beyond phrenology: what can neuroimaging tell us about distributed circuitry?. *Annual review of neuroscience*, 25(1), 221-250.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., Benedetti, F., Abbamonte, M., Gasparotti, R., Barale, F., Perez, J., McGuire, P., & Politi, P. (2009). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J Psychiatry Neurosci*, 34(6), 418–432.

Fuhrmann, D., Knoll, L. J., & Blakemore, S. J. (2015). Adolescence as a sensitive period of brain development. *Trends in cognitive sciences*, 19(10), 558-566.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., ... & Politi, P. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of psychiatry and neuroscience*, 34(6), 418-432.

G

Gabard-Durnam, L., & McLaughlin, K. A. (2020). Sensitive periods in human development: charting a course for the future. *Current Opinion in Behavioral Sciences*, 36, 120-128.

Gee, D. G., & Casey, B. J. (2015). The impact of developmental timing for stress and recovery. *Neurobiology of Stress*, 1, 184–194. <https://doi.org/10.1016/j.ynstr.2015.02.001>

Geerligs, L., Rubinov, M., Cam-CAN, & Henson, R. N. (2015). State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. *The Journal of Neuroscience*, 35(41), 13949–13961. <https://doi.org/10.1523/JNEUROSCI.1324-15.2015>

Gheorghie, D. A., Li, C., Gallacher, J., & Bauermeister, S. (2021). Associations of perceived adverse lifetime experiences with brain structure in UK Biobank participants. *Journal of Child Psychology and Psychiatry*, 62(7), 822–830. <https://doi.org/10.1111/jcpp.13298>

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, 2(10), 861–863. <https://doi.org/10.1038/13158>

Goemans, A., Viding, E., & McCrory, E. (2023). Child Maltreatment, Peer Victimization, and Mental Health: Neurocognitive Perspectives on the Cycle of Victimization. *Trauma, Violence, & Abuse*, 24(2), 530–548. <https://doi.org/10.1177/15248380211036393>

González-García, N., Buimer, E.E.L., Moreno-López, L., Sallie, S., Váša, F., Lim, S., . . . Van Harmelen, A. (2023). Resilient functioning is associated with altered structural brain network topology in adolescents exposed to childhood adversity. *Development and Psychopathology*, 1-11. doi: <https://doi.org/10.1017/S0954579423000901>

- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of child psychology and psychiatry*, 38(5), 581-586.
- Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.
- Gorgolewski, K. J., Storkey, A., Bastin, M. E., Whittle, I. R., Wardlaw, J. M., & Pernet, C. R. (2013a). A test-retest fMRI dataset for motor, language and spatial attention functions. *GigaScience*, 2(1), 6. <https://doi.org/10.1186/2047-217X-2-6>
- Gorgolewski, K. J., Storkey, A. J., Bastin, M. E., Whittle, I., & Pernet, C. (2013b). Single subject fMRI test-retest reliability metrics and confounding factors. *NeuroImage*, 69, 231-243. <https://doi.org/10.1016/j.neuroimage.2012.10.085>
- Graf, G. H. J., Biroli, P., & Belsky, D. W. (2021). Critical periods in child development and the transition to adulthood. *JAMA Network Open*, 4(1), e2033359-e2033359.
- Gratton, C., Nelson, S. M., & Gordon, E. M. (2022). Brain-behavior correlations: Two paths toward reliability. *Neuron*, 110(9), 1446-1449. <https://doi.org/10.1016/j.neuron.2022.04.018>
- Green, J. G., McLaughlin, K. A., Berglund, P. A., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., & Kessler, R. C. (2010). Childhood Adversities and Adult Psychiatric Disorders in the National Comorbidity Survey Replication I: Associations With First Onset of DSM-IV Disorders. *Archives of General Psychiatry*, 67(2), 113. <https://doi.org/10.1001/archgenpsychiatry.2009.186>
- Green, K. H., Van De Groep, I. H., Te Brinke, L. W., van der Crujisen, R., van Rossenberg, F., & El Marroun, H. (2022). A perspective on enhancing representative samples in developmental human neuroscience: Connecting science to society. *Frontiers in Integrative Neuroscience*, 16, 981657. <https://doi.org/10.3389/fnint.2022.981657>
- Green, K. H., Becht, A. I., van de Groep, S., van der Crujisen, R., Sweijen, S. W., & Crone, E. A. (2023). Socioeconomic hardship, uncertainty about the future, and adolescent mental wellbeing over a year during the COVID-19 pandemic. *Social Development*.
- Gulban, O. F., Nielson, D., Lee, J., Poldrack, R., Gorgolewski, C., Vanessasaurus, & Satrajit Ghosh. (2022). poldracklab/pydeface: V2.0.1 (v2.0.1). Zenodo. <https://doi.org/10.5281/ZENODO.6822992>
- Gur, R. C., Richard, J., Calkins, M. E., Chiavacci, R., Hansen, J. A., Bilker, W. B., Loughhead, J., Connolly, J. J., Qiu, H., Mentch, F. D., Abou-Sleiman, P. M., Hakonarson, H., & Gur, R. E. (2012). Age group and sex differences in performance on a computerized neurocognitive battery in children age 8-21. *Neuropsychology*, 26(2), 251-265. <https://doi.org/10.1037/a0026712>
- Gur, R. C., Richard, J., Hughett, P., Calkins, M. E., Macy, L., Bilker, W. B., Brensinger, C., & Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, 187(2), 254-262. <https://doi.org/10.1016/j.jneumeth.2009.11.017>

H

- Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial Emotion Recognition in Autism Spectrum Disorders: A Review of Behavioral and Neuroimaging Studies. *Neuropsychology Review*, 20(3), 290–322. <https://doi.org/10.1007/s11065-010-9138-6>
- Hart, & Rubia, K. (2012). Neuroimaging of child abuse: A critical review. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00052>
- Hart, S. A., Little, C., & van Bergen, E. (2021). Nurture might be nature: Cautionary tales and proposed solutions. *Npj Science of Learning*, 6(1), 2. <https://doi.org/10.1038/s41539-020-00079-z>
- Hawk, S. T., Keijsers, L., Branje, S. J., Graaff, J. V. D., Wied, M. D., & Meeus, W. (2013). Examining the interpersonal reactivity index (IRI) among early and late adolescents and their mothers. *Journal of personality assessment*, 95(1), 96-106.
- Heckendorf, E., Huffmeijer, R., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2016). Neural Processing of Familiar and Unfamiliar Children's Faces: Effects of Experienced Love Withdrawal, but No Effects of Neutral and Threatening Priming. *Frontiers in Human Neuroscience*, 10. <https://doi.org/10.3389/fnhum.2016.00231>
- Heim, C., & Binder, E. B. (2012). Current research trends in early life stress and depression: Review of human studies on sensitive periods, gene–environment interactions, and epigenetics. *Experimental Neurology*, 233(1), 102–111. <https://doi.org/10.1016/j.expneurol.2011.10.032>
- Hein, T. C., & Monk, C. S. (2017). Research Review: Neural response to threat in children, adolescents, and adults after child maltreatment - a quantitative meta-analysis. *Journal of Child Psychology and Psychiatry*, 58(3), 222–230. <https://doi.org/10.1111/jcpp.12651>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. *Journal of Child Psychology and Psychiatry*, 45(7), 1185–1198. <https://doi.org/10.1111/j.1469-7610.2004.00316.x>
- Herting, M. M., Gautam, P., Chen, Z., Mezher, A., & Vetter, N. C. (2018). Test-retest reliability of longitudinal task-based fMRI: Implications for developmental studies. *Developmental Cognitive Neuroscience*, 33, 17–26. <https://doi.org/10.1016/j.dcn.2017.07.001>
- Herting, M. M., Gautam, P., Spielberg, J. M., Kan, E., Dahl, R. E., & Sowell, E. R. (2014). The role of testosterone and estradiol in brain volume changes across adolescence: A longitudinal structural MRI study: Pubertal Hormones and Brain Volume. *Human Brain Mapping*, 35(11), 5633–5645. <https://doi.org/10.1002/hbm.22575>

Hoemann, K., Wu, R., LoBue, V., Oakes, L. M., Xu, F., & Barrett, L. F. (2020). Developing an understanding of emotion categories: Lessons from objects. *Trends in Cognitive Sciences*, 24(1), 39-51.

Holmes, A. J., Hollinshead, M. O., O'Keefe, T. M., Petrov, V. I., Fariello, G. R., Wald, L. L., Fischl, B., Rosen, B. R., Mair, R. W., Roffman, J. L., Smoller, J. W., & Buckner, R. L. (2015). Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures. *Scientific Data*, 2(1), 150031. <https://doi.org/10.1038/sdata.2015.31>

Hughes, K., Bellis, M. A., Hardcastle, K. A., Sethi, D., Butchart, A., Mikton, C., Jones, L., & Dunne, M. P. (2017). The effect of multiple adverse childhood experiences on health: A systematic review and meta-analysis. *The Lancet Public Health*, 2(8), e356–e366. [https://doi.org/10.1016/S2468-2667\(17\)30118-4](https://doi.org/10.1016/S2468-2667(17)30118-4)

Hulshoff Pol, H. E., van Baal, G. C. M., Schnack, H. G., Brans, R. G. H., van der Schot, A. C., Brouwer, R. M., van Haren, N. E. M., Lepage, C., Collins, D. L., Evans, A. C., Boomsma, D. I., Nolen, W., & Kahn, R. S. (2012). Overlapping and segregating structural brain abnormalities in twins with schizophrenia or bipolar disorder. *Archives of General Psychiatry*, 69(4), 349–359. <https://doi.org/10.1001/archgenpsychiatry.2011.1615>

Hunnikin, L. M., Wells, A. E., Ash, D. P., & van Goozen, S. H. M. (2020). The nature and extent of emotion recognition and empathy impairments in children showing disruptive behaviour referred into a crime prevention programme. *European Child & Adolescent Psychiatry*, 29(3), 363–371. <https://doi.org/10.1007/s00787-019-01358-w>

I

Ioannidis, K., Askelund, A. D., Kievit, R. A., & van Harmelen, A.-L. (2020). The complex neurobiology of resilient functioning after childhood maltreatment. *BMC Medicine*, 18(1), 32. <https://doi.org/10.1186/s12916-020-1490-7>

Iscan, Z., Jin, T. B., Kendrick, A., Szeglin, B., Lu, H., Trivedi, M., Fava, M., McGrath, P. J., Weissman, M., Kurian, B. T., Adams, P., Weyandt, S., Toups, M., Carmody, T., McInnis, M., Cusin, C., Cooper, C., Oquendo, M. A., Parsey, R. V., & DeLorenzo, C. (2015). Test-retest reliability of freesurfer measurements within and between sites: Effects of visual approval process: Test-Retest Reliability of FreeSurfer Measurements. *Human Brain Mapping*, 36(9), 3472–3485. <https://doi.org/10.1002/hbm.22856>

Israelashvili, J., Sauter, D., & Fischer, A. (2020). Two facets of affective empathy: Concern and distress have opposite relationships to emotion recognition. *Cognition and Emotion*, 34(6), 1112–1122. <https://doi.org/10.1080/02699931.2020.1724893>

J

Jansen, A. G., Mous, S. E., White, T., Posthuma, D., & Polderman, T. J. (2015). What twin studies tell us about the heritability of brain development, morphology, and function: a review. *Neuropsychology review*, 25, 27-46.

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>

Job, D. E., Dickie, D. A., Rodriguez, D., Robson, A., Danso, S., Pernet, C., Bastin, M. E., Boardman, J. P., Murray, A. D., Ahearn, T., Waiter, G. D., Staff, R. T., Deary, I. J., Shenkin, S. D., & Wardlaw, J. M. (2017). A brain imaging repository of normal structural MRI across the life course: Brain Images of Normal Subjects (BRAINS). *NeuroImage*, 144, 299–304. <https://doi.org/10.1016/j.neuroimage.2016.01.027>

Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Nobili, F., Hensch, T., Tränkner, A., Schönknecht, P., Leroy, M., Lopes, R., Bordet, R., Chanoine, V., Ranjeva, J.-P., Didic, M., ... Frisoni, G. B. (2013). Brain morphometry reproducibility in multi-center 3T MRI studies: A comparison of cross-sectional and longitudinal segmentations. *NeuroImage*, 83, 472–484. <https://doi.org/10.1016/j.neuroimage.2013.05.007>

Junge, C., Valkenburg, P. M., Deković, M., & Branje, S. (2020). The building blocks of social competence: Contributions of the Consortium of Individual Development. *Developmental Cognitive Neuroscience*, 45, 100861. <https://doi.org/10.1016/j.dcn.2020.100861>

K

Kalisch, R., Baker, D. G., Basten, U., Boks, M. P., Bonanno, G. A., Brummelman, E., Chmitorz, A., Fernández, G., Fiebach, C. J., Galatzer-Levy, I., Geuze, E., Groppa, S., Helmreich, I., Hendler, T., Hermans, E. J., Jovanovic, T., Kubiak, T., Lieb, K., Lutz, B., ... Kleim, B. (2017). The resilience framework as a strategy to combat stress-related disorders. *Nature Human Behaviour*, 1(11), 784–790. <https://doi.org/10.1038/s41562-017-0200-8>

Kalisch, R., Cramer, A. O. J., Binder, H., Fritz, J., Leertouwer, Ij., Lunansky, G., Meyer, B., Timmer, J., Veer, I. M., & van Harmelen, A.-L. (2019). Deconstructing and Reconstructing Resilience: A Dynamic Network Approach. *Perspectives on Psychological Science*, 14(5), 765–777. <https://doi.org/10.1177/1745691619855637>

Kalmakis, K. A., & Chandler, G. E. (2014). Adverse childhood experiences: Towards a clear conceptual meaning. *Journal of Advanced Nursing*, 70(7), 1489–1501. <https://doi.org/10.1111/jan.12329>

Kalmakis, K. A., & Chandler, G. E. (2015). Health consequences of adverse childhood experiences: A systematic review. *Journal of the American Association of Nurse Practitioners*, 27(8), 457–465. <https://doi.org/10.1002/2327-6924.12215>

Kellner, E., Dhital, B., Kiselev, V. G., & Reisert, M. (2016). Gibbs-ringing artifact removal based on local subvoxel-shifts: Gibbs-Ringing Artifact Removal. *Magnetic Resonance in Medicine*, 76(5), 1574–1581. <https://doi.org/10.1002/mrm.26054>

Kempton, M. J., Ettinger, U., Schmechtig, A., Winter, E. M., Smith, L., McMorris, T., Wilkinson, I. D., Williams, S. C. R., & Smith, M. S. (2009). Effects of acute dehydration on brain morphology in healthy humans. *Human Brain Mapping*, 30(1), 291–298. <https://doi.org/10.1002/hbm.20500>

- Kessler, R. C., McLaughlin, K. A., Green, J. G., Gruber, M. J., Sampson, N. A., Zaslavsky, A. M., Aguilar-Gaxiola, S., Alhamzawi, A. O., Alonso, J., Angermeyer, M., Benjet, C., Bromet, E., Chatterji, S., de Girolamo, G., Demyttenaere, K., Fayyad, J., Florescu, S., Gal, G., Gureje, O., ... Williams, D. R. (2010). Childhood adversities and adult psychopathology in the WHO World Mental Health Surveys. *British Journal of Psychiatry*, 197(5), 378–385. <https://doi.org/10.1192/bjp.bp.110.080499>
- Knafo-Noam, A., Uzefovsky, F., Israel, S., Davidov, M., & Zahn-Waxler, C. (2015). The prosocial personality and its facets: genetic and environmental architecture of mother-reported behavior of 7-year-old twins. *Frontiers in psychology*, 6, 112.
- Koenis, M. M., Brouwer, R. M., van Baal, G. C. M., van Soelen, I. L., Peper, J. S., van Leeuwen, M., ... & Hulshoff Pol, H. E. (2013). Longitudinal study of hormonal and physical development in young twins. *The Journal of Clinical Endocrinology & Metabolism*, 98(3), E518-E527.
- Koenis, M. M. G., Brouwer, R. M., van den Heuvel, M. P., Mandl, R. C. W., van Soelen, I. L. C., Kahn, R. S., Boomsma, D. I., & Hulshoff Pol, H. E. (2015). Development of the brain's structural network efficiency in early adolescence: A longitudinal DTI twin study. *Human Brain Mapping*, 36(12), 4938–4953. <https://doi.org/10.1002/hbm.22988>
- Kolhatkar, G., & Berkowitz, C. (2014). Cultural Considerations and Child Maltreatment. *Pediatric Clinics of North America*, 61(5), 1007–1022. <https://doi.org/10.1016/j.pcl.2014.06.005>
- Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsón, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdóttir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdóttir, G., Thorsteinsdóttir, U., & Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, 359(6374), 424–428. <https://doi.org/10.1126/science.aan6877>
- Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Koolschijn, P. C. M. P., Schel, M. A., de Rooij, M., Rombouts, S. A. R. B., & Crone, E. A. (2011). A Three-Year Longitudinal Functional Magnetic Resonance Imaging Study of Performance Monitoring and Test-Retest Reliability from Childhood to Early Adulthood. *Journal of Neuroscience*, 31(11), 4204–4212. <https://doi.org/10.1523/JNEUROSCI.6415-10.2011>
- Kraaijenvanger, E. J., Pollok, T. M., Monninger, M., Kaiser, A., Brandeis, D., Banaschewski, T., & Holz, N. E. (2020). Impact of early life adversities on human brain functioning: A coordinate-based meta-analysis. *Neuroscience & Biobehavioral Reviews*, 113, 62–76. <https://doi.org/10.1016/j.neubiorev.2020.03.008>
- Kuhn, M., Scharfenort, R., Schumann, D., Schiele, M. A., Münsterkötter, A. L., Deckert, J., Domschke, K., Haaker, J., Kalisch, R., Pauli, P., Reif, A., Romanos, M., Zwanzger, P., & Lonsdorf, T. B. (2016). Mismatch or allostatic load? Timing of life adversity differentially shapes gray matter volume and anxious temperament. *Social Cognitive and Affective Neuroscience*, 11(4), 537–547. <https://doi.org/10.1093/scan/nsv137>

L

- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Larivière, S., Paquola, C., Park, B., Royer, J., Wang, Y., Benkarim, O., Vos de Wael, R., Valk, S. L., Thomopoulos, S. I., Kirschner, M., Lewis, L. B., Evans, A. C., Sisodiya, S. M., McDonald, C. R., Thompson, P. M., & Bernhardt, B. C. (2021). The ENIGMA Toolbox: Multiscale neural contextualization of multisite neuroimaging datasets. *Nature Methods*, 18(7), 698–700. <https://doi.org/10.1038/s41592-021-01186-4>
- Laurienti, P. J., Field, A. S., Burdette, J. H., Maldjian, J. A., Yen, Y.-F., & Moody, D. M. (2002). Dietary Caffeine Consumption Modulates fMRI Measures. *NeuroImage*, 17(2), 751–757. <https://doi.org/10.1006/nimg.2002.1237>
- Leemans, A., & Jones, D. K. (2009). The B -matrix must be rotated when correcting for subject motion in DTI data. *Magnetic Resonance in Medicine*, 61(6), 1336–1349. <https://doi.org/10.1002/mrm.21890>
- Lees, B., Stapinski, L. A., Teesson, M., Squeglia, L. M., Jacobus, J., & Mewton, L. (2021). Problems experienced by children from families with histories of substance misuse: An ABCD study®. *Drug and Alcohol Dependence*, 218, 108403. <https://doi.org/10.1016/j.drugalcdep.2020.108403>
- Lenroot, R. K., & Giedd, J. N. (2008). The changing impact of genes and environment on brain development during childhood and adolescence: Initial findings from a neuroimaging study of pediatric twins. *Development and Psychopathology*, 20(4), 1161–1175. <https://doi.org/10.1017/S0954579408000552>
- Lenroot, R. K., Schmitt, J. E., Ordaz, S. J., Wallace, G. L., Neale, M. C., Lerch, J. P., Kendler, K. S., Evans, A. C., & Giedd, J. N. (2009). Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human Brain Mapping*, 30(1), 163–174. <https://doi.org/10.1002/hbm.20494>
- LeWinn, K. Z., Sheridan, M. A., Keyes, K. M., Hamilton, A., & McLaughlin, K. A. (2017). Sample composition alters associations between age and brain structure. *Nature Communications*, 8(1), 874. <https://doi.org/10.1038/s41467-017-00908-7>
- Liem, F., Mérillat, S., Bezzola, L., Hirsiger, S., Philipp, M., Madhyastha, T., & Jäncke, L. (2015). Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. *NeuroImage*, 108, 95–109. <https://doi.org/10.1016/j.neuroimage.2014.12.035>
- Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N., Lefebvre, S., Nakamura, W., Saldana, D., Schmiesing, A., Tran, C., Vo, D., Ard, T., Heydari, P., Kim, B., ... Stroud, A. (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Scientific Data*, 5(1), 180011. <https://doi.org/10.1038/sdata.2018.11>

- Lim, L., Howells, H., Radua, J., & Rubia, K. (2020). Aberrant structural connectivity in childhood maltreatment: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, 116, 406–414. <https://doi.org/10.1016/j.neubiorev.2020.07.004>
- Löytömäki, J., Laakso, M.-L., & Huttunen, K. (2022). Social-Emotional and Behavioural Difficulties in Children with Neurodevelopmental Disorders: Emotion Perception in Daily Life and in a Formal Assessment Context. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-022-05768-9>
- Luijten, M. A. J., van Muilekom, M. M., Teela, L., Polderman, T. J. C., Terwee, C. B., Zijlmans, J., Klaufus, L., Popma, A., Ostrom, K. J., van Oers, H. A., & Haverman, L. (2021). The impact of lockdown during the COVID-19 pandemic on mental and social health of children and adolescents. *Quality of Life Research*, 30(10), 2795–2804. <https://doi.org/10.1007/s11136-021-02861-x>

M

- Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., Barch, D. M., Archie, K. A., Burgess, G. C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., ... Van Essen, D. C. (2013). Human Connectome Project informatics: Quality control, database services, and data visualization. *NeuroImage*, 80, 202–219. <https://doi.org/10.1016/j.neuroimage.2013.05.077>
- Marcus, D. S., Olsen, T. R., Ramaratnam, M., & Buckner, R. L. (2007a). The extensible neuroimaging archive toolkit: An informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, 5(1), 11–33. <https://doi.org/10.1385/NI:5:1:11>
- Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., & Buckner, R. L. (2007b). Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, 19(9), 1498–1507. <https://doi.org/10.1162/jocn.2007.19.9.1498>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), Article 7902. <https://doi.org/10.1038/s41586-022-04492-9>
- Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and psychopathology*, 22(3), 491-495.
- Masten, A. S., Lucke, C. M., Nelson, K. M., & Stallworthy, I. C. (2021). Resilience in development and psychopathology: Multisystem perspectives. *Annual Review of Clinical Psychology*, 17, 521-549.

- Marzocchi, G. M., Capron, C., Di Pietro, M., Duran Tauleria, E., Duyme, M., Frigerio, A., ... & Théron, C. (2004). The use of the Strengths and Difficulties Questionnaire (SDQ) in Southern European countries. *European Child & Adolescent Psychiatry, 13*, ii40-ii46.
- Maurice-Stam, H., Haverman, L., Splinter, A., Van Oers, H. A., Schepers, S. A., & Grootenhuis, M. A. (2018). Dutch norms for the Strengths and Difficulties Questionnaire (SDQ)–parent form for children aged 2–18 years. *Health and quality of life outcomes, 16*(1), 1-11.
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin, 126*(3), 424–453. <https://doi.org/10.1037/0033-2909.126.3.424>
- McCrory, E., De Brito, S. A., & Viding, E. (2010). Research Review: The neurobiology and genetics of maltreatment and adversity: Research Review: The neurobiology and genetics of maltreatment and adversity. *Journal of Child Psychology and Psychiatry, 51*(10), 1079–1095. <https://doi.org/10.1111/j.1469-7610.2010.02271.x>
- McCrory, E., Foulkes, L., & Viding, E. (2022). Social thinning and stress generation after childhood maltreatment: A neurocognitive social transactional model of psychiatric vulnerability. *The Lancet Psychiatry, 9*(10), 828–837. [https://doi.org/10.1016/S2215-0366\(22\)00202-4](https://doi.org/10.1016/S2215-0366(22)00202-4)
- McCrory, E. J., & Viding, E. (2015). The theory of latent vulnerability: Reconceptualizing the link between childhood maltreatment and psychiatric disorder. *Development and Psychopathology, 27*(2), 493–505. <https://doi.org/10.1017/S0954579415000115>
- McCrory, E. J., Gerin, M. I., & Viding, E. (2017). Annual research review: childhood maltreatment, latent vulnerability and the shift to preventative psychiatry—the contribution of functional brain imaging. *Journal of child psychology and psychiatry, 58*(4), 338-357. <https://doi.org/10.1111/jcpp.12713>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>
- McIntosh, A. R. (2000). Towards a network theory of cognition. *Neural networks, 13*(8-9), 861-870.
- McLaughlin, K. A. (2016). Future Directions in Childhood Adversity and Youth Psychopathology. *Journal of Clinical Child & Adolescent Psychology, 45*(3), 361–382. <https://doi.org/10.1080/15374416.2015.1110823>
- McLaughlin, K. A., & Sheridan, M. A. (2016). Beyond cumulative risk: A dimensional approach to childhood adversity. *Current directions in psychological science, 25*(4), 239-245.
- McLaughlin, K. A., Weissman, D., & Bitrán, D. (2019). Childhood Adversity and Neural Development: A Systematic Review. *Annual Review of Developmental Psychology, 1*(1), 277–312. <https://doi.org/10.1146/annurev-devpsych-121318-084950>
- Merrick, M. T., Ford, D. C., Ports, K. A., Guinn, A. S., Chen, J., Klevens, J., Metzler, M., Jones, C. M., Simon, T. R., Daniel, V. M., Ottley, P., & Mercy, J. A. (2019). Vital Signs: Estimated Proportion of Adult Health Problems Attributable to Adverse Childhood Experiences and

Implications for Prevention—25 States, 2015–2017. *MMWR. Morbidity and Mortality Weekly Report*, 68(44), 999–1005. <https://doi.org/10.15585/mmwr.mm6844e1>

- Milchenko, M., & Marcus, D. (2013). Obscuring Surface Anatomy in Volumetric Imaging Data. *Neuroinformatics*, 11(1), 65–75. <https://doi.org/10.1007/s12021-012-9160-3>
- Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., Bartsch, A. J., Jbabdi, S., Sotiropoulos, S. N., Andersson, J. L. R., Griffanti, L., Douaud, G., Okell, T. W., Weale, P., Dragonu, I., Garratt, S., Hudson, S., Collins, R., Jenkinson, M., ... Smith, S. M. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience*, 19(11), 1523–1536. <https://doi.org/10.1038/nn.4393>
- Mills, K. L., Siegmund, K. D., Tamnes, C. K., Ferschmann, L., Wierenga, L. M., Bos, M. G. N., Luna, B., Li, C., & Herting, M. M. (2021). Inter-individual variability in structural brain development from late childhood to young adulthood. *NeuroImage*, 242, 118450. <https://doi.org/10.1016/j.neuroimage.2021.118450>
- Mills, K. L., & Tamnes, C. K. (2014). Methods and considerations for longitudinal structural brain imaging analysis across development. *Developmental Cognitive Neuroscience*, 9, 172–190. <https://doi.org/10.1016/j.dcn.2014.04.004>
- Mitchell, A. E., Dickens, G. L., & Picchioni, M. M. (2014). Facial Emotion Processing in Borderline Personality Disorder: A Systematic Review and Meta-Analysis. *Neuropsychology Review*, 24(2), 166–184. <https://doi.org/10.1007/s11065-014-9254-9>
- Morales, D. X., Morales, S. A., & Beltran, T. F. (2021). Racial/ethnic disparities in household food insecurity during the COVID-19 pandemic: a nationally representative study. *Journal of racial and ethnic health disparities*, 8, 1300-1314.
- Moreno-López, L., Ioannidis, K., Askelund, A. D., Smith, A. J., Schueler, K., & van Harmelen, A.-L. (2020). The Resilient Emotional Brain: A Scoping Review of the Medial Prefrontal Cortex and Limbic Structure and Function in Resilient Adults With a History of Childhood Maltreatment. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(4), 392–402. <https://doi.org/10.1016/j.bpsc.2019.12.008>
- Morey, R. A., Selgrade, E. S., Wagner, H. R., Huettel, S. A., Wang, L., & McCarthy, G. (2010). Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. *Human Brain Mapping*, NA-NA. <https://doi.org/10.1002/hbm.20973>
- Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A. V., Mahmood, A., Woods, R., Toga, A. W., Pike, G. B., Neto, P. R., Evans, A., Zhang, J., Huang, H., Miller, M. I., van Zijl, P., & Mazziotta, J. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *NeuroImage*, 40(2), 570–582. <https://doi.org/10.1016/j.neuroimage.2007.12.035>
- Mori, S., & van Zijl, P. (2007). Human White Matter Atlas. *American Journal of Psychiatry*, 164(7), 1005–1005. <https://doi.org/10.1176/ajp.2007.164.7.1005>

Murphy, K., Bodurka, J., & Bandettini, P. A. (2007). How long to scan? The relationship between fMRI temporal signal to noise ratio and necessary scan duration. *NeuroImage*, 34(2), 565–574. <https://doi.org/10.1016/j.neuroimage.2006.09.032>

Muris, P., Meesters, C., & van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ) further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European child & adolescent psychiatry*, 12, 1-8.

N

Nakamura, K., Brown, R. A., Araujo, D., Narayanan, S., & Arnold, D. L. (2014). Correlation between brain volume change and T2 relaxation time induced by dehydration and rehydration: Implications for monitoring atrophy in clinical studies. *NeuroImage: Clinical*, 6, 166–170. <https://doi.org/10.1016/j.nicl.2014.08.014>

Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, 203, 116157. <https://doi.org/10.1016/j.neuroimage.2019.116157>

O

Oliveira-Silva, P., Maia, L., & Coutinho, J. (2018). Empathy by default: Correlates in the brain at rest. *Psicothema*, 30.1, 97–103. <https://doi.org/10.7334/psicothema2016.366>

Onland-Moret, N. C., Buizer-Voskamp, J. E., Albers, M. E. W. A., Brouwer, R. M., Buimer, E. E. L., Hessels, R. S., de Heus, R., Huijding, J., Junge, C. M. M., Mandl, R. C. W., Pas, P., Vink, M., van der Wal, J. J. M., Hulshoff Pol, H. E., & Kemner, C. (2020). The YOUth study: Rationale, design, and study procedures. *Developmental Cognitive Neuroscience*, 46, 100868. <https://doi.org/10.1016/j.dcn.2020.100868>

Overgaauw, S., Rieffe, C., Broekhof, E., Crone, E. A., & Güroğlu, B. (2017). Assessing empathy across childhood and adolescence: Validation of the empathy questionnaire for children and adolescents (EmQue-CA). *Frontiers in psychology*, 8, 870.

Ozyurt, I. B., Keator, D. B., Wei, D., Fennema-Notestine, C., Pease, K. R., Bockholt, J., & Grethe, J. S. (2010). Federated Web-accessible Clinical Data Management within an Extensible NeuroImaging Database. *Neuroinformatics*, 8(4), 231–249. <https://doi.org/10.1007/s12021-010-9078-6>

P

Paine, A. L., van Goozen, S. H. M., Burley, D. T., Anthony, R., & Shelton, K. H. (2023). Facial emotion recognition in adopted children. *European Child & Adolescent Psychiatry*, 32(1), 87–99. <https://doi.org/10.1007/s00787-021-01829-z>

- Panchal, U., Salazar de Pablo, G., Franco, M., Moreno, C., Parellada, M., Arango, C., & Fusar-Poli, P. (2021). The impact of COVID-19 lockdown on child and adolescent mental health: Systematic review. *European Child & Adolescent Psychiatry*. <https://doi.org/10.1007/s00787-021-01856-w>
- Panizzon, M. S., Fennema-Notestine, C., Eyler, L. T., Jernigan, T. L., Prom-Wormley, E., Neale, M., Jacobson, K., Lyons, M. J., Grant, M. D., Franz, C. E., Xian, H., Tsuang, M., Fischl, B., Seidman, L., Dale, A., & Kremen, W. S. (2009). Distinct Genetic Influences on Cortical Surface Area and Cortical Thickness. *Cerebral Cortex*, 19(11), 2728–2735. <https://doi.org/10.1093/cercor/bhp026>
- Pas, P., Pol, H. H., Raemaekers, M., & Vink, M. (2021). Self-regulation in the pre-adolescent brain. *Developmental Cognitive Neuroscience*, 51, 101012.
- Paquola, C., Bennett, M. R., & Lagopoulos, J. (2016). Understanding heterogeneity in grey matter research of adults with childhood maltreatment—A meta-analysis and review. *Neuroscience & Biobehavioral Reviews*, 69, 299–312. <https://doi.org/10.1016/j.neubiorev.2016.08.011>
- Parker, D. B., & Razlighi, Q. R. (2019). The Benefit of Slice Timing Correction in Common fMRI Preprocessing Pipelines. *Frontiers in Neuroscience*, 13, 821. <https://doi.org/10.3389/fnins.2019.00821>
- Parker, D., Liu, X., & Razlighi, Q. R. (2017). Optimal slice timing correction and its interaction with fMRI parameters and artifacts. *Medical Image Analysis*, 35, 434–445. <https://doi.org/10.1016/j.media.2016.08.006>
- Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, 171, 415–436. <https://doi.org/10.1016/j.neuroimage.2017.12.073>
- Passarotti, A. M., Paul, B. M., Bussiere, J. R., Buxton, R. B., Wong, E. C., & Stiles, J. (2003). The development of face and location processing: An fMRI study. *Developmental Science*, 6(1), 100–117. <https://doi.org/10.1111/1467-7687.00259>
- Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, 9(12), 947–957. <https://doi.org/10.1038/nrn2513>
- Peper, J. S., Schnack, H. G., Brouwer, R. M., Van Baal, G. C. M., Pjetri, E., Székely, E., van Leeuwen, M., van den Berg, S. M., Collins, D. L., Evans, A. C., Boomsma, D. I., Kahn, R. S., & Hulshoff Pol, H. E. (2009). Heritability of regional and global brain structure at the onset of puberty: A magnetic resonance imaging study in 9-year-old twin pairs. *Human Brain Mapping*, 30(7), 2184–2196. <https://doi.org/10.1002/hbm.20660>
- Perrone, D., Aelterman, J., Pižurica, A., Jeurissen, B., Philips, W., & Leemans, A. (2015). The effect of Gibbs ringing artifacts on measures derived from diffusion MRI. *NeuroImage*, 120, 441–455. <https://doi.org/10.1016/j.neuroimage.2015.06.068>

- Pierce, M., McManus, S., Jessop, C., John, A., Hotopf, M., Ford, T., Hatch, S., Wessely, S., & Abel, K. M. (2020). Says who? The significance of sampling in mental health surveys during COVID-19. *The Lancet Psychiatry*, 7(7), 567–568. [https://doi.org/10.1016/S2215-0366\(20\)30237-6](https://doi.org/10.1016/S2215-0366(20)30237-6)
- Poldrack, R. A. (2010). Interpreting developmental changes in neuroimaging signals. *Human Brain Mapping*, 31(6), 872–878. <https://doi.org/10.1002/hbm.21039>
- Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M.-Y., Gorgolewski, K. J., Luci, J., Joo, S. J., Boyd, R. L., Hunicke-Smith, S., Simpson, Z. B., Caven, T., Sochat, V., Shine, J. M., Gordon, E., Snyder, A. Z., Adeyemo, B., Petersen, S. E., ... Mumford, J. A. (2015). Long-term neural and physiological phenotyping of a single human. *Nature Communications*, 6(1), 8885. <https://doi.org/10.1038/ncomms9885>
- Poldrack, R. A., Paré-Blagoev, E. J., & Grant, P. E. (2002). Pediatric Functional Magnetic Resonance Imaging: Progress and Challenges: Topics in Magnetic Resonance Imaging, 13(1), 61–70. <https://doi.org/10.1097/00002142-200202000-00005>
- Poline, J.-B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., Haselgrove, C., Helmer, K. G., Keator, D. B., Marcus, D. S., Poldrack, R. A., Schwartz, Y., Ashburner, J., & Kennedy, D. N. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*, 6. <https://doi.org/10.3389/fninf.2012.00009>
- Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59(3), 2142–2154. <https://doi.org/10.1016/j.neuroimage.2011.10.018>
- Prior, F. W., Brunnsden, B., Hildebolt, C., Nolan, T. S., Pringle, M., Vaishnavi, S. N., & Larson-Prior, L. J. (2009). Facial Recognition From Volume-Rendered Magnetic Resonance Imaging Data. *IEEE Transactions on Information Technology in Biomedicine*, 13(1), 5–9. <https://doi.org/10.1109/TITB.2008.2003335>

Q

R

- Raman, S., & Hodes, D. (2012). Cultural issues in child maltreatment: Cultural issues in child maltreatment. *Journal of Paediatrics and Child Health*, 48(1), 30–37. <https://doi.org/10.1111/j.1440-1754.2011.02184.x>
- Reuter, M., Schmansky, N. J., Rosas, H. D., & Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage*, 61(4), 1402–1418. <https://doi.org/10.1016/j.neuroimage.2012.02.084>
- Roalf, D. R., Quarmley, M., Elliott, M. A., Satterthwaite, T. D., Vandekar, S. N., Ruparel, K., Gennatas, E. D., Calkins, M. E., Moore, T. M., Hopson, R., Prabhakaran, K., Jackson, C. T., Verma, R., Hakonarson, H., Gur, R. C., & Gur, R. E. (2016). The impact of quality assurance assessment

on diffusion tensor imaging outcomes in a large-scale population-based cohort. *NeuroImage*, 125, 903–919. <https://doi.org/10.1016/j.neuroimage.2015.10.068>

Ruba, A. L., & Pollak, S. D. (2020). The development of emotion reasoning in infancy and early childhood. *Annual Review of Developmental Psychology*, 2, 503–531.

S

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., Gennatas, E. D., Jackson, C., Prabhakaran, K., Smith, A., Hakonarson, H., Verma, R., Davatzikos, C., Gur, R. E., & Gur, R. C. (2013). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *NeuroImage*, 83, 45–57. <https://doi.org/10.1016/j.neuroimage.2013.06.045>

Savalia, N. K., Agres, P. F., Chan, M. Y., Feczko, E. J., Kennedy, K. M., & Wig, G. S. (2017). Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. *Human Brain Mapping*, 38(1), 472–492. <https://doi.org/10.1002/hbm.23397>

Scheinost, D., Pollatou, A., Dufford, A. J., Jiang, R., Farruggia, M. C., Rosenblatt, M., ... & Westwater, M. L. (2023). Machine learning and prediction in fetal, infant, and toddler neuroimaging: A review and primer. *Biological psychiatry*, 93(10), 893–904.

Schimke, N., & Hale, J. (2011). Quickshear Defacing for Neuroimages. 5.

Schmitt, J. E., Neale, M. C., Fassassi, B., Perez, J., Lenroot, R. K., Wells, E. M., & Giedd, J. N. (2014). The dynamic role of genetics on cortical patterning during childhood and adolescence. *Proceedings of the National Academy of Sciences*, 111(18), 6774–6779. <https://doi.org/10.1073/pnas.1311630111>

Schnack, H. G. (2019). Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophrenia Research*, 214, 34–42. <https://doi.org/10.1016/j.schres.2017.10.023>

Schnack, H. G., van Haren, N. E. M., Brouwer, R. M., Evans, A., Durston, S., Boomsma, D. I., Kahn, R. S., & Hulshoff Pol, H. E. (2015). Changes in Thickness and Surface Area of the Human Cortex and Their Relationship with Intelligence. *Cerebral Cortex*, 25(6), 1608–1617. <https://doi.org/10.1093/cercor/bht357>

Schreuders, E., Smeekens, S., Cillessen, A. H. N., & Güroğlu, B. (2019). Friends and foes: Neural correlates of prosocial decisions with peers in adolescence. *Neuropsychologia*, 129, 153–163. <https://doi.org/10.1016/j.neuropsychologia.2019.03.004>

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P. J., Dalley, J. W., Flor, H., Gallinat, J., Garavan, H., Heinz, A., Itterman, B., Lathrop, M., Mallik, C., Mann, K., Martinot, J.-L., Paus, T., Poline, J.-B., ... the IMAGEN consortium. (2010). The IMAGEN study: Reinforcement-related behaviour in normal brain function and

psychopathology. *Molecular Psychiatry*, 15(12), 1128–1139.
<https://doi.org/10.1038/mp.2010.4>

Schwarz, C. G., Kremers, W. K., Therneau, T. M., Sharp, R. R., Gunter, J. L., Vemuri, P., Arani, A., Spychalla, A. J., Kantarci, K., Knopman, D. S., Petersen, R. C., & Jack, C. R. (2019). Identification of Anonymous MRI Research Participants with Face-Recognition Software. *New England Journal of Medicine*, 381(17), 1684–1686.
<https://doi.org/10.1056/NEJMc1908881>

Schwarz, C. G., Kremers, W. K., Wiste, H. J., Gunter, J. L., Vemuri, P., Spychalla, A. J., Kantarci, K., Schultz, A. P., Sperling, R. A., Knopman, D. S., Petersen, R. C., & Jack, C. R. (2021). Changing the face of neuroimaging research: Comparing a new MRI de-facing technique with popular alternatives. *NeuroImage*, 231, 117845.
<https://doi.org/10.1016/j.neuroimage.2021.117845>

Schwarzenberg, S. J., Georgieff, M. K., Daniels, S., Corkins, M., Golden, N. H., Kim, J. H., ... & Magge, S. N. (2018). Advocacy for improving nutrition in the first 1000 days to support childhood development and adult health. *Pediatrics*, 141(2).

Shah, L. M., Cramer, J. A., Ferguson, M. A., Birn, R. M., & Anderson, J. S. (2016). Reliability and reproducibility of individual differences in functional connectivity acquired during task and resting state. *Brain and Behavior*, 6(5). <https://doi.org/10.1002/brb3.456>

Sheridan, M. A., & McLaughlin, K. A. (2022). Introduction to the special issue on childhood adversity and neurodevelopment. *Developmental Cognitive Neuroscience*, 54, 101082.
<https://doi.org/10.1016/j.dcn.2022.101082>

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, 31(4), 1487–1505. <https://doi.org/10.1016/j.neuroimage.2006.02.024>

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>

Smith, K. E., & Pollak, S. D. (2021). Rethinking concepts and categories for understanding the neurodevelopmental effects of childhood adversity. *Perspectives on psychological science*, 16(1), 67-93.

Staff, A. I., Luman, M., van der Oord, S., Bergwerff, C. E., van den Hoofdakker, B. J., & Oosterlaan, J. (2022). Facial emotion recognition impairment predicts social and emotional problems in children with (subthreshold) ADHD. *European Child & Adolescent Psychiatry*, 31(5), 715–727. <https://doi.org/10.1007/s00787-020-01709-y>

- Stinson, E. A., Sullivan, R. M., Petee, B. J., Tapert, S. F., Baker, F. C., Breslin, F. J., ... & Lisdahl, K. M. (2021). Longitudinal impact of childhood adversity on early adolescent mental health during the COVID-19 pandemic in the ABCD study cohort: does race or ethnicity moderate findings?. *Biological psychiatry global open science*, 1(4), 324-335.
- St Pourcain, B., Haworth, C. M. A., Davis, O. S., Wang, K., Timpson, N. J., Evans, D. M., ... & Davey Smith, G. (2015). Heritability and genome-wide analyses of problematic peer relationships during childhood and adolescence. *Human genetics*, 134, 539-551.
- Streitbürger, D.-P., Möller, H. E., Tittgemeyer, M., Hund-Georgiadis, M., Schroeter, M. L., & Mueller, K. (2012). Investigating Structural Brain Changes of Dehydration Using Voxel-Based Morphometry. *PLoS ONE*, 7(8), e44195. <https://doi.org/10.1371/journal.pone.0044195>
- Stuhrmann, A., Suslow, T., & Dannlowski, U. (2011). Facial emotion processing in major depression: A systematic review of neuroimaging findings. *Biology of Mood & Anxiety Disorders*, 1(1), 10. <https://doi.org/10.1186/2045-5380-1-10>
- Sui, J., Adali, T., Yu, Q., Chen, J., & Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*, 204(1), 68–81. <https://doi.org/10.1016/j.jneumeth.2011.10.031>
- Svatkova, A., Mandl, R. C. W., Scheewe, T. W., Cahn, W., Kahn, R. S., & Hulshoff Pol, H. E. (2015). Physical Exercise Keeps the Brain Connected: Biking Increases White Matter Integrity in Patients With Schizophrenia and Healthy Controls. *Schizophrenia Bulletin*, 41(4), 869–878. <https://doi.org/10.1093/schbul/sbv033>
- Swagerman, S. C., de Geus, E. J. C., Kan, K.-J., van Bergen, E., Nieuwboer, H. A., Koenis, M. M. G., Hulshoff Pol, H. E., Gur, R. E., Gur, R. C., & Boomsma, D. I. (2016). The Computerized Neurocognitive Battery: Validation, aging effects, and heritability across cognitive domains. *Neuropsychology*, 30(1), 53–64. <https://doi.org/10.1037/neu0000248>
- Székely, E., Tiemeier, H., Jaddoe, V. W. V., Hofman, A., Verhulst, F. C., & Herba, C. M. (2014). Associations of Internalizing and Externalizing Problems with Facial Expression Recognition in Preschoolers: The Generation R Study: Problem Behavior and Emotion Recognition in Preschoolers. *Social Development*, 23(3), 611–630. <https://doi.org/10.1111/sode.12070>

T

- Tamnes, C. K., Herting, M. M., Goddings, A.-L., Meuwese, R., Blakemore, S.-J., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., Crone, E. A., & Mills, K. L. (2017). Development of the Cerebral Cortex across Adolescence: A Multisample Study of Inter-Related Longitudinal Changes in Cortical Volume, Surface Area, and Thickness. *The Journal of Neuroscience*, 37(12), 3402–3412. <https://doi.org/10.1523/JNEUROSCI.3302-16.2017>
- Tamnes, C. K., Walhovd, K. B., Dale, A. M., Østby, Y., Grydeland, H., Richardson, G., Westlye, L. T., Roddey, J. C., Hagler, D. J., Due-Tønnessen, P., Holland, D., & Fjell, A. M. (2013). Brain

development and aging: Overlapping and unique patterns of change. *NeuroImage*, 68, 63–74. <https://doi.org/10.1016/j.neuroimage.2012.11.039>

- Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., Tyler, L. K., Cam-CAN, & Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, 144, 262–269. <https://doi.org/10.1016/j.neuroimage.2015.09.018>
- Teeuw, J., Brouwer, R. M., Guimarães, J. P. O. F. T., Brandner, P., Koenis, M. M. G., Swagerman, S. C., Verwoert, M., Boomsma, D. I., & Hulshoff Pol, H. E. (2019). Genetic and environmental influences on functional connectivity within and between canonical cortical resting-state networks throughout adolescent development in boys and girls. *NeuroImage*, 202, 116073. <https://doi.org/10.1016/j.neuroimage.2019.116073>
- Teicher, M. H., & Samson, J. A. (2016). Annual Research Review: Enduring neurobiological effects of childhood abuse and neglect. *Journal of Child Psychology and Psychiatry*, 57(3), 241–266. <https://doi.org/10.1111/jcpp.12507>
- Telzer, E. H., McCormick, E. M., Peters, S., Cosme, D., Pfeifer, J. H., & van Duijvenvoorde, A. C. K. (2018). Methodological considerations for developmental longitudinal fMRI research. *Developmental Cognitive Neuroscience*, 33, 149–160. <https://doi.org/10.1016/j.dcn.2018.02.004>
- Termenon, M., Jaillard, A., Delon-Martin, C., & Achard, S. (2016). Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project. *NeuroImage*, 142, 172–187. <https://doi.org/10.1016/j.neuroimage.2016.05.062>
- Thomas, K. M., King, S. W., Franzen, P. L., Welsh, T. F., Berkowitz, A. L., Noll, D. C., Birmaher, V., & Casey, B. J. (1999). A Developmental Functional MRI Study of Spatial Working Memory. *NeuroImage*, 10(3), 327–338. <https://doi.org/10.1006/nimg.1999.0466>
- Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M. J., Martin, N. G., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N. C., ... Alzheimer's Disease Neuroimaging Initiative, EPIGEN Consortium, IMAGEN Consortium, Saguenay Youth Study (SYS) Group. (2014). The ENIGMA Consortium: Large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging and Behavior*, 8(2), 153–182. <https://doi.org/10.1007/s11682-013-9269-5>
- Tottenham, N., & Sheridan. (2009). A review of adversity, the amygdala and the hippocampus: A consideration of developmental timing. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/neuro.09.068.2009>
- Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202, 116137. <https://doi.org/10.1016/j.neuroimage.2019.116137>

- Tremblay, M.-P., Deschamps, I., Tousignant, B., & Jackson, P. L. (2022). Functional connectivity patterns of trait empathy are associated with age. *Brain and Cognition*, 159, 105859. <https://doi.org/10.1016/j.bandc.2022.105859>
- Trentacosta, C. J., & Fine, S. E. (2010). Emotion Knowledge, Social Competence, and Behavior Problems in Childhood and Adolescence: A Meta-analytic Review. *Social Development*, 19(1), 1–29. <https://doi.org/10.1111/j.1467-9507.2009.00543.x>
- Tulay, E. E., Metin, B., Tarhan, N., & Arıkan, M. K. (2019). Multimodal neuroimaging: basic concepts and classification of neuropsychiatric diseases. *Clinical EEG and neuroscience*, 50(1), 20–33.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1), 273–289. <https://doi.org/10.1006/nimg.2001.0978>

U

V

- van den Berg, G., Donker, A., van Hummel, N., Tuenter, T., Branje, S., Finkenauer, C., & Polderman, T. (2023). Mentaal welbevinden van de jeugd: lessen uit de coronacrisis. Een nieuw overzicht van de onderzoeksliteratuur. Nederlands Jeugdinstituut. <https://www.nji.nl/publicaties/mentaal-welbevinden-van-de-jeugd-lessen-uit-de-coronacrisis>
- van der Laan, S. E. I., Finkenauer, C., Lenters, V. C., van Harmelen, A.-L., van der Ent, C. K., & Nijhof, S. L. (2021). Gender-Specific Changes in Life Satisfaction After the COVID-19–Related Lockdown in Dutch Adolescents: A Longitudinal Study. *Journal of Adolescent Health*, 69(5), 737–745. <https://doi.org/10.1016/j.jadohealth.2021.07.013>
- van der Meulen, M., Wierenga, L. M., Achterberg, M., Drenth, N., van IJzendoorn, M. H., & Crone, E. A. (2020). Genetic and environmental influences on structure of the social brain in childhood. *Developmental Cognitive Neuroscience*, 44, 100782.
- Van Dijk, K. R. A., Sabuncu, M. R., & Buckner, R. L. (2012). The influence of head motion on intrinsic functional connectivity MRI. *NeuroImage*, 59(1), 431–438. <https://doi.org/10.1016/j.neuroimage.2011.07.044>
- van Harmelen, A. L., Blakemore, S. J., Goodyer, I. M., & Kievit, R. A. (2021). The Interplay Between Adolescent Friendship Quality and Resilient Functioning Following Childhood and Adolescent Adversity. *Adversity and Resilience Science*, 2(1), 37–50. <https://doi.org/10.1007/s42844-020-00027-1>
- van Harmelen, A. L., Kievit, R. A., Ioannidis, K., Neufeld, S., Jones, P. B., Bullmore, E., Dolan, R., The NSPN Consortium, Fonagy, P., & Goodyer, I. (2017). Adolescent friendships predict later

- resilient functioning across psychosocial domains in a healthy community cohort. *Psychological Medicine*, 47(13), 2312–2322. <https://doi.org/10.1017/S0033291717000836>
- van Soelen, I. L. C., Brouwer, R. M., Peper, J. S., van Leeuwen, M., Koenis, M. M. G., van Beijsterveldt, T. C. E. M., Swagerman, S. C., Kahn, R. S., Hulshoff Pol, H. E., & Boomsma, D. I. (2012a). Brain SCALE: Brain Structure and Cognition: an Adolescent Longitudinal Twin Study into the Genetic Etiology of Individual Differences. *Twin Research and Human Genetics*, 15(3), 453–467. <https://doi.org/10.1017/thg.2012.4>
- van Soelen, I. L. C., Brouwer, R. M., van Baal, G. C. M., Schnack, H. G., Peper, J. S., Chen, L., Kahn, R. S., Boomsma, D. I., & Pol, H. E. H. (2011). Heritability of volumetric brain changes and height in children entering puberty. *Human Brain Mapping*, n/a-n/a. <https://doi.org/10.1002/hbm.21468>
- van Soelen, I. L. C., Brouwer, R. M., van Baal, G. C. M., Schnack, H. G., Peper, J. S., Collins, D. L., Evans, A. C., Kahn, R. S., Boomsma, D. I., & Hulshoff Pol, H. E. (2012b). Genetic influences on thinning of the cerebral cortex during development. *NeuroImage*, 59(4), 3871–3880. <https://doi.org/10.1016/j.neuroimage.2011.11.044>
- Veldkamp, C. L. S., & Kemner, C. (2023). Sharing is caring for developmental psychology. *Infant and Child Development*. Online Version of Record before inclusion in an issue, e2438. <https://doi.org/10.1002/icd.2438>.
- Veraart, J., Novikov, D. S., Christiaens, D., Ades-aron, B., Sijbers, J., & Fieremans, E. (2016). Denoising of diffusion MRI using random matrix theory. *NeuroImage*, 142, 394–406. <https://doi.org/10.1016/j.neuroimage.2016.08.016>
- Verpaalen, I. A. M., Bijsterbosch, G., Mobach, L., Bijlstra, G., Rinck, M., & Klein, A. M. (2019). Validating the Radboud faces database from a child's perspective. *Cognition and Emotion*, 33(8), 1531–1547. <https://doi.org/10.1080/02699931.2019.1577220>
- Vetter, N. C., Pilhatsch, M., Weigelt, S., Ripke, S., & Smolka, M. N. (2015). Mid-adolescent neurocognitive development of ignoring and attending emotional stimuli. *Developmental Cognitive Neuroscience*, 14, 23–31. <https://doi.org/10.1016/j.dcn.2015.05.001>
- Vetter, N. C., Steding, J., Jurk, S., Ripke, S., Mennigen, E., & Smolka, M. N. (2017). Reliability in adolescent fMRI within two years – a comparison of three tasks. *Scientific Reports*, 7(1), 2287. <https://doi.org/10.1038/s41598-017-02334-7>
- Vijayakumar, N., Mills, K. L., Alexander-Bloch, A., Tamnes, C. K., & Whittle, S. (2018). Structural brain development: A review of methodological approaches and best practices. *Developmental Cognitive Neuroscience*, 33, 129–148. <https://doi.org/10.1016/j.dcn.2017.11.008>
- Vink, M., Zandbelt, B. B., Gladwin, T., Hillegers, M., Hoogendam, J. M., van den Wildenberg, W. P. M., Du Plessis, S., & Kahn, R. S. (2014). Frontostriatal activity and connectivity increase during proactive inhibition across adolescence and early adulthood: Activity and Functional Connectivity of Frontostriatal Network. *Human Brain Mapping*, 35(9), 4415–4427. <https://doi.org/10.1002/hbm.22483>

Vink, R. M., van Dommelen, P., van der Pal, S. M., Eekhout, I., Pannebakker, F. D., Klein Velderman, M., Haagmans, M., Mulder, T., & Dekker, M. (2019). Self-reported adverse childhood experiences and quality of life among children in the two last grades of Dutch elementary education. *Child Abuse & Neglect*, 95, 104051. <https://doi.org/10.1016/j.chiabu.2019.104051>

W

Waheed, S. H., Mirbagheri, S., Agarwal, S., Kamali, A., Yahyavi-Firouz-Abadi, N., Chaudhry, A., DiGianvittorio, M., Gujar, S. K., Pillai, J. J., & Sair, H. I. (2016). Reporting of Resting-State Functional Magnetic Resonance Imaging Preprocessing Methodologies. *Brain Connectivity*, 6(9), 663–668. <https://doi.org/10.1089/brain.2016.0446>

Warrier, V., Toro, R., Chakrabarti, B., iPSYCH-Broad Autism Group, Børglum, A. D., Grove, J., ... & Baron-Cohen, S. (2018). Genome-wide analyses of self-reported empathy: correlations with autism, schizophrenia, and anorexia nervosa. *Translational psychiatry*, 8(1), 35.

Warrier, V., Kwong, A. S. F., Luo, M., Dalvie, S., Croft, J., Sallis, H. M., Baldwin, J., Munafò, M. R., Nievergelt, C. M., Grant, A. J., Burgess, S., Moore, T. M., Barzilay, R., McIntosh, A., van IJzendoorn, M. H., & Cecil, C. A. M. (2021). Gene–environment correlations and causal effects of childhood maltreatment on physical and mental health: A genetically informed approach. *The Lancet Psychiatry*, 8(5), 373–386. [https://doi.org/10.1016/S2215-0366\(20\)30569-1](https://doi.org/10.1016/S2215-0366(20)30569-1)

Weisskoff, R. M. (1996). Simple measurement of scanner stability for functional NMR imaging of activation in the brain. *Magnetic Resonance in Medicine*, 36(4), 643–645. <https://doi.org/10.1002/mrm.1910360422>

Welton, T., Kent, D. A., Auer, D. P., & Dineen, R. A. (2015). Reproducibility of Graph-Theoretic Brain Network Metrics: A Systematic Review. *Brain Connectivity*, 5(4), 193–202. <https://doi.org/10.1089/brain.2014.0313>

Wendelken, C., Ferrer, E., Ghetti, S., Bailey, S. K., Cutting, L., & Bunge, S. A. (2017). Frontoparietal Structural Connectivity in Childhood Predicts Development of Functional Connectivity and Reasoning Ability: A Large-Scale Longitudinal Investigation. *The Journal of Neuroscience*, 37(35), 8549–8558. <https://doi.org/10.1523/JNEUROSCI.3726-16.2017>

White, T., Blok, E., & Calhoun, V. D. (2022). Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Human Brain Mapping*, 43(1), 278–291. <https://doi.org/10.1002/hbm.25120>

White, T., Marroun, H. E., Nijs, I., Schmidt, M., van der Lugt, A., Wielopolki, P. A., Jaddoe, V. W. V., Hofman, A., Krestin, G. P., Tiemeier, H., & Verhulst, F. C. (2013). Pediatric population-based neuroimaging and the Generation R Study: The intersection of developmental neuroscience and epidemiology. *European Journal of Epidemiology*, 28(1), 99–111. <https://doi.org/10.1007/s10654-013-9768-0>

- Whitfield-Gabrieli, S., & Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, 2(3), 125–141. <https://doi.org/10.1089/brain.2012.0073>
- Wilke, M., Holland, S. K., Altaye, M., & Gaser, C. (2008). Template-O-Matic: A toolbox for creating customized pediatric templates. *NeuroImage*, 41(3), 903–913. <https://doi.org/10.1016/j.neuroimage.2008.02.056>
- Wilke, M., Schmithorst, V. J., & Holland, S. K. (2002). Assessment of spatial normalization of whole-brain magnetic resonance images in children. *Human Brain Mapping*, 17(1), 48–60. <https://doi.org/10.1002/hbm.10053>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Winters, D. E., Pruitt, P. J., Fukui, S., Cyders, M. A., Pierce, B. J., Lay, K., & Damoiseaux, J. S. (2021). Network functional connectivity underlying dissociable cognitive and affective components of empathy in adolescence. *Neuropsychologia*, 156, 107832. <https://doi.org/10.1016/j.neuropsychologia.2021.107832>
- Wonderlick, J., Ziegler, D., Hosseinivarnamkhasti, P., Locascio, J., Bakkour, A., Vanderkouwe, A., Triantafyllou, C., Corkin, S., & Dickerson, B. (2009). Reliability of MRI-derived cortical and subcortical morphometric measures: Effects of pulse sequence, voxel geometry, and parallel imaging. *NeuroImage*, 44(4), 1324–1333. <https://doi.org/10.1016/j.neuroimage.2008.10.037>
- Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. *NeuroImage*, 91, 412–419. <https://doi.org/10.1016/j.neuroimage.2013.12.058>
- X**
- Y**
- Yan, C.-G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., Li, Q., Zuo, X.-N., Castellanos, F. X., & Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, 76, 183–201. <https://doi.org/10.1016/j.neuroimage.2013.03.004>
- Yap, P.-T., Fan, Y., Chen, Y., Gilmore, J. H., Lin, W., & Shen, D. (2011). Development Trends of White Matter Connectivity in the First Years of Life. *PLoS ONE*, 6(9), e24678. <https://doi.org/10.1371/journal.pone.0024678>

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>

Z

Zandbelt, B. B., Bloemendaal, M., Neggers, S. F. W., Kahn, R. S., & Vink, M. (2013). Expectations and violations: Delineating the neural network of proactive inhibitory control: Neural Network of Proactive Inhibition. *Human Brain Mapping*, 34(9), 2015–2024. <https://doi.org/10.1002/hbm.22047>

Zandbelt, B. B., & Vink, M. (2010). On the Role of the Striatum in Response Inhibition. *PLoS ONE*, 5(11), e13848. <https://doi.org/10.1371/journal.pone.0013848>

Zanolie, K., Ma, I., Bos, M. G. N., Schreuders, E., Vandenbroucke, A. R. E., van Hoorn, J., van Duijvenvoorde, A. C. K., Wierenga, L., Crone, E. A., & Güroğlu, B. (2022). Understanding the Dynamics of the Developing Adolescent Brain Through Team Science. *Frontiers in Integrative Neuroscience*, 16, 827097. <https://doi.org/10.3389/fnint.2022.827097>

Zijlmans, J., Luijten, M., van Oers, H., Tieskens, J., Alrouh, H., Haverman, L., ... & Dreams Consortium. (2022). 3.78 Mental Health Problems During the COVID-19 Pandemic in Dutch Children and Adolescents With and Without Pre-existing Mental Health Problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, 61(10), S253.

Zondergeld, J. J., Scholten, R. H. H., Vreede, B. M. I., Hessels, R. S., Pijl, A. G., Buizer-Voskamp, J. E., Rasch, M., Lange, O. A., & Veldkamp, C. L. S. (2020). FAIR, safe and high-quality data: The data infrastructure and accessibility of the YOUth cohort study. *Developmental Cognitive Neuroscience*, 45, 100834. <https://doi.org/10.1016/j.dcn.2020.100834>

A1 - Nederlandse wetenschappelijke samenvatting

Algemene introductie

Hersenontwikkeling wordt gedreven door interacties tussen genen en de omgeving. De wisselwerking tussen hersenontwikkeling, genen en omgeving vormt de basis voor de variaties die worden waargenomen tussen individuen. Het bestuderen van hersenscans van het ontwikkelende brein stelt onderzoekers in staat om een dieper inzicht te krijgen in de factoren die jou maken tot wie je bent. De pre-adolescentie is een interessante periode van de ontwikkeling. Bij een deel van de kinderen beginnen tijdens de pre-adolescentie puberteitshormonen een rol te spelen (Koenis et al., 2013; Pas et al., 2021). Daarnaast ontwikkelen kinderen zich op sociaal vlak (Blakemore, 2012; Junge et al., 2020). Psychosociale klachten ontstaan ook vaak tijdens de (pre-)adolescentie of verergeren in die periode (Paus et al., 2008). Al deze ontwikkelingen gaan samen met veranderingen in de structuur van de hersenen.

Op verschillende plekken in de wereld zijn de afgelopen jaren studies gestart die hersenontwikkeling in kaart brengen in grote groepen kinderen. Deze cohort studies laten zien dat de hersenen snel groeien tijdens de zwangerschap en de eerste jaren na de geboorte (Dubois et al., 2021). Tijdens de kindertijd en de adolescentie groeien de subcorticale gebieden (Mills et al., 2021), wordt de hersenschors dunner (Teeuw et al., 2019), en laat de oppervlakte van de hersenschors eerst groei zien gevolgd door krimp terwijl het volume van de witte stof door blijft groeien (Frangou et al., 2022; Giedd et al., 1999; Koenis et al., 2015; Tamnes et al., 2017). Door trajecten van hersenontwikkeling in kaart te brengen, kan gezocht worden naar waarom individuen in het latere leven verschillen, bijvoorbeeld op het gebied van cognitie (Schnack et al., 2015) of kwetsbaarheid voor psychiatrische aandoeningen (Paus et al., 2008).

Ingrijpende jeugdervaringen blijken een belangrijke risico-factor te zijn voor psychosociale problemen tijdens de volwassenheid (Green et al., 2010; Kessler et al., 2010; McLaughlin, 2016). Aan de andere kant kan de sociale omgeving ook een beschermende werking hebben (van Harmelen et al., 2017, 2021). De omgeving waarin een kind opgroeit biedt dus aanknopingspunten voor het beter begrijpen van mechanismen van psychische kwetsbaarheid.

Dit proefschrift

Dit proefschrift richt zich op verschillen tussen individuen in de structuur en functie van de hersenen in de pre-adolescentie. Het proefschrift bestaat uit meerdere deelstudies die verschillende onderwerpen behandelen.

Om naar subtiele verschillen in hersenmaten tussen kinderen te kunnen kijken, is het belangrijk om eerst te onderzoeken hoe betrouwbaar deze hersenmaten zijn. Deze betrouwbaarheid kan onderzocht worden door bij een groep mensen twee keer dezelfde hersenscan te maken. Vervolgens kan dan gekwantificeerd worden wat het verschil is tussen de hersenmaten verworven tijdens de eerste en de tweede sessie (test-hertest betrouwbaarheid). De eerste twee deelstudies staan in het teken van deze test-hertest betrouwbaarheid en zijn tevens belangrijk in het kader van het delen van onderzoeksdata met andere wetenschappers om de wetenschap vooruit te helpen (*open science*). Transparantie over het onderzoeksprotocol, waarborging van de privacy van deelnemers en goede kwaliteit van de data staan in deze deelstudies centraal.

Nadat de betrouwbaarheid van de hersenmaten is onderzocht, focussen de laatste twee deelstudies op de relatie tussen de hersenen en de sociale omgeving. Hiervoor is data gebruikt van 1000 kinderen (8-, 9- en 10-jarigen) die deel hebben genomen aan het YOUth-onderzoek (Onland et al., 2020) in de periode van 2016 tot begin 2020. Eerst is onderzocht of er een verband is tussen ingrijpende jeugdervaringen en de structuur van de hersenen. Dit is belangrijk, gegeven de eerder besproken link tussen ingrijpende jeugdervaringen en mentale problemen tijdens de volwassenheid. Daarna is gekeken of de neurale verwerking van foto's van gezichten met een

emotionele expressie (neutraal, blij of bang) samenhangt met sociaal functioneren en het identificeren van emoties op gezichten. Dit is belangrijk, omdat een beter begrip van mechanismen van sociaal functioneren kan helpen om kinderen te ondersteunen die hier moeite mee ervaren.

Test-hertest betrouwbaarheid

De eerste deelstudie in dit proefschrift (Buimer et al., 2020; **Hoofdstuk 2**) beschrijft het MRI-protocol gebruikt in het YOUth-onderzoek en de test-hertest betrouwbaarheid van hersenmaten die verkregen kunnen worden met dit protocol. Dit onderzoek laat zien dat verschillende MRI-scans een verschillende test-hertest betrouwbaarheid hebben en anatomische hersenmaten betrouwbaarder gemeten kunnen worden dan functionele hersenmaten. Variatie tussen beide metingen kan bijvoorbeeld ontstaan door meetruis of beweging tijdens het maken van de hersenscan. Bij het meten van hersenactiviteit spelen meer factoren mee dan bij anatomie, zoals of de deelnemer zich tijdens beide sessies even goed kan concentreren. De tweede deelstudie (Buimer et al., 2021; **Hoofdstuk 3**) evalueert de-identificatiemethoden voor MRI-scans (het verwijderen van tot de persoon herleidbare kenmerken op anatomische scans) en vindt minimale effecten van deze technieken op de uitkomstmaten.

Ingrijpende jeugdervaringen en de structuur van de hersenen

In de derde deelstudie (Buimer et al., 2022; **Hoofdstuk 4**) wordt het verband tussen ingrijpende jeugdervaringen en hersenstructuur onderzocht. Dit onderzoek laat zien dat kinderen die opgroeien in een gezin waar alcohol of drugsproblematiek speelt, een grotere oppervlakte van de hersenschors hebben in frontale gebieden. Dit is in overeenstemming met een grote Amerikaanse studie van een cohort van dezelfde leeftijd (Lees et al., 2021). Daarnaast laat ik zien, dat blootstelling aan geweld of criminaliteit in het gezin dan wel de buurt waarin het kind opgroeit, samenhangt met een lagere witte stof integriteit in een gedeelte van het cingulum.



Figuur 1. Samenvatting van de hoofdeffecten in het onderzoek naar de relatie tussen ingrijpende jeugdervaringen en hersenstructuur.

Een belangrijke kanttekening bij dit onderzoek is dat de effecten klein waren en voor veel ingrijpende jeugdervaringen geen effect is gevonden. Daarnaast is niet onderzocht of de verschillen in de anatomie van de hersenen bij kinderen met deze ingrijpende jeugdervaringen samenhangen met hun functioneren. In de toekomst zou het interessant zijn om te kijken naar de relatie tussen deze neurobiologische correlaten en de kwetsbaarheid voor psychopathologie in het latere leven. Een interessante nieuwe richting is om te focussen op verschillen tussen mensen in hun mate van veerkrachtig functioneren na ingrijpende jeugdervaringen. Zo wordt er nu onderzoek gedaan naar de neurobiologische correlaten van veerkrachtig functioneren na jeugdtrauma (González-García et al., 2023).

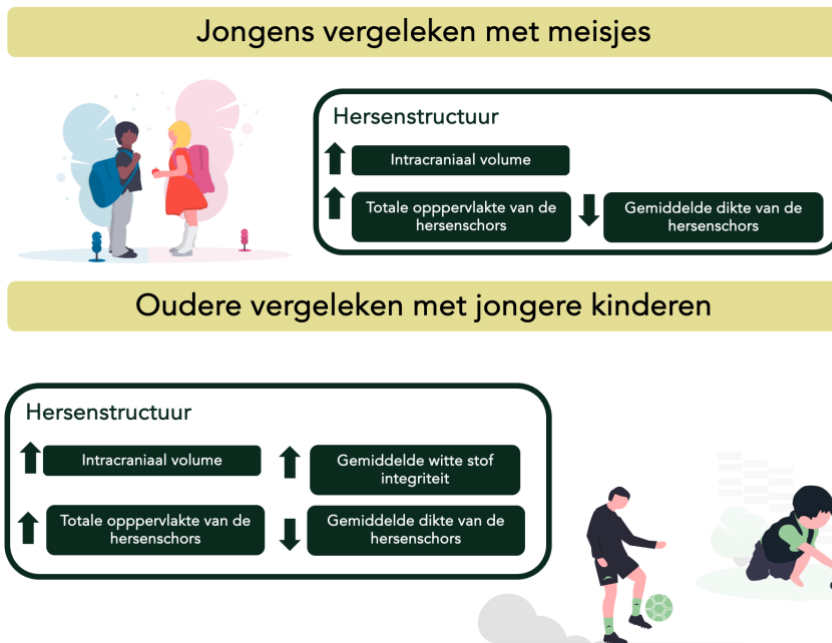
Sociale cognitie en hersenactiviteit

De vierde deelstudie (**Hoofdstuk 5**) onderzoekt de onderlinge relatie van factoren onderliggend aan sociale cognitie. Deze factoren waren: 1) het correct en snel identificeren van emoties, 2) hersenactiviteit tijdens het verwerken van emoties op gezichten en 3) sociale vaardigheid. In tegenstelling tot mijn verwachting vond ik geen bewijs voor een sterk verband tussen deze factoren.

Effecten van leeftijd en geslacht op primaire uitkomstmaten

De statistische modellen in **Hoofdstuk 4** en **Hoofdstuk 5** zijn telkens gecorrigeerd voor leeftijd en geslacht. Maar wat voor leeftijds- en geslachtseffecten zijn er eigenlijk geobserveerd in deze studies?

In **Hoofdstuk 4** laat ik zien dat er veel variatie is in de hersenstructuur tussen kinderen onderling en dat effecten van leeftijd en geslacht hier een grote rol in spelen (Figuur 2). Leeftijd associeert positief met de volgende globale hersenmaten: intracraniaal volume, de oppervlakte van de hersenschors en witte stof integriteit. Een negatief verband met leeftijd werd gevonden voor corticale dikte. Wanneer jongens en meisjes met elkaar worden vergeleken valt op dat jongens een groter intracraniaal volume hebben, een grotere corticale oppervlakte en een dunnere cortex dan meisjes. Er werd geen effect van geslacht gevonden op globale witte stof integriteit.



Figuur 2. Samenvatting van de effecten van leeftijd en geslacht op globale maten van hersenstructuur.

In **Hoofdstuk 5** laat ik de effecten van leeftijd en geslacht op de identificatie van emoties en sociale vaardigheid zien (Figuur 3). Meisjes en oudere kinderen waren over het algemeen sneller en accurater in het correct identificeren van een emotie. Ook lieten meisjes betere sociale vaardigheden zien op elke gemeten subschaal. Een positieve verband met leeftijd werd gevonden voor de subschalen die meten hoe goed een kind het perspectief van een ander in kan nemen en hoeveel pro sociaal gedrag het kind vertoont. De effecten van leeftijd en geslacht op hersenactiviteit waren echter minimaal.

Voor het sociaal functioneren en het identificeren van emoties spelen leeftijds- en geslachtseffecten dus een grote rol bij kinderen tussen de 8 en 11 jaar oud. Ook variatie in de anatomie van de hersenen wordt deels verklaard door leeftijd en geslacht. Longitudinale data is nodig om uit te wijzen of de gevonden effecten van leeftijd en geslacht te verklaren zijn door (hersenen)ontwikkeling.



Figuur 3. Samenvatting van de effecten van leeftijd en geslacht op identificatie van emoties op gezichten en sociale vaardigheid.

A

Slotopmerking

Dit proefschrift benadrukt de complexiteit van ontwikkelingsneurowetenschap. Dit vakgebied vraagt om grote steekproeven en geavanceerde statistische methodes om kleine effecten te kunnen detecteren. Grote cohort studies wereldwijd bieden een schat aan informatie over het opgroeiende kind. Het veilig delen van data wordt steeds meer de norm. Uitdagingen voor de toekomst zijn het zorgen voor een deelnemerspopulatie die een goede weerspiegeling is van de maatschappij. Het onderzoeken van neurobiologische mechanismen die ervaringen tijdens de jeugd linken aan psychosociaal functioneren biedt in de toekomst mogelijk aanknopingspunten om opgroeiende kinderen beter te ondersteunen.

A2 - Open science statement

Availability of publications

All published articles in my PhD thesis are open-access and thus publicly available via the DOI-links on the chapter title pages. The last chapter is not published yet but will be submitted to an open access journal as well.

Availability of data and code

The YOUth cohort study encourages and facilitates bona fide use of its data. Researchers who wish to use YOUth data, can request the data for free: <https://www.uu.nl/en/research/youth-cohort-study/data-access>. Please note that anatomical T1-weighted MRI data will be subjected to a de-identification method. The raw test-retest data of young adults is not available, but the processed data can be requested via the UMC Utrecht: <https://www.umcutrecht.nl/en/data-request-form-umc-utrecht>. Requests for MRI scans of the data from older adults can be submitted here: <http://adni.loni.usc.edu/data-samples/access-data/>. Scripts and short README files will be available on my osf page: <http://doi.org/10.17605/OSF.IO/M5R3U>

Usage of open access tools and resources

I gratefully used many open access tools and resources in this PhD thesis. Some examples: 1) Illustrations from Katerina Limpitsouni's undraw.co; 2) Paul Tol's color palette for improved accessibility from <https://personal.sron.nl/~pault/>; 3) Coding using R and RStudio <https://www.r-project.org>; 4) Processing of neuroimaging data using FreeSurfer <http://freesurfer.net> and SPM <https://www.fil.ion.ucl.ac.uk/spm/>; 5) Visualizing neuroimaging data using the ENIGMA toolbox <https://enigma-toolbox.readthedocs.io>, MRICroGL & Surf Ice <https://www.nitrc.org/projects/mricrogl> & <https://www.nitrc.org/projects/surfice/>.

A3 - Scientific publications

Published

- González-García, N., **Buimer, E.E.L.**, Moreno-López, L., Sallie, S., Váša, F., Lim, S., . . . Van Harmelen, A. (2023). Resilient functioning is associated with altered structural brain network topology in adolescents exposed to childhood adversity. *Development and Psychopathology*, 1-11. doi: <https://doi.org/10.1017/S0954579423000901>
- Buimer, E.E.L.**, Brouwer, R.M., Mandl, R.C., Pas, P., Schnack, H.G., & Hulshoff Pol, H.E. (2022). Adverse childhood experiences and fronto-subcortical structures in the developing brain. *Frontiers in psychiatry*, 13. doi: <https://doi.org/10.3389/fpsy.2022.955871>
- Brouwer, R.M., Klein, M., Grasby, K.L., Schnack, H.G., Jahanshad, N., Teeuw, J., [and 196 others including **Buimer, E.E.L.**] & Hulshoff Pol, H.E. (2022). Age-dependent genetic variants associated with longitudinal changes in brain structure across the lifespan. *Nature Neuroscience*, 25, 421–432. doi: <https://doi.org/10.1038/s41593-022-01042-4>
- Janssen, J., Alloza, C., Díaz-Caneja, C. M., Santonja, J., Pina-Camacho, L., Gordaliza, P.M., [and 11 others, including **Buimer, E.E.L.**] & Schnack, H.G. (2022). Longitudinal allometry of sulcal morphology in health and schizophrenia. *Journal of Neuroscience*, 42(18), 3704-3715. doi: <https://doi.org/10.1523/JNEUROSCI.0606-21.2022>
- de Zwarte, S.M., Brouwer, R.M., Agartz, I., Alda, M., Alonso-Lana, S., Bearden, C.E., [and 95 others, including **Buimer, E.E.L.**] & van Haren, N.E. (2022). Intelligence, educational attainment, and brain structure in those at familial high-risk for schizophrenia or bipolar disorder. *Human brain mapping*, 43 (1), 414-430. doi: <https://doi.org/10.1002/hbm.25206>
- Buimer, E.E.L.**, Schnack, H.G., Caspi, Y, van Haren, N.E.M., Milchenko, M., Pas, P, Alzheimer's Disease Neuroimaging Initiative, Hulshoff Pol, H.E., Brouwer, R.M. (2021). De-identification procedures for magnetic resonance images and the impact on structural brain measures at different ages. *Human Brain Mapping*, 42(11), 3643-3655. doi: <https://doi.org/10.1002/hbm.25459>
- Janssen, J., Díaz-Caneja, C.M., Alloza, C., Schippers, A., De Hoyos, L., Santonja, J., [and 7 others, including **Buimer, E.E.L.**] & Schnack, H.G. (2021). Dissimilarity in sulcal width patterns in the cortex can be used to identify patients with schizophrenia with extreme deficits in cognitive performance. *Schizophrenia Bulletin*, 47(2), 552-561. doi: <https://doi.org/10.1093/schbul/sbaa131>
- Díaz-Caneja, C.M., Alloza, C., Gordaliza, P.M., Fernández-Pena, A., De Hoyos, L., Santonja, J., [and 7 others, including **Buimer, E.E.L.**] & Janssen, J. (2021). Sex differences in lifespan trajectories and variability of human sulcal and gyral morphology. *Cerebral Cortex*, 31(11), 5107-5120. doi: <https://doi.org/10.1093/cercor/bhab145>

- Buimer, E.E.L.***, Pas, P. *, Brouwer, R.M., Froeling, M., Hoogduin, H., Leemans, A., ... & Mandl, R.C. (2020). The YOUth cohort study: MRI protocol and test-retest reliability in adults. *Developmental cognitive neuroscience*, 45, 100816. doi: <https://doi.org/10.1016/j.dcn.2020.100816> * indicates joint first author
- Onland-Moret, N.C., Buizer-Voskamp, J.E., Albers, M.E., Brouwer, R.M., **Buimer, E.E.L.**, Hessels, R. S., ... & Kemner, C. (2020). The YOUth study: rationale, Design, and study procedures. *Developmental cognitive neuroscience*, 46, 100868. doi: <https://doi.org/10.1016/j.dcn.2020.100868>
- Grasby, K.L., Jahanshad, N., Painter, J.N., Colodro-Conde, L., Bralten, J., Hibar, D P., [and 361 others, including **Buimer, E.E.L.**] (2020). The genetic architecture of the human cerebral cortex. *Science*, 367(6484). doi: <https://doi.org/10.1126/science.aay6690>
- Hofer, E., Roshchupkin, G.V., Adams, H.H., Knol, M.J., Lin, H., Li, S., [and 92 others, including **Buimer, E.E.L.**, as part of the ENIGMA consortium] & Seshadri, S. (2020). Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nature communications*, 11(1), 1-16. doi: <https://doi.org/10.1038/s41467-020-18367-y>
- de Zwarte, S.M., Brouwer, R.M., Agartz, I., Alda, M., Aleman, A., Alpert, K.I., [and 84 others, including **Buimer, E.E.L.**] & van Haren, N.E. (2019). The association between familial risk and brain abnormalities is disease specific: An ENIGMA-relatives study of schizophrenia and bipolar disorder. *Biological psychiatry*, 86(7), 545-556. doi: <https://doi.org/10.1016/j.biopsych.2019.03.985>
- Boedhoe, P.S., Schmaal, L., Abe, Y., Alonso, P., Ameis, S.H., Anticevic, A., [and 96 others including **Buimer, E.E.L.** as part of the ENIGMA OCD working group] & van den Heuvel, O.A. (2018). Cortical abnormalities associated with pediatric and adult obsessive-compulsive disorder: findings from the ENIGMA Obsessive-Compulsive Disorder Working Group. *American Journal of Psychiatry*, 175(5), 453-462. doi: <https://doi.org/10.1176/appi.ajp.2017.17050485>

Forthcoming

- Buimer, E.E.L.**, Pas, P., van den Boomen, C., Raemaekers, M., Brouwer, R.M., & Hulshoff Pol, H.E. Inter-individual differences in facial-emotion processing in pre-adolescence. *In preparation*.

A4 - Societal publications

Voor volwassenen

In juli 2022 heb ik een week lang mijn leven als onderzoeker en mijn onderzoek mogen delen via sociale media via het @NL_wetenschap account van de Universiteiten van Nederland. Een live interview op NPO-radio 1 over deze ervaring en over mijn promotieonderzoek vind je [hier](#).

In dit [interview](#) geef ik meer achtergrondinformatie over de verzameling van hersenscans in het YOUth-onderzoek.

Een infographic met de eerste resultaten van de COVID-19 deelstudie van het YOUth-onderzoek is [hier](#) te vinden.

Een onlineversie van het volledige proefschrift is [hier](#) terug te vinden.

A

Voor kinderen en tieners

Wil je weten hoe het is om hersenonderzoeker te zijn? Je ziet in deze [ylog](#) waarom ik mijn werk zo leuk vind en wat voor onderzoek ik doe. Deze video is opgenomen voor het programma Slimme Gasten van het Wetenschapsknooppunt Universiteit Utrecht.

Informatie over het maken van hersenfoto's vind je in deze [speciale uitgave](#) van New Scientist NL voor jongeren over het Consortium on Individual Development!

Wist je dat ik een samenvatting van mijn proefschrift heb gemaakt speciaal voor jongeren? Kijk snel op de volgende bladzijde of klik [hier](#)!

A5 - Samenvatting voor jongeren

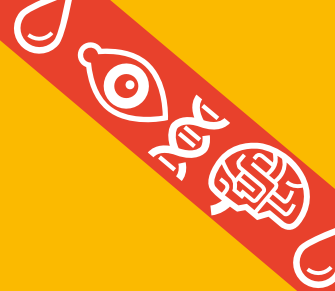
Beelden uit de kindertijd

Jeugdervaringen en hersenen
in de groei

Het proefschrift van hersenwetenschapper Elizabeth Buimer



Samengevat voor jongeren





Hoi, ik ben Elizabeth en ik ben hersenwetenschapper.

Ik heb onderzoek gedaan naar de hersenen van kinderen en tieners. De resultaten van dit onderzoek heb ik opgeschreven in een proefschrift, dat is een werkstuk maar dan met wat meer pagina's (er staan resultaten van vier jaar onderzoek in). Dat proefschrift heb ik nodig om te promoveren. Promoveren betekent dat ik de titel doctor mag gebruiken. Proefschriften worden geschreven in wetenschappelijke taal en zijn daardoor niet zo makkelijk te begrijpen. Daarom heb ik deze samenvatting voor jou geschreven. Ik vind het namelijk belangrijk dat iedereen kan snappen hoe ik mijn onderzoek naar hersenen heb gedaan en wat het heeft opgeleverd.



Het onderzoek dat ik de afgelopen vier jaar deed hoort bij een heel groot onderzoek, namelijk het YOUTH-onderzoek. Mijn onderzoek is een klein stukje van dat hele grote onderzoek. Misschien ken je het YOUTH-onderzoek omdat je mee hebt gedaan. Bij het YOUTH-onderzoek worden duizenden baby's en jongeren een paar jaar gevolgd om te onderzoeken wat er gebeurt als zij opgroeien.

Het YOUth-onderzoek

Jongeren die meedoen aan het YOUth-onderzoek doen tijdens een onderzoeksdag verschillende testjes. Ze krijgen bijvoorbeeld vragen over hun jeugd en er worden hersenfoto's gemaakt met een MRI-scanner. Met die hersenfoto's kan je zien uit welke structuren de hersenen zijn opgebouwd. En met de MRI-scanner kan je ook de activiteit van de hersenen onderzoeken. Meer informatie hierover vind je verderop in deze samenvatting. De gegevens die via YOUth bij jongeren zijn verzameld, heb ik gebruikt voor mijn onderzoek.

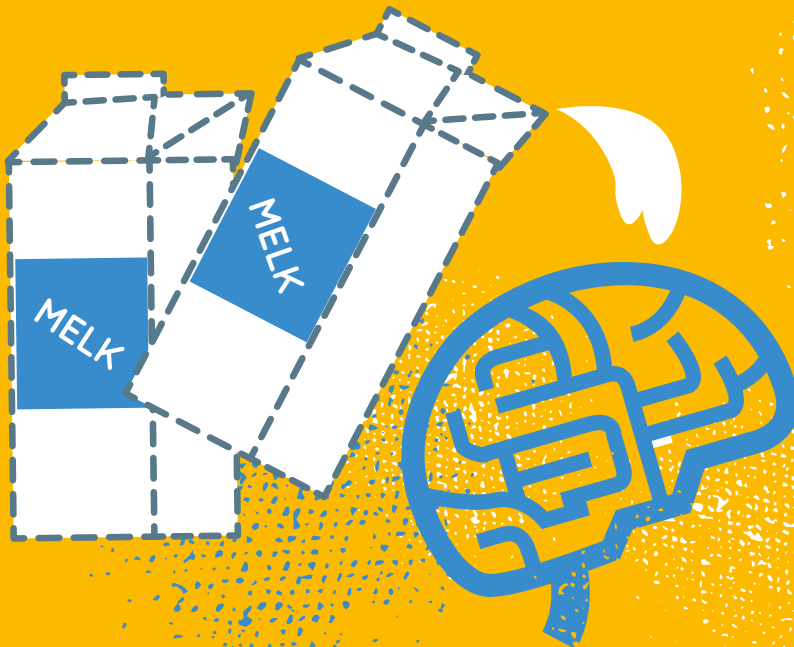
Hersenfoto's zijn interessant, omdat de hersenen van ieder mens uniek zijn. Net zoals jouw vingerafdrukken uniek zijn. Er zijn veel verschillende dingen die allemaal een klein beetje bepalen hoe jouw hersenen eruitzien en hoe jouw hersenen werken. Dat zijn bijvoorbeeld dingen zoals je leeftijd, of je een jongen of meisje bent, het gezin waarin je opgroeit en je erfelijk materiaal (DNA). Zo ontstaan verschillen tussen de hersenen van jongeren.

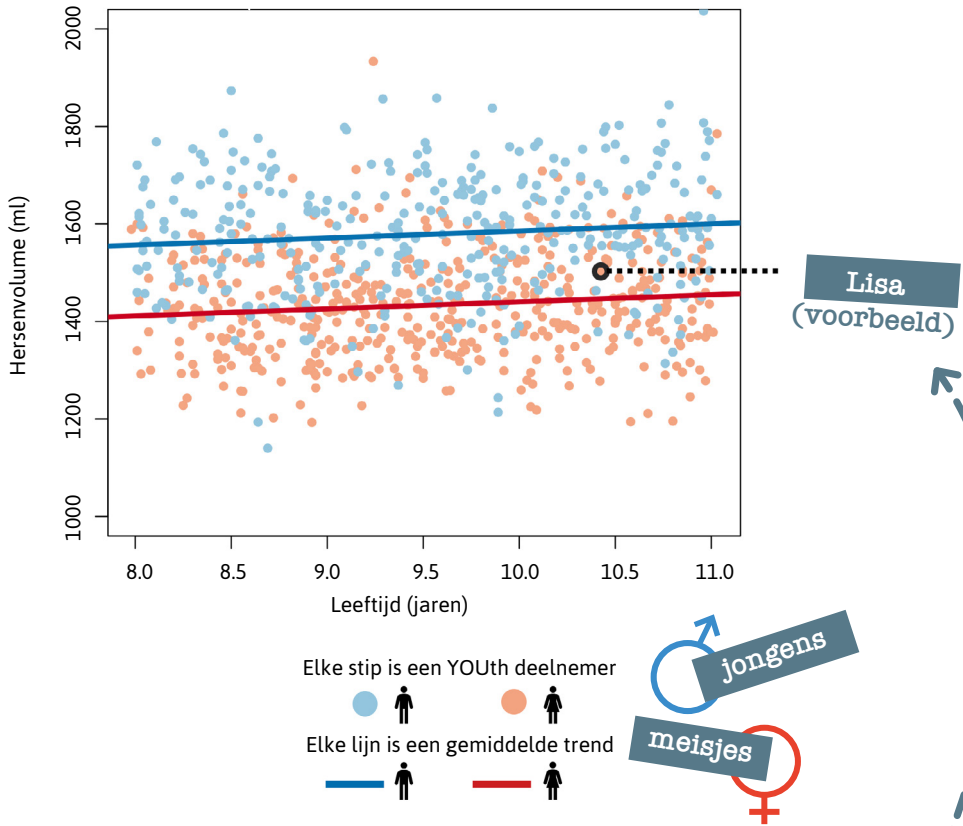




Hoe ouder, hoe groter de hersenen

Als je groeit veranderen je hersenen met je mee. De hersenen van een baby zien er anders uit dan de hersenen van een puber. Al voor je geboorte, dus in de buik, beginnen je hersenen te groeien. En tijdens je kindertijd en puberteit blijven ze doorgroeien. In het plaatje met de roze en blauwe stipjes zie je de hersenvolumes – dus hoe groot de hersenen zijn - van de deelnemers aan het YOUTH-onderzoek. Je ziet dat er veel verschil is tussen de jongeren. Sommige jongeren hebben een hersenvolume tot wel 2.000 milliliter (dat is zo groot als twee melkpakken), terwijl andere jongeren een hersenvolume hebben van zo'n 1.200 milliliter (iets meer dan één pak melk). In dit plaatje zie je ook dat oudere kinderen grotere hersenvolumes hebben.





Laten we eens kijken naar één deelnemer, bijvoorbeeld Lisa (die naam heb ik verzonnen, want ik mag de namen van deelnemers aan het onderzoek niet zomaar delen). In het plaatje zie je dat Lisa grotere hersenen heeft dan een gemiddeld meisje van haar leeftijd. Haar stip ligt namelijk boven de rode lijn. Dat komt omdat hersenvolume niet alleen wordt bepaald door hoe oud je bent of doordat je een jongen of meisje bent. Ook wat je meemaakt in je jeugd en erfelijke factoren bepalen hoe groot je hersenen zijn.



DNA: instructieboekje voor je hersenen

Ik heb je net uitgelegd dat er verschillende dingen zijn die allemaal een beetje bepalen hoe je hersenen eruitzien. Een van die dingen is je DNA. In je DNA liggen alle erfelijke eigenschappen opgeslagen die jouw vader en moeder aan jou hebben doorgegeven. Of je blond of donker haar hebt bijvoorbeeld. In jouw DNA vind je ook een soort instructieboekje waarin staat hoe je hersenen gebouwd moeten worden. Dit unieke instructieboekje wordt gemaakt door informatie te combineren uit het instructieboekje van je vader én je moeder.

Je hersenen zijn een soort fotoalbum

Ook jeugdervaringen en de omgeving waarin je opgroeit hebben invloed op hersenen in de groei. Je kan je vast veel momenten van vroeger goed herinneren zonder dat je daarvoor in je fotoalbum hoeft te kijken. Dat komt omdat ze zijn opgeslagen in je hersenen. Je hersenen zijn dus veranderd door al jouw eigen ervaringen. Die verandering in je hersenen is wel heel klein. Maar als een ervaring heel belangrijk was of heel lang heeft geduurd – bijvoorbeeld als je ouders zijn gescheiden of als er iemand in je omgeving heel ziek is geweest – dan is het effect op je hersenen iets groter. Daardoor zijn die veranderingen makkelijker te meten door wetenschappers. Als je als kind lange tijd veel stress hebt gehad of iets naars hebt meemaakt, kan het zijn dat je hersenen zich daaraan aanpassen. De ontwikkeling van je hersenen verloopt dan misschien wel anders dan bij je leeftijdsgenoten. Deze veranderingen in de hersenen kunnen positieve effecten hebben. Iemand kan bijvoorbeeld door die ervaringen misschien wel beter omgaan met stress, omdat de hersenen eraan gewend zijn. Maar, je raadt het al, de veranderingen in de hersenen kunnen helaas ook negatieve effecten hebben. Zo kan iemand bijvoorbeeld gevoeliger worden voor stress.



Hoe worden hersenfoto's gemaakt?

MRI is een techniek waarmee de binnenkant van je lichaam wordt gefotografeerd. Een MRI-scanner is een tunnel met daarin sterke magneten. De kleinste bouwstenen van het menselijk lichaam zijn ook een beetje magnetisch en reageren op de magneten in de MRI-scanner met een signaal. Dat signaal wordt opgevangen door antennes die verstopt zitten in een helm. Tijdens het scannen ligt iemand in de magneettunnel met de helm op en dan wordt er een foto van de hersenen (hersenscan) gemaakt. Bij het YOUTH-onderzoek maken we verschillende soorten MRI-scans.

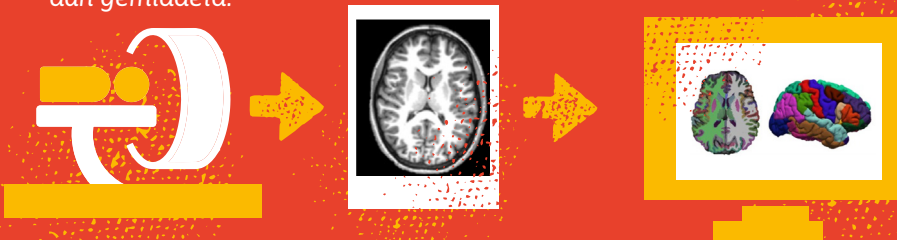


Ik gebruik voor mijn onderzoek drie soorten scans:

Hoe groot zijn de hersengebieden?

1 Anatomische scan

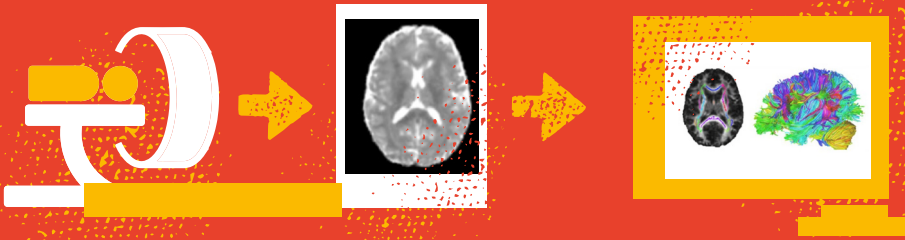
Een anatomische scan is een hele scherpe foto waarop de bouw van de hersenen (de anatomie) gedetailleerd te zien is. Hieronder zie je een voorbeeld van zo'n foto. De grijze stof aan de buitenkant van de hersenen noemen we de hersenschors. De witte stof verbindt de verschillende hersenschors gebieden met elkaar. Rechts op het plaatje zie je hoe de hersenscan eruitziet nadat we een computerprogramma hebben gebruikt dat de hersenen indeelt in kleine gebieden (elke kleur is een gebied). Daarna meten we hoe groot elk gebied is. Bijvoorbeeld om te onderzoeken of bepaalde gebieden bij een jongere groter zijn dan gemiddeld.



Hoe sterk zijn de verbindingen tussen hersengebieden?

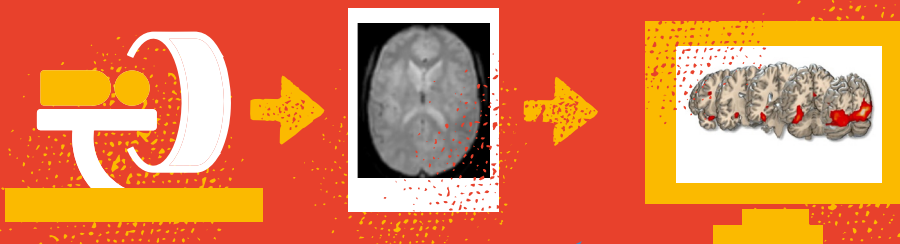
2 Witte stof scan

Een witte stof scan kun je gebruiken om te bepalen hoe sterk de hersengebieden met elkaar verbonden zijn. De witte stof bestaat uit verbindingskabels tussen hersengebieden. Sterke kabels maken het makkelijker voor hersengebieden om met elkaar te kunnen praten en onderling samen te werken. Zoals je ziet op de foto hieronder is bij een witte stof scan, de witte stof juist donkergrijs en de hersenschors wit. Rechts zie je hoe we de scan weer opdelen in verschillende verbindingskabels die elk een eigen kleur krijgen. Daarna meten we hoe sterk de verbindingskabels zijn.



3 Hersenactiviteit scan

Een hersenactiviteit scan is eigenlijk meer een filmpje dan een foto van de hersenen. De jongere die in de MRI-scanner ligt, krijgt verschillende plaatjes te zien. In dit onderzoek zijn het plaatjes van gezichten en van huizen. Daarna meten we of een hersengebied actiever is bij het kijken naar gezichten dan tijdens het kijken naar huizen. Ook kunnen we kijken of de hersenen anders reageren op een bang gezicht dan op een vrolijk gezicht door de hersenactiviteit met elkaar te vergelijken.



Welke gebieden zijn actief tijdens het herkennen van emotie?



Onderzoeksvragen

In mijn proefschrift zoek ik naar een verklaring voor de verschillen in de bouw en activiteit van de hersenen tussen jongeren. In mijn onderzoek probeer ik antwoord te geven op twee vragen:

- 1 Is er een verband tussen de bouw van de hersenen en negatieve jeugdervaringen? Om deze vraag te beantwoorden heb ik gebruik gemaakt van anatomische scans en witte stof scans.
- 2 Is de hersenactiviteit anders in jongeren die snel emoties op gezichten kunnen herkennen (bang of blij)? Om deze vraag te beantwoorden heb ik gebruik gemaakt van hersenactiviteit scans.

De belangrijkste onderzoeksresultaten

Vraag 1: Is er een verband tussen de bouw van de hersenen en negatieve jeugdervaringen?

De hersenen van jongeren die kortgeleden een nare levenservaring hebben meegemaakt (zoals pesten of een scheiding) lijken heel erg op die van jongeren zonder deze ervaringen. De invloed van nare levenservaringen op hersenstructuur lijkt dus mee te vallen. Het kan ook zijn dat de deelnemers aan het YOUth-onderzoek sommige hele nare levenservaringen niet hebben meegemaakt, of dat we die niet hebben kunnen meten. Daarom is het goed om hier nog verder onderzoek naar te doen.

Vraag 2: Is de hersenactiviteit anders bij jongeren die snel emoties in gezichten herkennen?

Jongeren die snel kunnen herkennen of iemand bang of blij is, hebben geen andere activiteit in hersengebieden die zorgen voor het verwerken van emoties in gezichten. Opvallend was dat oudere kinderen het makkelijker vinden om emoties te herkennen dan jongere kinderen. Tot slot, was een leuk resultaat dat de hersenactiviteit groter is tijdens het kijken naar blij gezichten dan naar bange gezichten. De hersenen van jongeren worden dus helemaal wakker van blij gezichten.



Toekomst

Is het je opgevallen dat ik in mijn resultaten wel heel vaak schrijf 'het kan zijn dat...' en 'misschien'. Dat betekent dat mijn onderzoek nog niet klaar is en dat ik – of andere onderzoekers – nog verder zullen gaan met dit onderzoek.

Onderzoek naar hersenontwikkeling kan helpen bij het ondersteunen van jongeren die tegen problemen aanlopen tijdens het opgroeien. Daarom is het fijn dat er zoveel deelnemers mee hebben gedaan aan YOUth. Zo helpen we samen de wetenschap een stukje verder!

Mijn proefschrift is nu af, maar ik ben nog lang niet klaar met hersenonderzoek. Tegenwoordig werk ik bij de Universiteit Leiden. Daar ga ik verder onderzoek doen naar de invloed van de sociale omgeving op hersenontwikkeling. Ook wil ik graag onderzoeken of de hersenen van jongens en meisjes anders reageren op positieve en negatieve gebeurtenissen in de jeugd.

Heb je vragen over deze samenvatting van mijn proefschrift? Dan mag je me altijd een mail sturen: e.e.l.buimer@fsw.leidenuniv.nl

Meer weten?



Scan de QR-code voor een video met nog meer informatie hersenen en MRI.



Scan de QR-code voor meer informatie over YOUth.



Redactie

YOUth-jongerenpanel
Marije Witsenboer
Elizabeth Buimer

Ontwerp

Taluut

Fotografie

Ivar Pel

A6 - Dankwoord

Aan het einde van mijn promotietraject kijk ik met veel plezier en dankbaarheid terug op de afgelopen jaren. In dit dankwoord wil ik graag alle mensen bedanken die hebben bijgedragen aan de totstandkoming van dit proefschrift.

Om te beginnen wil ik alle vrijwilligers bedanken die aan de studies in dit proefschrift hebben meegedaan. In het bijzonder wil ik de duizenden kinderen en ouders bedanken die deel hebben genomen aan het YOUth onderzoek. Dankzij jullie durf en inzet tijdens de testdagen hebben we nu een schat aan data voor huidige en toekomstige onderzoekers.

Hilleke, jouw enthousiasme is absoluut aanstekelijk. Ik heb het brain plasticity lab ervaren als een erg fijn lab om in te werken door de goede sfeer onderling en de wetenschappelijke vrijheid die onderzoekers genoten. In onze meetings probeerde je altijd de inhoud te laten winnen van het bureaucratische geneuzel, maar er was ook altijd ruimte voor persoonlijke gesprekken en humor. Bedankt voor je duidelijke adviezen gedurende mijn hele traject. Ik heb van je geleerd hoe belangrijk *storytelling* is en om te geloven in mijn eigen kunnen.

Rachel, het meest kenmerkend aan jou vind ik jouw integriteit en bescheidenheid. Daarnaast waardeerde ik het hoge tempo van onze meetings en dat jij altijd op detailniveau wist waar ik mee bezig was. Ik heb ongelofelijk veel vaardigheden van jou geleerd. Als begeleider vond ik je geduldig, diplomatiek en eerlijk. In het bijzonder waardeer ik dat de continuïteit van begeleiding niet op het spel kwam te staan bij het starten van jouw nieuwe baan.

Neeltje, dankzij jou kwam ik in Utrecht te werken en zo opende je voor mij de toegangsdeuren naar een academische loopbaan. Toen ik voor je werkte als onderzoeksassistent leerde je me dat ik niet alles meteen hoefde te kunnen, zolang ik er maar plezier in had. Ik vind je een ontzettend warm persoon en ik dank je voor je hulp op verschillende momenten in mijn carrière. **Hugo**, aan het begin van mijn promotietraject vroeg ik of je mijn peet-dagelijks-begeleider wilde zijn. Rachel was zo

consistent in haar begeleiding dat je die rol nooit echt hebt hoeven waarnemen maar we hebben gelukkig veel samen kunnen werken. Ik heb onze persoonlijke en inhoudelijke gesprekken altijd erg gewaardeerd. Ik heb veel van je geleerd qua vaardigheden en manier van nadenken over dingen. Ik hoop dat je verder gaat met cello spelen. Wie weet kunnen we in de toekomst nog samenwerken want elk project kan wel wat *machine learning* gebruiken. **René**, bedankt voor de gezellige lunch en koffie momentjes. Fijn dat Pascal en ik op jouw hulp konden rekenen voor YOUth MRI.

Pascal, bedankt dat je paranimf bent bij de verdediging van dit proefschrift. We waren een goed team binnen YOUth MRI. Gelukkig ging het nooit te lang over werk. Liever praten we bij over de nieuwste aflevering van *Succession*. Ik vond het erg mooi om de belangrijke events in jouw leven van de zijlijn mee te mogen maken (vaak ondersteund met beeldmateriaal).

Sonja, bedankt voor je gezelligheid en energie. Ik heb veel bewondering voor je stressbestendigheid en durf. Knap hoe je je carrière en leven hebt weten in te richten. Leuk dat we samen nog in Rome zijn geweest en hebben geproost op jouw promotie op de trap van het academiegebouw. **Jalmar**, wat was het fijn om jou in het lab te hebben met jouw rust, kennis en gezelligheid. **Janneke**, bedankt voor al je behulpzaamheid en zorgvuldigheid tijdens mijn promotietraject, met name bij de voltooiing ervan. **Jeanette**, bedankt voor hoe je iedereen welkom hebt laten voelen binnen Hilleke's groep. **Zimbo**, bedankt voor al het werk dat je hebt gedaan om onze (legacy) data veilig te stellen. **Yumas**, bedankt dat je zo'n goede admin was op onze superclusters.

Graag bedank ik de leden van de leescommissie voor hun tijd om het proefschrift te lezen en beoordelen: **Prof. dr. Durston**, **Prof. dr. Lucrez Nauta-Jansen**, **Prof. dr. Kenemans**, **Prof. dr. Ramsey** en **Prof. dr. Kemner**.

Big thanks to all the researchers that I had the pleasure of working with during my PhD, such as **Xiao**, **Nikita**, **Daniël**, **Mosi**, **Karis**, **Merel**, **Martijn**, **Natascha**, **Bram**, **Gijs**, **Ties**, **Marissa**, **Anika**, **Carlijn**, **Caroline**, **Roy** etc. Also thanks to the amazing interns: **Vera**, **Eva** en **Julia**.

Bij zo'n groot onderzoek als het YOUth-onderzoek zijn heel veel mensen betrokken. Om te beginnen, alle mensen die hebben gewerkt bij het **KinderKennisCentrum**. In het speciaal wil ik hiervan bedanken **Lilli** en **Coosje** voor Parijs en zoveel meer, **Marije** voor je enthousiaste reacties op mijn communicatie ideeën en je geduldige hulp bij de uitvoering ervan, **Gwen** en **Femke** voor de fantastische manier waarop jullie de KKC praktijken soepel lieten lopen, **Ron** voor je vriendelijkheid en dat je altijd klaar stond voor onderzoekers. En natuurlijk wil ik ook bedanken: **Jacobine**, **Dienke**, **Juliëtte**, **Jolien**, **Liset**, **Leon**, **Elysia**, **Alice**, **Djoya**, **Lesley**, **Mark** en **Danny**. Als laatste wil ik hier **Chantal** bedanken. Wat mij betreft heb je wetenschap echt heel erg vooruit gebracht door binnen het YOUth-onderzoek *team science* en *open science* principes centraal te stellen. Je gaf altijd veel ruimte aan jonge onderzoekers en deelde het podium altijd met het YOUth-team.

De inzet van heel veel medewerkers van het **Team Radiologie van het UMC Utrecht** waren essentieel bij het verzamelen van MRI-data binnen het YOUth-onderzoek. In het bijzonder wil ik hier **Guus** bedanken voor zijn betrokkenheid. Ook wil ik de medewerkers bedanken die hielpen met het veiligstellen van de MRI-data en dan vooral **Gerard**.

Van het **WKZ** wil ik **Sabine**, **Sanne** en **Anouk** hartelijk bedanken.

Het geweldige **SYNC lab** wil ik bedanken voor de fijne sfeer op congressen en tijdens samenkomsten van CID. In het bijzonder wil ik **Eveline** bedanken voor het feit dat ze altijd klaar staat om met jonge onderzoekers mee te denken. **Michelle** en **Suzanne** wil ik bedanken voor het meedenken in mijn zoektocht naar een nieuwe baan.

Christopher, it was a pleasure to find another researcher that was equally excited about privacy, structural MRI and defacing.

I want to thank my new academic family for their support the past year and for making me really excited about the future, **Anne-Laura**, **Maxi**, **Geert-Jan** and **Pauline**. I am also very happy to have met the extended academic family, **Rogier**, **Léa** and **Ivan**. Also, many thanks to the really cool interns that were so patient with

me and boosted my energy levels when needed, **Florencia, Robin, Marjolein, Afroditi** and **Zenna**.

Thanks for the shared dinners and fun activities that distracted me from my PhD thesis, **Stefanie, Rymke, Johanneke, Renske, Aline, Heleen, Anna, Claudia, Jilles, Claudia, Caroline, Femke, Yuri, Signe, Flore, Vinnie, Giel, Yalda, Sven, Alex, Maria** and **Matteo**.

Bedankt familie Buimer (in het bijzonder **opa** en **oma**) en familie Hokke (in het bijzonder **Gert** en **Lida**) voor jullie support en liefde de afgelopen jaren.

Wat een fantastische schoonfamilie heb ik erbij gekregen. **Tijn**, bedankt voor alle gezelligheid. Van jou heb ik geleerd om nooit op te geven. **Tjeerd**, ik ben heel erg blij met de goede band die we hebben en dankbaar dat je onderdeel van mijn leven bent. **Nanda**, wat ben jij een fijne aanwinst als schoonzus en ook nog eens als buurvrouw. **Tini**, leuk dat je mijn onderzoek altijd met zoveel interesse hebt gevolgd en bedankt voor de heerlijkheden die je ons altijd voorschotelt. **Tom**, bedankt voor je gezelligheid en positiviteit.

Lieve **papa** en **mama**, bedankt voor het feit dat jullie me altijd zoveel kansen hebben gegeven en mijn ontwikkeling zo hebben gestimuleerd. Met jullie als vangnet, wist ik dat ik nooit diep kon vallen. **Papa**, bedankt voor jouw onuitputtelijke geloof en vertrouwen in mij. **Mama**, bedankt voor het intense meeleven met elk moment in mijn leven en voor het luisteren naar mijn oneindige stroom verhalen.

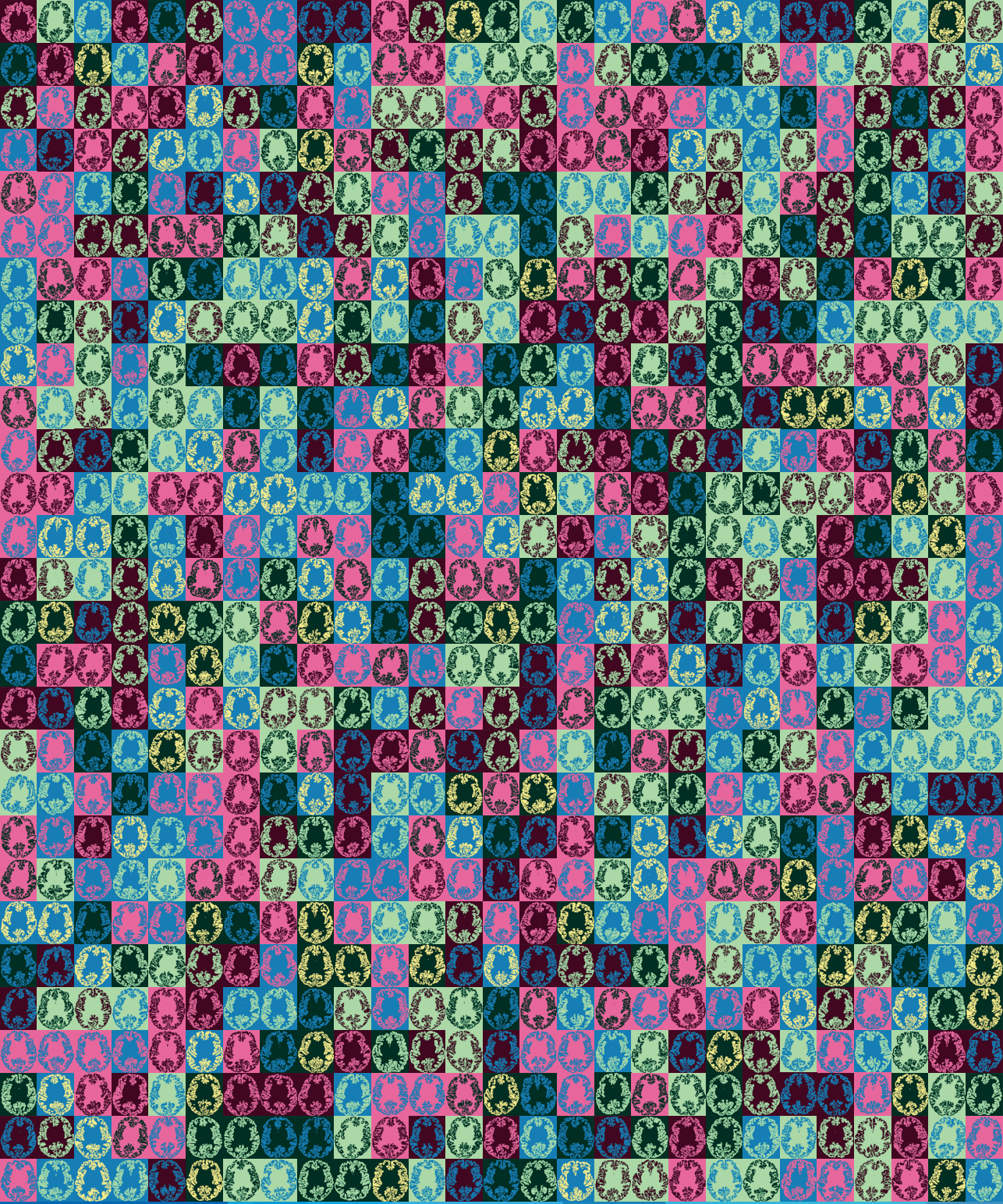
Lieve **Laura**, bedankt dat je er altijd bent op de momenten dat ik je het hardst nodig heb. Ik ben super trots op je. Je hebt me ongelooflijk geholpen als paranimf. Ik had niemand liever naast me gewild tijdens de verdediging.

Frerik. Lieve **Freer**, *God only knows what I'd be without you*. Maar als ik een gok mag doen, dan zou ik denken dat ik ook doctor zou zijn maar een stuk minder gelukkige doctor. Bedankt voor de ontspanning, steun, liefde en leuke grapjes. Het komt niet goed, het is al goed.

A7 - Curriculum Vitae

Elizabeth Buimer was born in Aberdeen, Scotland. She grew up in Leeuwarden, the Netherlands. In 2011 she graduated from high school (Christelijk Gymnasium Beyers Naudé, Leeuwarden). She obtained a Bachelor of Science degree in Psychobiology at the University of Amsterdam in 2014. During her Bachelor, she went to Germany for a research internship in a rehabilitation clinic to investigate the effects of transcranial direct-current stimulation on craving in alcohol addiction. Next, she obtained a Research Master of Science degree in Biomedical Science - Neurobiology at the University of Amsterdam in 2016. During the final year of the Master's program, she contributed to a publication of the ENIGMA-OCD working group, a large world-wide collaboration to unravel the neuroimaging and genetic correlates of obsessive-compulsive disorder, under the supervision of prof. dr. Odile van den Heuvel and dr. Premika Boedhoe at the Amsterdam University Medical Center. In 2016 she got the opportunity to contribute to other ENIGMA working groups and assist other neuroimaging studies as a research assistant under supervision of prof. dr. Neeltje van Haren and dr. Rachel Brouwer at the University Medical Center Utrecht. In the same research group, she started to work on the YOUth cohort study (part of the Consortium on Individual Development) in 2017, first as a research assistant and from 2018 as a PhD-candidate supervised by prof. dr. Hilleke Hulshoff Pol and dr. Rachel Brouwer. Her PhD project aimed to study childhood experiences and the relation to brain structure and function in pre-adolescence. From October 2022 to September 2023, she worked as a postdoctoral researcher at Leiden University in the lab of Prof. Anne-Laura van Harmelen to study neurobiological mechanisms underlying resilient functioning after childhood adversity. Currently, she works as an Assistant Professor at Leiden University, Faculty of Social and Behavioral Sciences, Institute of Education and Child Studies.

A



UMC Utrecht



Universiteit Utrecht

ISBN 978-94-6483-418-5

