



Best Practices in Structural Neuroimaging of Neurodevelopmental Disorders

Lea L. Backhausen¹ · Megan M. Herting² · Christian K. Tamnes^{3,4,5} · Nora C. Vetter¹

Received: 18 August 2020 / Accepted: 2 March 2021 / Published online: 24 April 2021
© The Author(s) 2021

Abstract

Structural magnetic resonance imaging (sMRI) offers immense potential for increasing our understanding of how anatomical brain development relates to clinical symptoms and functioning in neurodevelopmental disorders. Clinical developmental sMRI may help identify neurobiological risk factors or markers that may ultimately assist in diagnosis and treatment. However, researchers and clinicians aiming to conduct sMRI studies of neurodevelopmental disorders face several methodological challenges. This review offers hands-on guidelines for clinical developmental sMRI. First, we present brain morphometry metrics and review evidence on typical developmental trajectories throughout adolescence, together with atypical trajectories in selected neurodevelopmental disorders. Next, we discuss challenges and good scientific practices in study design, image acquisition and analysis, and recent options to implement quality control. Finally, we discuss choices related to statistical analysis and interpretation of results. We call for greater completeness and transparency in the reporting of methods to advance understanding of structural brain alterations in neurodevelopmental disorders.

Keywords Neurodevelopmental disorders · Children · Structural MRI · Study design · Quality control · FreeSurfer

Introduction

Knowledge about the prevalence, clinical diagnosis and treatment of frequent neurodevelopmental disorders has increased considerably over the past two decades. These

comprise, for example, attention deficit hyperactivity disorder (ADHD; prevalence approx. 6 %; Barkley, 2014), conduct disorder (CD; prevalence approx. 3 %; Canino et al., 2010), and oppositional defiant disorder (ODD; prevalence approx. 3 %; Canino et al., 2010). These externalizing disorders often persist into adulthood, impacting quality of life dramatically (Kessler et al., 2005). To prevent worsening of symptoms and life-long suffering from these disorders, it is crucial to investigate and understand their complex neurobiology, which may enable early diagnosis, prevention, prognosis, and treatment during childhood and adolescence.

Certainly, several challenges remain in improving prevention and treatment of these disorders. First, the etiology of neurodevelopmental disorders is still largely unknown, but likely comprises a mix of factors ranging from genetics (Blesson & Cohen, 2020) to epigenetics and environmental influences (Nigg, 2012). Moreover, to date, neurodevelopmental disorders are phenomenologically defined and diagnosed by clinical observations and reports from the child and caregivers. However, classification into clinical categories may not match individual circumstances, problems, and needs (Hyman, 2010). Consequently, psychopharmacological and psychotherapeutic treatments do not help all patients. Therefore, researchers and clinicians should consider neurobiological processes, together with psychological mechanisms, to gain a more holistic picture of neurodevelopmental disorders (see also Research Domain Criteria; RDoC, Sanislow

✉ Lea L. Backhausen
lea_luise.backhausen@tu-dresden.de

✉ Nora C. Vetter
nora.vetter@tu-dresden.de

Megan M. Herting
herting@usc.edu

Christian K. Tamnes
c.k.tamnes@psykologi.uio.no

¹ Department of Child and Adolescent Psychiatry, Faculty of Medicine of the Technische Universitaet Dresden, Dresden, Germany

² Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

³ PROMENTA Research Center, Department of Psychology, University of Oslo, Oslo, Norway

⁴ NORMENT, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

⁵ Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, Norway

et al., 2019). One invaluable tool that has the potential to identify neurobiological markers, and inform diagnosis and therapy is structural magnetic resonance imaging (sMRI). sMRI allows researchers and clinicians to quantify aspects of brain morphometry, such as the size and shape of specific brain structures and regions. By showing how structural brain maturation in neurodevelopmental disorders deviates from typical developmental trajectories, sMRI may advance understanding of disorder etiology and ontogeny. Beyond increasing knowledge about the neurobiological underpinnings of particular neurodevelopmental disorders, a possible research and clinical application could be to directly compare and contrast brain developmental patterns across disorders and subgroups (see Opel et al., 2020), and address questions about transdiagnostic versus disorder-specific developmental abnormalities. Further, it has been suggested that sMRI one day might even provide brain growth charts, similar to existing charts for height, weight and head circumference, which may be used at the individual level to help identify developing disorders early on. Pioneering studies in this field are emerging (Marquand et al., 2019). For example, Dong et al., (2020) analyzed data from two accelerated longitudinal cohorts from China and the United States, including in total 590 typically developing children with 864 scans, and generated brain volumetric growth charts. Importantly, certain growth differences between the two cohorts were observed. Two possible causes of such differences are methodological differences between sites and ethnicity differences, with both of these possibilities requiring increased attention in future endeavors to develop brain growth charts. Moreover, in this context where individual brain developmental differences, including infrequent atypically developmental trajectories, are of interest, existing studies are still limited in terms of sample size or cross-sectional design. For brain growth charts to be useful, it might also, at least in some situations, be necessary to scan the individual child undergoing clinical examination repeatedly over time. This may not only provide a static comparison with the variation in brain structure across children at the same age, but may also allow mapping her or his brain developmental trajectories, similar to how the body growth charts typically are used.

In order to achieve these goals, the design, implementation, and interpretation of sMRI studies require particular attention. First, theoretical foundations of different brain morphometry metrics need to be considered when formulating hypotheses and choosing data analysis tools. Second, the recruitment of children and adolescents with neurodevelopmental disorders is often challenging. Third, these participants tend to move more during data acquisition, which often produces artifacts in the data. Trying to minimize movement by ensuring compliance, training of study personnel regarding movement, controlling data quality and employing quality control software tools are important steps. Further important challenges comprise the choice of statistical analysis approaches, including selecting appropriate thresholds for multiple comparison correction. In this paper

we focus on mid-childhood (5 years) to early adulthood (24 years), as there are almost no longitudinal sMRI data on neurodevelopmental disorders for early childhood due to several methodological challenges. These comprise techniques to manage participant anxiety or excessive movement, technical obstacles like availability of child-appropriate equipment, and special data analysis techniques like the use of pediatric atlases to parcellate brain structure; as discussed in detail by Raschle et al. (2012). Moreover, we focus on brain's gray matter and the semi-automated processing tool FreeSurfer, but refer to other tools where appropriate.

After giving an overview of metrics commonly used to probe brain morphometry, we will review typical neuromaturation from childhood to early adulthood and structural alterations in selected neurodevelopmental disorders, that is, ADHD, CD, and ODD. These frequent externalizing disorders with childhood onset exemplify key challenges in sMRI studies with neurodevelopmental disorders like younger age and hyperactivity in the scanner. For reviews on structural alterations of the less frequent, yet extensively studied autism spectrum disorder, please refer to Bednarz and Kana (2018) and Ecker et al. (2015). Please note that the methodological considerations throughout this paper are not restricted to these disorders but instead apply to the diverse range of neurodevelopmental disorders and studies of typical development.

In this review we provide a hands-on, start-to-finish overview of challenges in clinical neurodevelopmental sMRI research, including suggestions on how to improve practices. We will discuss these challenges following the chronological order of any sMRI study: study design, image acquisition, image processing, quality control of data, and finally, statistical analysis and interpretation.

Brain Morphometry Metrics

To display brain structure, T1-weighted sequences depict high signal (i.e. lighter areas) for fat content as in white matter, and lower signal (i.e. darker areas) for more water content as in cerebrospinal fluid, skull and gray matter (Westbrook & Talbot, 2018). Gray matter consists of neuron bodies, glial cells, dendrites, blood vessels, extracellular space, unmyelinated and myelinated axons. It is found in the outer layer of the cerebrum (cerebral cortex), as well as in subcortical structures, and in the cerebellar cortex. White matter consists chiefly of long-range myelinated axons and is found in the cerebrum and the cerebellum (Mills & Tamnes, 2014). As the majority of sMRI studies of neurodevelopmental disorders focused on gray matter, this will also be the main focus in this paper.

One established method to characterize the volume and density of gray matter in brain structures is voxel-based morphometry (VBM; Ashburner & Friston, 2000). However, these volume output metrics consist of a mixture of complex underlying

effects, which has also lead to criticism (Ashburner, 2009; Davatzikos, 2004). In the following, we will hence focus on surface-based and volume-based analysis in semi-automated tools like FreeSurfer (Fischl, 2012), which allow the individual investigation of several morphometry metrics (see Textbox 1). For a short review of major differences between VBM and surface-based analysis, please see Greve (2011).

Surface-based and Volume-based Analysis

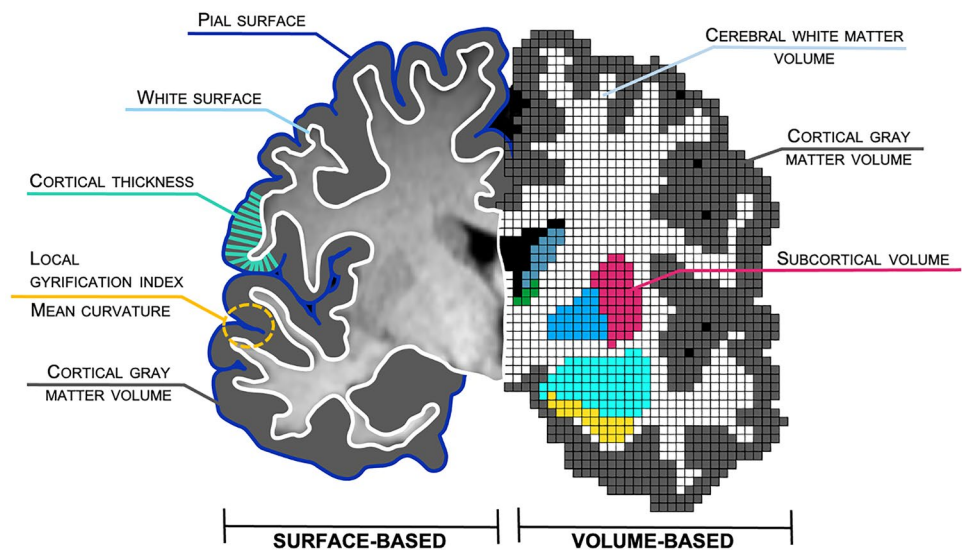
FreeSurfer (Fischl, 2012) is widely used, documented, freely available (<http://surfer.nmr.mgh.harvard.edu/>) and widely supported by the neuroimaging community. FreeSurfer calculates various morphometry metrics in two processing streams: a surface-based and a volume-based stream (see Textbox 1 and Fig. 1 for an overview of these metrics, their computation, and format). Cortical surface-based reconstruction (left side of Fig. 1) is based on geometric models of the cortical surface and the identification of the borders between certain tissue types (described in detail in Dale et al., 1999). The boundary between the pia mater and cortical gray matter forms the pial surface (medium blue in Fig. 1), while the boundary between cortical gray and white matter represents the white surface (light blue in Fig. 1). The cortex is further modeled as a surface with a mesh of triangles (not depicted in Fig. 1). Each meeting point of triangles (called a vertex) has exact coordinates which allows various non-linear manipulations, like inflation, to perform spatial normalization and group analysis, and to improve visualization. These reconstructions allow for a differentiation of cortical volume, thickness, surface area, mean curvature, and local gyrification patterns. For cortical thickness (turquoise in Fig. 1), FreeSurfer calculates the distance between the pial and white surface (Fischl & Dale, 2000). Cortical volume represents a product of cortical thickness and surface area (gray in Fig. 1). Local gyrification index quantifies the gyrification at

each vertex on the surface (yellow in Fig. 1), and is computed in a 3D fashion using a circular region of interest (ROI; 20 to 25 mm) around each vertex (Schaer et al., 2008). Recent studies indicated different genetic, cognitive, and clinical correlates for different cortical morphometry metrics (Nissim et al., 2017; Raznahan et al., 2011; Winkler et al., 2010), emphasizing their independent development.

The volume-based stream (right side of Fig. 1) applies five stages when analyzing T1-weighted raw data (Fischl et al., 2002, 2004) and was developed independently from the surface-based stream. It labels each voxel in cortical and subcortical tissue in a skull-stripped mask of the brain depending on voxel intensity and probability maps. Consequently, the volume of subcortical structures (e.g. nucleus caudatus, thalamus, putamen, globus pallidus, amygdala, and hippocampus), cerebellar gray and white matter, cortical gray matter, and cerebral white matter is calculated (see Fig. 1). FreeSurfer also calculates “estimated total intracranial volume” (eTIV; also called intracranial volume or ICV), using an atlas-representative template and the Atlas Scaling Factor (ASF) which represents “the whole-brain volume expansion (or contraction) required to register each individual to the template” (Buckner et al., 2004, p. 725). For visualization of eTIV, please see Fig. 3 in Buckner et al. (2004, p. 728). This metric may also be used in cross-sectional comparative studies to adjust for individual differences in brain size (see section [Statistical analysis](#)). Alternatively, whole brain volume may be used for adjustment. Typically, it is calculated by summing the gray and white matter volumes, excluding the brainstem. Still, whole brain volume may vary depending on researchers’ choice to include non-brain matter such as cerebrospinal fluid, ventricles, and choroid plexus.

Other semi-automated tools exist that analyze several of the aforementioned metrics, for example, AFNI (<https://afni.nimh.nih.gov/>), Brain Visa (brainvisa.info), Brain Voyager (brainvoyager.com), CARET (brainvis.wustl.edu/wiki/index.php/Main_Page), CAT12 (<http://www.neuro.uni-jena.de/cat/>),

Fig. 1 Overview of structural brain metrics. Coronal slice of an individual brain indicating metrics according to the surface-based (left) and volume-based (right) processing streams as implemented in FreeSurfer. Depicted are the subcortical structures nucleus caudatus (blue-gray), thalamus (green), putamen (magenta), globus pallidus (deep sky blue), amygdala (cyan), and hippocampus (yellow). Black represents cerebrospinal fluid. Corpus callosum and ventricles are not labeled. For illustration purposes, the graphic and scaling was simplified and does not claim anatomical correctness. Figure courtesy of Anna Backhausen



FSL (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>), and MindBoggle (<https://mindboggle.info/>). Please see Mills and Tamnes (2014) and Popescu et al. (2016) for overviews. Moreover, processing pipelines like the Human Connectome Project (HCP) pipeline (Glasser et al., 2013) combine several tools to facilitate multi-modal neuroimaging analysis, including sMRI, resting-state MRI, task functional MRI, and diffusion MRI.

Textbox 1. Brain Morphometry Metrics Given by FreeSurfer (Fischl, 2012).

ATLAS-based Spatial Normalization

Estimated Total Intracranial Volume (eTIV) Also called intracranial volume (ICV); may be used for global brain size adjustment when analyzing cortical volume, subcortical volume, and surface area in cross-sectional group comparisons. Reported in mm^3 or ml.

Surface-based Analysis Surface area Area of the brain surface, spanning two metrics with the same topology:

- 1) **WHITE SURFACE** (inner surface area): area of the boundary between white and gray matter. Reported in mm^2 .
- 2) **PIAL SURFACE** (outer surface area): area of the boundary between gray matter and pia mater. Reported in mm^2 .

Cortical thickness Distance between white surface and pial surface; calculated by finding the closest point on the opposite surface. Reported in mm.

Cortical gray matter volume Also called cortex volume; represents the volume inside the pial surface minus the volume inside the white surface minus tissue inside the ribbon that is not part of the cortex (e.g. hippocampus). Reported in mm^3 or ml.

Mean curvature Indicator for degree of cortex folding with increased curvature indicating increased or sharper folding; Calculated by the average of two principal curvatures of white matter or pial surface; measured in $1/r$, where r is the radius of an inscribed circle. Reported in mm^{-1} .

Local gyrification index Quantifies the amount of cortex buried within the sulcal folds as compared with the amount of cortex on the outer visible cortex; extensive folding indicates large gyrification index whereas limited folding indicates small gyrification index. Has no unit.

Volume-based Analysis Subcortical volumes Gray matter volume of various segmented subcortical structures (e.g. nucleus caudatus, thalamus, putamen, globus pallidus, amygdala, hippocampus, ventral diencephalon, and substantia nigra). Reported in mm^3 or ml.

Cerebellar gray and white matter volume Gray and white matter volume of the cerebellum. Reported in mm^3 or ml.

Cortical gray matter volume Also called cortex volume; sum of all cortical label voxels. Reported in mm^3 or ml.

Cerebral white matter volume Total volume inside the white surface minus anything that is not white matter, does neither include cerebellar white matter nor brain stem. Reported in mm^3 or ml.

Typical and Atypical Structural Brain Development

The objective to track structural brain alterations and understand the underlying mechanisms of neurodevelopmental disorders requires researchers and clinicians to first characterize typical neurodevelopment. Components of cortical and subcortical morphometry develop along different trajectories throughout childhood into adolescence and early adulthood (see e.g. Herting et al., 2018; Tamnes et al., 2017; Vijayakumar et al., 2016). In the next section we will first review typical neurodevelopmental trajectories, followed by a review on alterations in the externalizing disorders ADHD, CD, and ODD. For more details and schematic illustrations of typical neurodevelopmental trajectories for various brain measures, we refer to Mills and Tamnes (2014).

Typical Structural Brain Development

Mills et al. (2016) identified different trajectories for *whole brain volume* (here a composite measure of white and gray matter including the cerebellum) and *eTIV* (atlas-based spatial normalization procedure using a scaling factor) across four longitudinal samples. While whole brain volume increased across childhood, peaked at age 13 followed by a gradual decrease in adolescence and stabilized in the early twenties, eTIV showed an annual increase of around one percent between late childhood and mid-adolescence, followed by a stabilization in late adolescence (Mills et al., 2016).

Cortical gray matter volume, which comprises both cortical thickness and surface area, follows an inverted U-shape trajectory with an increase during early childhood followed by decrease during later childhood and adolescence, and into adulthood (Gilmore et al., 2012; Mills et al., 2016). The decrease of gray matter volume from childhood into adulthood seems to be largely driven

by decreases in cortical thickness (Storsve et al., 2014; Tamnes et al., 2017). According to pioneering studies, regions serving primary functions such as the visual and motor cortex show a gray matter peak earlier, while the peak in higher-order association areas such as the prefrontal and temporal cortex appears later (Gogtay et al., 2004; Sowell et al., 2004). However, more recent longitudinal studies did not replicate a peak of gray matter during late childhood or adolescence. Instead, results suggest that gray matter volume is highest before 8 years of age (Mills et al., 2016), decreases across the second decade (Tamnes et al., 2013; Wierenga et al., 2014), and stabilizes in the third decade (Mills et al., 2016). For *cortical thickness*, recent studies reported a monotonic decrease from childhood to early adulthood (Tamnes et al., 2017; Walhovd et al., 2017). Generally, *cortical surface area* showed smaller changes relative to cortical thickness and volume (Tamnes et al., 2017). Increase was reported until about 9 years of age (Wierenga et al., 2014) followed by stability or relatively smaller decrease during adolescence (Amlien et al., 2016; Tamnes et al., 2017; Vijayakumar et al., 2016; Wierenga et al., 2014). Although being one of the most prominent features of the human brain, a unified model on the development and role of *cortical folding* does not yet exist. Originally, cortical shape was seen as a product of underlying patterns of connectivity (see axon tension-based model of convolutional development; Essen, 1997). However, more recent investigations on the development of cortical folding focus on genetic influences (Alexander-Bloch et al., 2020), as well as mechanical (stiffness and elasticity), and cellular mechanisms (Llinares-Benadero & Borrell, 2019). So far, a smaller number of studies investigated the development of curvature and gyrification patterns. These mostly found decreasing gyrification from childhood to adulthood (Mutlu et al., 2013; Raznahan et al., 2011). Concerning *subcortical brain structures*, the basal ganglia (i.e. nucleus caudatus, putamen, and globus pallidus), nucleus accumbens, and cerebellar gray matter showed volume decreases from ages 8 to 22 (Tamnes et al., 2013), while the amygdala and hippocampus appeared to increase with age during adolescence (Durston et al., 2001; Giedd et al., 1996) or showed little to no change (Tamnes et al., 2013). Consistent with these data, a recent study reported slight nonlinear increase for hippocampus and amygdala volume from age 10 to 22 (Herting et al., 2018). In contrast to cortical, subcortical and cerebellar gray matter volumes, which for the most part decrease in adolescence, *white matter volume* has been shown to increase throughout childhood and adolescence, and possibly even further (Mills et al., 2016; Westlye et al., 2010).

Although longitudinal studies are the gold standard and were widely used in recent years to characterize development, please note that contradictory findings exist. These likely stem from specific methodological challenges of longitudinal projects like age range, number of assessments, sample characteristics, image processing techniques, and longitudinal statistical analytic methods. These challenges and best practices have been recently reviewed by Vijayakumar et al. (2018).

Alterations in the Externalizing Disorders ADHD, CD, and ODD

In contrast to typical development, almost all studies on the neurodevelopmental disorders ADHD, CD, and ODD are cross-sectional, that is, focus on age-independent or age-related differences between patient and control groups, as opposed to differences in development (but see longitudinal studies from Shaw et al., 2012, 2013, 2014). Overall, distinct alterations in neurodevelopmental disorders emerge for cortical and subcortical metrics (see e.g. meta-analysis on ADHD: Hoogman et al., 2017; Nakao et al., 2011 and meta-analysis on CD or ODD: Noordermeer et al., 2016). Please note that results for these neurodevelopmental disorders are biased given high comorbidity rates between ADHD and CD or ODD in most studies, complicating interpretation (Vetter et al., 2020).

For ADHD, reductions have recently been found in eTIV (5%; Vetter et al., 2020), as well as in total brain volume (2.5%; Greven et al., 2015). Concerning *cortical gray matter volume*, a meta-analysis with children and adolescents with CD or ODD revealed reductions in the bilateral insula and the left middle/superior frontal gyrus, possibly indicating compromised hot executive functions (e.g. emotion processing, empathy, and introspection) in CD or ODD (Noordermeer et al., 2016). In ADHD, studies showed decreased (Ambrosino et al., 2017), but also increased (Semrud-Clikeman et al., 2014) volume of the prefrontal cortex. Moreover, delayed *cortical thinning* in prefrontal areas in ADHD has been interpreted as a maturational delay in regions important for cognitive control processes, including attention and motor planning (Shaw et al., 2007). Children and adolescents with CD showed reduced cortical thickness within the parietal lobe, paracentral lobule, precuneus (Hyatt et al., 2012), and superior temporal cortex (Wallace et al., 2014). The thickness of the latter region also correlated with callous-unemotional traits, the core emotional component of psychopathy (Wallace et al., 2014). Concerning *surface area*, a reduced total, frontal, temporal, and parietal area has been found in children and adolescents with ADHD (Noordermeer et al., 2017), as well as a delayed developmental trajectory of the prefrontal cortex surface area (Shaw et al., 2012). For CD or ODD, a reduced prefrontal surface area (Fairchild et al., 2015; Sarkar et al., 2015)

or no surface area alterations have been reported (Wallace et al., 2014). Furthermore, Shaw et al. (2012) reported no alterations in *cortical gyrification* for children and adolescents with ADHD from ages 10 to 18, neither in baseline gyrification nor in developmental trajectories. Similarly, no alterations in developmental trajectories for intrinsic curvature and local gyrification index were found for adolescents with ADHD (Forde et al., 2017). For 22 children and adolescents with CD, decreased gyrification has been shown in the right ventromedial prefrontal cortex (Wallace et al., 2014), yet this finding did not survive cluster correction. Moreover, Hyatt et al. (2012) found widespread folding deficits in a similar sample, located mainly in anterior brain regions, including the left anterior insular cortex. In line with the somatic marker hypothesis, alterations in this region may be related to deficits in the formation of subjective feeling and empathy (Medford & Critchley, 2010). Concerning *white matter volume*, no alterations were found neither for ADHD, nor for CD (Greven et al., 2015; Stevens & Haney-Caron, 2012).

For *subcortical brain structures*, a recent mega-analysis spanning 1713 participants with ADHD aged 4 to 63 years found volume reductions in the amygdala, nucleus accumbens, hippocampus, and putamen (Hoogman et al., 2017). Here, the biggest effects emerged in the amygdala, which emphasizes the role of emotional dysregulation in ADHD (Hoogman et al., 2017). In children with CD or ODD, volume reductions in the striatum, amygdala, and hippocampus (Noordermeer et al., 2017; Rogers & Brito, 2016) seem to mirror deficits in emotion processing and decision-making. In sum, there is evidence for alterations in cortical and subcortical structures in children and adolescents with neurodevelopmental disorders such as ADHD and CD or ODD. Affected regions seem to reflect distinct mechanisms underlying the symptomatology of each disorder, that is, attention processes, motor planning, and emotional regulation in ADHD, as well as empathy, introspection, and emotion processing in CD and ODD.

Still, longitudinal studies with large samples are mostly missing, presumably in part due to challenges involved in conducting longitudinal studies with patient groups. However, longitudinal studies are critically needed to probe brain developmental trajectories in neurodevelopmental disorders, both at the group and individual level, as these may yield different results than case-control studies testing for age-effects (see e.g. discussion about limitations with cross-sectional designs in Kraemer et al., 2000). Longitudinal studies are needed to inform us about the ontogeny of structural brain alterations in neurodevelopmental disorders, and ultimately to track developmental trajectories in individuals with psychopathological risk factors or a clinical diagnosis.

Finally, several findings remain inconclusive, especially when trying to unravel the unique role of the different brain structure metrics in atypical neuromaturation. Divergent findings might arise, among other reasons, from

methodological heterogeneity in image acquisition, image data processing, and statistical analysis. Hence, we will now present guidelines with the goal to help clinical researchers improve their practices and ultimately to gain more robust, cross-study comparable results for the field. In Textbox 2 we include a non-exhaustive summary of important considerations in study design and guidelines for image acquisition and analysis, quality control, and statistical analysis, which apply to all studies in clinical developmental structural neuroimaging research.

Textbox 2. Guidelines for Reporting Methodological Details in Clinical Developmental sMRI Research

Study Design ✓ Consider generalizability of sample during recruitment and report details (i.e. age, IQ, socioeconomic status, pubertal status, and ethnicity)

✓ Report physical and psychiatric comorbidities

✓ For the patient group:

- Report characteristics for (previous) medication and therapy, age of onset and duration of illness for neurodevelopmental disorder
- Report diagnostic procedure

- confirmed by whom and how (e.g. through questionnaires and clinical interviews by registered psychologist or study staff)
- cut-offs to define clinical psychopathology
- specify subtypes and severity

✓ Match groups according to e.g. sex and age; provide information on matching strategy

✓ Provide information on missing data (e.g. questionnaires, medication status, IQ)

Image Acquisition

✓ Implement and report protocols to improve comfort and thus reduce motion

✓ Report on participants' motion

✓ Consider acquisition techniques (e.g. fMRI as proxy, PROMO) for motion-correction

✓ Avoid change in scanner hardware, sequences and protocols across sites and participants

- If not possible, account for differences in all analyses

Image Processing and Quality Control

✓ Employ same software (and version) across all participants within a study and report details

✓ Give preference to software that covers brain metrics and regions of interest, chose these a priori based on literature and hypotheses, and report details

✓ Report on quality control procedure

- Inspection of the quality of raw and processed images with specification of exclusion criteria
- Tools or algorithms used during quality control procedure
- Manual changes/trouble shooting techniques

Statistical Analysis

- ✓ Consider e.g. sex, age, and global brain size as covariates
 - If correcting for global brain size:
 - Report the brain metric and correction method used
 - Report results from both raw and corrected regional brain measures
 - Take into account relationships between sex, age, and global brain size
- ✓ Appropriately account for the multivariate nature of the data by:
 - Correcting for multiple comparisons applying suggested thresholds for different brain metrics
 - Conducting multivariate analyses

Conclusions

- ✓ Interpret findings within the bounds of the analytic technique and in line with the statistical analyses performed (e.g. as relative differences if results were adjusted for global brain size)

Steps to Successful Clinical Developmental Structural Neuroimaging Research

Study Design

Decisions on the study design depend on the goal of the clinical developmental sMRI study (Greene et al., 2016). If one seeks to understand underlying mechanisms of a disorder, a “clean” sample, that is, participants with symptoms of the neurodevelopmental disorder of interest who do not receive medication or treatment, is preferable (Greene et al., 2016). However, this is seldom feasible as most patients with neurodevelopmental disorders receive some kind of treatment. In addition, samples that are more naturalistic allow for investigation of treatment effects (e.g. medication). Further, studies investigating neurodevelopmental disorders are often not capable of including large enough patient groups to compare effects of important, possibly confounding variables such as disorder subtypes, comorbidities, medication, and other treatments. Decisions made at this stage regarding recruitment and study focus will substantially affect the capability to conduct statistical analyses and reliably interpret results.

Many challenges of study design in clinical neurodevelopmental sMRI studies have already been discussed in detail elsewhere, spanning participant inclusion, sample composition (sample size, high versus low-functioning participants, subtypes, and comorbidities), medication, and other treatment history, as well as considerations for control groups (Bednarz & Kana, 2018; Greene et al., 2016). Therefore, we kindly refer to these publications. For discussion on statistical power and selective, small or non-representative samples, please refer to Klapwijk et al. (2019a, 2019b). We will further present considerations for image acquisition, quality control and image processing.

Image Acquisition and Immediate Quality Control

Most studies including MRI examinations (i.e. functional task-based/resting-state MRI or diffusion MRI) also run a T1-weighted sequence by default to check for neuroanatomical anomalies and to coregister data. The FreeSurfer Wiki (<https://surfer.nmr.mgh.harvard.edu/fswiki>) recommends T1-weighted sequence acquisition protocols for use with the FreeSurfer processing pipeline. Contrary to functional MRI, which aims to monitor brain activity, the measurement of brain morphometry is not timing sensitive during one MRI session. In addition, the acquisition time of typical sequences (≈ 1 mm isotropic) is relatively short (5–10 minutes) and users may inspect quality of T1-weighted images immediately after scan completion. Hence, if data quality was not satisfactory, such a short T1-weighted sequence can often be repeated (see also Backhausen et al., 2016) and a time buffer of 15 minutes included in the MRI session schedule may help reduce data loss in challenging samples like children and adolescents with neurodevelopmental disorders. Therefore, personnel present at the scanner should learn about different types of artifacts (mostly technical and motion artifacts) and their impact on the data to be able to identify compromised data quality and decide on the necessity to rescan. Technical artifacts include head coverage, radiofrequency noise, signal inhomogeneity, and susceptibility (Costa et al., 2009; Reuter et al., 2015; Wood & Henkelman, 1985). Motion artifacts result from the participant swallowing, blinking, chewing, turning, fidgeting or repositioning a limb (Bellon et al., 1986; Zaitsev et al., 2015). Younger age groups tend to produce more motion artifacts (Blumenthal et al., 2002; Satterthwaite et al., 2013; Van Dijk et al., 2012; Yuan et al., 2009). Moreover, images of children and adolescents with neurodevelopmental disorders might be particularly prone to motion artifacts due to the symptoms themselves (see for ADHD: Backhausen et al., 2016; Rauch, 2005; Vetter et al., 2020, and for CD: Huebner et al., 2008). Our group recently demonstrated that good quality

of structural data can be acquired in clinical developmental samples such as ADHD and CD or ODD when applying a systematic and detailed workflow (Backhausen et al., 2016). A further important procedure to reduce movement is to ensure compliance by age-appropriate instructions with film clips, a playful approach like the “statue game”, or by using a mock scanner (Raschle et al., 2009), allowing the participants to understand and practice the importance of holding still. These preparations can even be individualized in the form of personalized familiarization strategies and rewards to enable children and adolescents with neurodevelopmental disorders to better tolerate the MRI scanner environment and ensure high quality images (Pua et al., 2020).

Moreover, prospective motion correction (PMC) techniques exist which track the participants’s head motion during the scan (e.g. through volumetric navigators) and modify pulse sequences to correct for participant motion (Tisdall et al., 2016). PMC techniques specific to the three main scanner manufacturers include PROMO for GE (White et al., 2010) and motion tracking systems for Siemens (Zaitsev et al., 2006) and Philips (Ooi et al., 2009). To evaluate the costs and benefits of PMC, a recent study compared reliability and quality of structural imaging in traditional Magnetization-Prepared Rapid Gradient-Echo (MPRAGE) versus MPRAGE with PMC in a developmental data set (Ai et al., 2020). They reported higher intra-sequence reliability but poorer quality metrics in scans with PMC compared to the traditional MPRAGE and noted that scans with PMC were robust but not fully resistant to high head motion (Ai et al., 2020). In conclusion, Ai et al. (2020) recommend the use for hyperkinetic populations when increased motion can be expected (i.e. children, patients with neurodevelopmental disorders).

Still, from our experience, data quality can and should be evaluated at different stages in an sMRI study. First, check acquired images at the scanner console after running the structural scan to allow for re-scan if needed and possible. Second, visually rate data quality of acquired raw T1-weighted data sets according to a standard rating system or using an automated method, and third, check data sets after they have been processed, again, either visually or using an automated method (see below and Backhausen et al. (2016) for a detailed workflow suggestion).

Another important concern of image acquisition is the reliability, that is, overall consistency of MRI-derived output metrics. Reliability might vary across MRI scanners, field-strength and head-coils (Heinen et al., 2016). As higher field strength improves signal-to-noise ratio and spatial resolution (Tijssen et al., 2009), 3 Tesla MRI scanners are preferable as compared to 1.5 Tesla MRI scanners. More recently, ultrahigh-field 7 Tesla MRI has been utilized given it provides greater signal-to-noise ratio, contrast-to-noise ratio, and increased spatial resolution as compared to lower magnetic fields (Barisano et al., 2018). Although 7 Tesla MRI scanners became FDA approved for clinical use in October 2017 (Barisano et al., 2018), and

a few studies, including the Human Connectome Lifespan (<https://www.humanconnectome.org/study-hcp-lifespan-pilot/phase1b-pilot-parameters>), have successfully imaged children 8 years and older on 7 Tesla MRIs (albeit not structural MRI but rather to collect resting-state and diffusion MRI), 3 Tesla MRI has been more widely used world-wide. Moreover, 7 Tesla human brain imaging still has some limitations and faces technical challenges, including increased specific absorption rate (SAR) and increased sensitivity of inhomogeneity and motion artifacts (Barisano et al., 2018). Participant protections are also unique given potential increases in nausea, claustrophobia or dizziness. As advances in magnetic resonance technology tackle these issues, mainstream imaging with 3 Tesla may eventually be replaced by 7 Tesla protocols. Regardless of field strength, MRI technologists or physicists should regularly perform data quality checks (e.g. signal-to-noise) to ensure stable machine characteristics and performance. It is imperative that all participants are scanned using the same hardware, software, and MRI sequence to avoid increased variability or even systematic bias. If unavoidable, these effects should be estimated and statistically accounted for (Lee et al., 2019), either by including site, scanner or sequence as a covariate in statistical analysis or by applying tools developed for multi-site harmonization. The recent algorithm ComBat removes unwanted non-biological variability associated with different properties of MRI scanners from cortical thickness data in cross-sectional (Fortin et al., 2018) and longitudinal studies (Beer et al., 2020), which may increase power and reproducibility of subsequent statistical analysis.

Image Processing

The choice of an sMRI data processing tool depends on the research question and regions or metrics of interest. Since tools differ with regard to reconstruction algorithms and brain atlases, it is important to learn about the specific nomenclature of brain regions available for ROI-based analysis. Importantly, all tools, although appearing to offer automatic processing, require thorough quality control and other “manual” decisions. We will thus term them “semi-automatic”.

As previously outlined in detail, FreeSurfer (Fischl, 2012) computes volume-based and surface-based metrics. Using nonlinear transformations, cortical measurements are spatially re-sampled onto a standard surface-based template (fsaverage), which represents an average brain (Sabuncu et al., 2014). By default, FreeSurfer provides outputs of these estimates using the option of two standard atlases: the Destrieux atlas (Destrieux et al., 2010) and the Desikan-Killiany atlas (Desikan et al., 2006; see Fig. 2). In adults, the test-retest reliability tends to be higher when using the Desikan-Killiany compared to the Destrieux atlas (Iscan et al., 2015); albeit to our knowledge a similar

comparison has not yet been performed in children and adolescents.

For more than one data set per participant, researchers and clinicians should use the FreeSurfer longitudinal stream (Reuter et al., 2010). This pipeline includes surface- and volume-based image processing in three steps: 1) CROSS: processing of every data set cross-sectionally, 2) BASE: template creation, and 3) LONG: reanalyzing cross-sectional data sets with information from the BASE template. This process improves robustness compared to conventional cross-sectional analysis pipelines (Jovicich et al., 2013).

To sum up, researchers and clinicians need to keep in mind that output metrics and nomenclatures vary between image processing tools and their brain atlases, complicating interpretation of results. When comparing results of structural alterations in clinical developmental samples, it is thus crucial to consider the size and location of brain regions used for ROI-based analysis and how they compare to similar regions from other brain atlases implemented in different semi-automated processing tools. Moreover, for all neuroimaging studies, but perhaps especially for developmental clinical sMRI, the next important step is to check data to ensure good data quality for valid results. We will thus present considerations on quality control of both raw and processed data in the next passage.

Quality Control After Image Acquisition

Bad image quality can lead to poor image reconstruction with semi-automated tools and, importantly, to underestimation of gray matter volumes and cortical thickness (Backhausen et al., 2016; Blumenthal et al., 2002; Reuter et al., 2015). One

study also demonstrated the impact of quality control procedures on developmental trajectories, as quadratic trajectories of cortical thickness across ages 5–22 years could no longer be identified when applying a more stringent quality control as opposed to the standard form (Ducharme et al., 2016). Hence, stringent quality control procedures are of critical importance in clinical developmental sMRI studies to prevent biased results on case-control differences or on developmental trajectories.

Conventional quality control procedures include visual inspection of raw or processed T1-weighted images, without knowing patient versus control participants to reduce inherent bias. First, evaluating them according to a rating system including several quality markers (e.g. blurring, ringing, signal-to-noise ratio of the gray and white matter borders and subcortical structures, see rating system proposed by Backhausen et al., 2016) and then sorting scans into categories (e.g. as “pass”, “warn” or “fail” by Backhausen et al., 2016; Reuter et al., 2015). Those categorized as “fail” should then be excluded from further data analysis. Implementing visual quality control based on rating systems like these can be challenging as it usually requires experienced raters and is excessively time consuming, especially for large data sets. Yet, the need of increased transparency and detailed reports of quality control protocols has to be emphasized, as to this point several previous studies did not include any description of quality control (Vijayakumar et al., 2018). Arguably, both inter-rater as well as intra-rater variability of visual quality control is generally high but may still lead to inclusion of poor quality scans and exclusion of scans of usable quality (Klapwijk et al., 2019a, 2019b). In an attempt to expand the conventional visual quality control, we present a non-exhaustive selection of recently developed techniques in

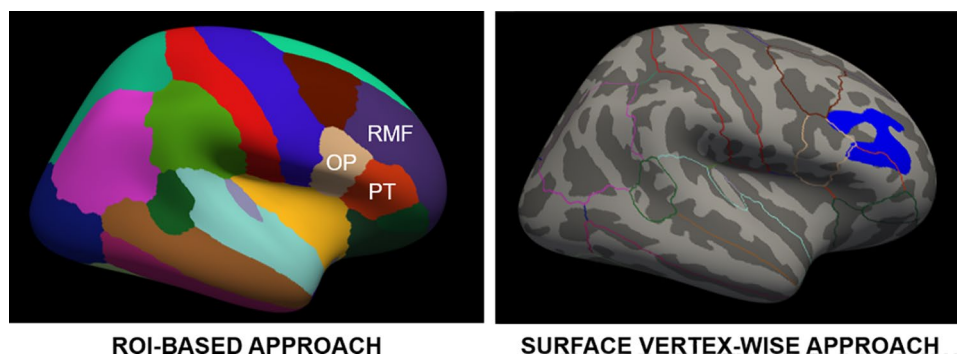


Fig. 2 ROI-based versus surface vertex-wise approach as implemented in FreeSurfer. ROI-based approach depicting the lateral part of the right hemisphere with Desikan-Kiliany atlas regions (left) and surface vertex-wise approach (right) for statistical analysis of clinical developmental sMRI data. The hemisphere is inflated for a better view of gyri and sulci. The blue area on the right picture highlights a region with significant differences in cortical thickness between two groups, which

falls partly into pars triangularis and rostral middle frontal cortex as indicated by Desikan-Kiliany atlas region outlines. Using the ROI-based approach this difference may or may not contribute to significant cortical thickness differences in the rostral middle frontal cortex, pars triangularis, or pars opercularis without the possibility of localizing the area more precisely. OP = pars opercularis; RMF = rostral middle frontal cortex; PT = pars triangularis

Table 1. They are evaluated concerning quality control input measures, visual quality control and classifier categories, techniques, quality control outputs, and performance.

The first approach shown in Table 1 uses the Euler number from the standard FreeSurfer output as a proxy for cortical reconstruction quality. It represents the number of topological defects which may occur during the surface retessellation step when structural data sets are reconstructed by the FreeSurfer pipeline. An Euler number of 2 (equal to that of a sphere) indicates a smooth cortical reconstruction. However, as stated by Fischl, Liu, & Dale (2001), two types of defects may occur during the reconstruction process, termed “holes” and “handles”. Holes consist of small perforations in planar sheets of white matter and handles are bridges between nonadjacent points in the cortex. Each defect of either type reduces the Euler number and FreeSurfer seeks to maximize the Euler characteristic to a value of 2 in subsequent topology correction steps. This quality measure was recently found to be consistently correlated with visual quality control ratings across samples, being able to identify images scored “unusable” by human raters with a high degree of accuracy (area under the curve (AUC): 0.98–0.99; Rosen et al., 2018). Rosen et al. (2018) reported z-scores for Euler numbers, and advised researchers to determine classification thresholds for each specific data set. This approach was followed by Yu et al. (2018) who identified three cases as outliers according to the averaged Euler number across the left and right hemispheres and thus excluded them from further analyses. Here, Yu et al. (2018) set the threshold for outliers to averaged Euler numbers < 3.29 standard deviations below the sample mean.

Furthermore, the frame-by-frame displacement (FD) approach uses head movement parameters of functional MRI scanning (average FD calculated for each individual; Power et al., 2014) during the same scan session as a proxy for movement during the structural scan (Savalia et al., 2017). As FD is derived from functional scan output it may be seen as an objective quantitative quality measure in comparison to qualitative visual quality control rating, which is arguably subjective. It is attempted to hereby identify and reduce potential motion-induced bias. Individuals showing excessive head motion during the functional scan combined with bad visual quality control ratings on the structural scans themselves can be “flagged” and subsequently removed from further statistical analysis. Nevertheless, measures derived from functional scans remain imprecise estimations as the severity of participants’ movement may vary across the scanning session.

Using automated machine-learning algorithms, the MRI Quality Control tool (MRIQC; Esteban et al., 2017) extracts 64 image quality metrics (IQM) from raw T1-weighted images, including measurements of variability and specific artifacts in the images, and gives a recommendation whether

to include or exclude each image. The tool also provides individual anatomical reports including calculated IQM and metadata, as well as a series of image mosaics and plots designed to assist visual assessment of structural images. MRIQC can be accessed freely online via a web interface using the OpenNeuro.org portal. Moreover, the source code is publicly accessible through GitHub (<https://github.com/poldracklab/mriqc>) providing a maximum of transparency.

Moreover, the supervised-learning model Qoala-T tool (Klapwijk et al., 2019a, 2019b) predicts manual quality control ratings from FreeSurfer processed metrics including the Euler number (namely “Surface holes” in the FreeSurfer output files), as well as subcortical volumes, cortical thickness, and cortical surface estimates. Researchers can choose either to predict scan quality by using the BrainTime data model provided by Qoala-T or to train the algorithm using ratings of their own data set. Qoala-T scores ranging from 0 to 100 are provided for every individual scan (scores smaller than 50 are advised to be excluded from further analyses). Still, as the algorithm had been originally trained based on subjective visual quality control ratings, the authors recommend to visually check scans with a score between 30 and 70 as misclassified scans are more likely to fall within these boundaries (Klapwijk et al., 2019a, 2019b). The use of the Qoala-T tool could greatly reduce the time needed for quality control as only a part of the data set has to be visually checked. More importantly, this procedure could further help to reduce variability related to visual quality control, thereby benefiting the comparability of data quality between studies (Klapwijk et al., 2019a, 2019b). Qoala-T scores could even be used as a covariate for head motion, that is, motion artifacts in statistical analysis. In sum, these tools provide researchers with a good overview of the quality of their data set and assist decisions about including or excluding data sets.

After good data quality has been insured, more challenges await when statistically testing for brain structure differences in neurodevelopmental disorders. We will address challenges concerning ROI-based versus vertex-wise analysis, covariates, multiple comparison correction, and generalization of results in the next section.

Statistical Analysis

Statistical analysis affects interpretation of results and may potentially bias decisions concerning diagnosis and treatment of a disorder. For a detailed review on statistical inferences and pitfalls in neuroimaging, see Hupé (2015).

FreeSurfer offers two possibilities to statistically analyze group comparisons and correlations with other variables of interest, namely the *ROI-based and surface vertex-wise approach*. ROI-based includes analyzing the subcortical and cortical volumes, cortical thickness, surface area, and local gyrification

Table 1 Quality control approaches for sMRI data

Method	QC input metrics	Visual QC/ classifier categories	Technique	QC output	Performance
FD (Savalia et al., 2017)	FD from functional MRI scan of the same session as proxy for motion in T1-weighted images	Three categories: pass, warn, fail	Flagging procedure; combining visual QC and estimates of head motion from functional MRI scans	FD estimates and visual QC ratings	FD estimates complement visual QC rating
Euler number (Rosen et al., 2018)	Euler number outputted by FreeSurfer	Three categories: 0 (gross artifacts/fail), 1 (some artifacts but usable), 2 (no artifacts)	/	Euler number, no specific recommendations	Euler number as most accurate quality measure/highest correlation with visual QC
MRI-QC (Esteban et al., 2017)	Raw T1-weighted images, 64 IQMs per input image	Binary classifier: include, exclude	random forests classifier trained on a publicly available, multi-site data set (17 sites, $N = 1102$)	individual anatomical reports (calculated IQMs and meta-data in the summary, as well as a series of image mosaics and plots designed for the visual assessment of images)	Intra-site prediction: high accuracy; Unseen site prediction: leaves space for improvement (76 % \pm 13 % accuracy)
Qoala-T (Klapwijk et al., 2019a, 2019b)	Metrics form the FreeSurfer output files aseg.txt, aparc_area.txt and aparc_thickness.txt (all for both hemispheres) including the variable surface holes	Four categories: 1 (excellent), 2 (good), 3 (poor), 4 (failed)	supervised-learning model, random forests classifier trained on the BrainTime data	Qoala-T score (ranging from 0 to 100), recommendation whether to visually check and whether to include or exclude each data set from further analyses	Intra-site prediction: high accuracy (mean AUC = 0.98); Unseen site prediction: similar accuracy (mean AUC = 0.95)

AUC area under the curve, *QC* quality control, *FD* frame-by-frame displacements, *IQM* image quality metrics, *MRI* magnetic resonance imaging

values according to pre-defined atlases (Destrieux or Desikan-Kikiliani; see section [Image processing](#)). ROIs should be derived a priori based on literature review and with specific hypotheses, as an excessive number of ROIs may lead to problems during multiple comparisons correction. By contrast, vertex-wise group analysis fits a between-subject general linear model at each particular surface vertex to compare values of cortical thickness, surface area, volume or local gyrification. Statistical maps are then overlaid on a template brain as a surface map representing contrast estimates in different colors. This approach may be useful in explorative studies without a priori hypotheses on affected brain regions. Note that FreeSurfer users can conduct a ROI-based analysis with segmented subcortical and parcellated cortical metrics, while the vertex-wise analysis is only available for cortical metrics. Choosing one or the other approach depends on the effects of interest. If regions with significant group differences follow gyral borders defined in a ROI-based atlas, using these pre-defined parcellations may increase statistical power. On the downside, effects in smaller or more specific regions might not be detected as values are averaged within one parcellated ROI (see [Fig. 2](#)). Finally, both methods may be used in one study to complement each other.

Furthermore, researchers and clinicians need to consider several *covariates* during statistical analysis of clinical neurodevelopmental sMRI data as they may confound between-subject comparisons. These closely intertwined variables include age, sex, and different global brain size measures. Researchers and clinicians should thoroughly look into the methods and metrics to use for correction and take them into account when interpreting results. The participants' age matters especially during periods of dynamic neurodevelopment like childhood and puberty, as regional brain metrics change with time (see section [Typical structural brain development](#)). Therefore, researchers and clinicians should examine scaling of regional metrics of interest with age within each group and interpret group differences with respect to the metrics' underlying developmental trajectory. Concerning sex differences in structural brain development, boys have larger head and brain sizes (De Bellis et al., 2001; Paus et al., 2017; Sowell et al., 2002) as compared to girls. Moreover, subtle sex differences have been reported for cortical thinning during adolescence (higher rate of cortical thinning in females in right temporal regions and the left temporoparietal junction; Mutlu et al., 2013). As sex differences during neurodevelopment may be associated with a different age of onset and clinical presentation of neurodevelopmental disorders, sex should always be considered in statistical analysis, for example, by including it as a covariate.

Moreover, accounting for global brain size or not is a particularly important issue in cross-sectional comparative studies, since differences between groups may be driven by interindividual global brain size differences and may not be inherent to the regional metric of interest. In longitudinal developmental

studies, this issue is more complex, as global brain size measures may themselves change with age in children and adolescents. The decision whether or not to correct for global brain size (O'Brien et al., 2011) and which metric and method to use for this correction (Mills et al., 2016) may affect both results and their interpretation. As already discussed (section [Typical and atypical structural brain development](#)), estimations for global brain size differed significantly in past sMRI research, including, for example, cerebral volume (De Bellis et al., 2001), total brain volume (Sowell et al., 2007), whole brain volume, and eTIV (Mills et al., 2016).

Moreover, researchers and clinicians may use three main adjustment methods to correct regional metrics for these global sizes: proportion, analysis of covariance, and residual approach (although the last is rarely used anymore as it may be difficult to interpret; O'Brien et al., 2011). In the proportional method, regional brain metrics of interest are divided by global brain size, leaving a proportional value for further analyses. In group comparisons, this implies the assumption of an identical linear relationship of each brain region and total eTIV or whole brain volume between the groups. If these conditions are not met, calculating a proportion may introduce a bias. The analysis of covariance method accounts for shared variance with global brain size by regression statistics through the inclusion of global brain size as a covariate into the analysis. Due to its simpler implementation the analysis of covariance method is preferred, especially in vertex-wise analysis, as the proportional method would imply adjusting each cortical vertex prior to statistical analysis (Vijayakumar et al., 2018). Notably, these variables may affect results of corrected regional brain metrics differently, as developmental trajectories of cortical gray matter volume from childhood to adulthood differed depending on the adjustment method and metric for brain size (eTIV or whole brain volume; Mills et al., 2016). Furthermore, eTIV and whole brain volume have both been found to account for sex differences in gray matter development during adolescence using the proportional adjustment method, yet only whole brain volume did so when applying the analysis of covariance adjustment method (Mills et al., 2016). Hence, the relationship between brain size and sex should be carefully investigated when including these two covariates in statistical analysis. For a further discussion of the assumption of linear scaling between regional and global brain size, please also see Vijayakumar et al. (2018).

Importantly, an adjustment for global brain sizes is not always necessary. eTIV should be included as a covariate only when analyzing metrics that scale with global brain size, including subcortical volume, cortical volume, and surface area (see FreeSurfer Wiki; <https://surfer.nmr.mgh.harvard.edu/fswiki/eTIV>). As cortical thickness is thought to be independent of global brain size, eTIV should not be considered and the majority of research on cortical thickness chose not to control for global brain size (Vijayakumar et al., 2018).

Correspondingly, Barnes et al. (2010) propose to adjust for age, sex, and global brain size when analyzing gray matter volume and an adjustment for age and sex when analyzing cortical thickness. Taken together, researchers and clinicians are strongly advised to present results for both raw and corrected brain metrics, report metrics and methods used for correction in detail, and discuss how correction may affect interpretation of results.

A big challenge, especially in vertex-wise analysis, is controlling the false positive rate (FPR) of results. The FPR represents the rate of type-I errors (assuming a group difference when in truth there is none), which should be arbitrarily but conventionally less than 5 % (Benjamin et al., 2018). In clinical developmental structural neuroimaging this is aggravated by the tens of thousands single measurements (vertices) in an image of the brain, where a large number of hypotheses are tested simultaneously. For a summary of approaches to this *multiple comparisons* problem, see Bender and Lange (2001) on when and how to adjust for multiple testing and Greve and Fischl (2018) on specifics in surface-based analysis.

Considering vertex-wise group comparisons in FreeSurfer, parametric gaussian-based Monte Carlo (MC) simulation is used by default to compute the FPR. MC simulation extracts sets of contiguous vertices (clusters) after smoothing and thresholding white noise (cluster forming threshold; CFT) in many iterations, usually 10,000. Subsequently, this algorithm computes the p-value of the clusters in the real data (Greve & Fischl, 2018). This parametric approach relies on gaussian spatial smoothness of data and a gaussian distributed underlying noise, which is not always met in neuroimaging data (Eklund et al., 2016). Non-gaussian permutation offers an alternative approach to multiple comparison correction and was previously found to adequately control FPRs (Winkler et al., 2014). Calculations of both MC simulation and permutation depend on the smoothing kernel (2, 4, 6, 8, 10, or 12 mm full-width/ half-max; FWHM) and CFT (0.05, 0.01, 0.005, or 0.001). In a study testing the performance of these two approaches in vertex-wise analysis of cortical thickness, surface area, and volume, Greve and Fischl (2018) suggested the following when applying MC simulation:

- thickness or volume: $CFT \leq 0.001$ and any smoothness level OR $CFT \leq 0.005$ with smoothness level $FWHM > 10$ mm;
- surface area: $CFT \leq 0.001$ and smoothness level $FWHM > 10$ mm

Still, often no vertices survive with such stringent CFTs, greatly reducing power. Permutation allows less stringent CFTs as it adequately controls FPR with all combinations of CFT and smoothness levels. However, permutation comes with some disadvantages including more complicated set up, high computation time, and it requires the data to be exchangeable across participants, a topic discussed in detail

by Winkler et al. (2014). Both MC simulation and permutation are available within FreeSurfer statistical analysis using the `mri_glmfit-sim` script.

For ROI-based analysis, false discovery rate (FDR) is frequently used to handle the multiple-testing problem (Benjamini & Hochberg, 1995). This method can be easily applied with tools like the False Discovery Rate Online Calculator (<https://tools.carbocation.com/FDR>). Similarly, researchers and clinicians may apply Bonferroni adjustment procedures (Bland & Altman, 1995). The original Bonferroni method is fairly simple but at the same time tends to have low power and should be used for a small number of tests (Bender & Lange, 2001). The Bonferroni method and some improvements like the more powerful Holm method (Aickin & Gensler, 1996), and correction for correlated outcome variables are implemented and easy to apply (e.g. via Simple Interactive Statistical Analysis; <https://www.quantitativeskills.com/sisa/calculations/bonhlp.htm>). As a weakness of these procedures, the interpretation of a finding depends on the number of other ROIs going into the FDR or Bonferroni correction. Hence, excessive numbers of ROIs reduce the probability of any significant result and truly important differences may be deemed non-significant. In any case, researchers and clinicians should include and transparently report steps taken to correct for multiple comparisons, and discuss possible interpretations of each result to ease comparability between studies.

Results on altered brain structure in neurodevelopmental disorders, even when statistically significant and corrected for multiple comparisons, often need to be interpreted with caution due to small sample sizes and redundant samples when the same participants are utilized in several reports (Anderson & Kiehl, 2012). As mentioned earlier, recruiting enough children or adolescents with neurodevelopmental disorders who fulfill inclusion criteria is often difficult for smaller laboratories. Smaller samples are usually more manageable, and researchers are able to thoroughly check for exclusion criteria in each participant. Concerning the goal to obtain results from clean samples, this is beneficial. Yet, small samples have been shown to more likely yield false significant findings and produce type-I errors with relatively large effect sizes (Ingre, 2013). Larger sample sizes allow for a better estimation of effect sizes (Ingre, 2013), which are usually quite small in sMRI research.

One way to increase the value of small clinical developmental samples is to work together across multiple centers to increase total sample sizes. Researchers and clinicians can pursue this goal via collaborations in multi-center studies. Another option is to take part in global alliances like the ENIGMA (Enhancing NeuroImaging Genetics through Meta-Analysis) consortium, where over 50 diverse working groups participate in post-hoc data pooling and analyses (Thompson et al., 2020). These efforts accounted for the largest sMRI studies to date in several neurodevelopmental disorders (e.g. Hoogman et al., 2017 for ADHD). In doing so, large scale study findings from the ENIGMA-ADHD working group with

36 cohorts from around the world indicated significant cortical thickness reductions in the fusiform, parahippocampal, and precentral gyrus, as well as the temporal pole in 1018 children with ADHD compared to 1048 controls (Hoogman et al., 2019). Likewise, surface area reductions were found with the largest effect size d being - 0.21 for total surface area (Hoogman et al., 2019). Importantly, such collaborative approaches heavily rely on transparent standardized procedures (study protocol, hardware and sequences used for acquiring sMRI data, quality control procedure, processing tools, statistical analysis methods etc.). These measures are vital to enable replication studies and to include studies in meta-analyses to combine several smaller studies on the same research questions. To further promote collaboration, the Brain Imaging Data Structure (BIDS) has been developed to organize and describe neuroimaging and behavioral data sets (i.e. file naming and folder organization) prior to image processing to facilitate automatic pipelines and quality control protocols in shared data (Gorgolewski et al., 2016).

Please note that all aforementioned statistical analysis approaches fall under easier implemented univariate methods which assume that differences between groups can be observed in isolated ROIs or vertices but ignore relationships between them (Davis & Poldrack, 2013). Additional multivariate methods are also valuable as they allow for more complex analyses concerning cortical networks like identifying cortical regions that co-vary together (Alexander-Bloch et al., 2013) using, for example, nonnegative matrix factorization (Ball, Beare, and Seal 2019) or machine learning (Peng et al., 2013).

To sum up, it is highly advised to follow guidelines for reporting methodological details in clinical developmental sMRI research. An overview of such guidelines was presented by Vijayakumar et al. (2018) spanning sample, acquisition, processing, analysis, and conclusion. As the authors formulated the guidelines specifically for longitudinal sMRI studies investigating typical brain development, we modified them to apply to clinical developmental sMRI (see Textbox 2). They represent comprehensive guidance when implementing transparency of practices in order to reach accurate understanding of neurodevelopmental disorders.

Conclusion

In conclusion, we have discussed several important methodological challenges of clinical developmental sMRI research and provided step-by-step hands-on guidelines how to approach them in the order of an sMRI study: study design, image acquisition, image processing, quality control at different stages, and statistical analysis and interpretation. Variation in these approaches at each step may have contributed to differing results and interpretations of

typical neuromaturation from childhood to adolescence, and of alterations in these processes in neurodevelopmental disorders. Further research should seek for three things: 1) adopt greater transparency of practice and rationales for decisions on study design, image acquisition, image processing, and statistical analysis, 2) conduct analyses and report findings for each brain morphometry metric to achieve a complete picture of brain structure and maturation and possible alterations, 3) empirically examine the effects of varying methods on results, in order to promote the most robust results possible. Adoption of these criteria will help to ensure findings do not only apply to specific samples or methods, but are robust enough to ultimately promote a solid and accurate understanding of neurodevelopmental disorders.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Forschungsgemeinschaft [grant number VE 892/2-1]; the National Institutes of Health [grant number K01MH1087610]; the Research Council of Norway [grant numbers 230345, 288083, 223273]; and the South-Eastern Norway Regional Health Authority [grant number 2019069]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ai, L., Craddock, R. C., Tottenham, N., Dyke, J. P., Lim, R., Colcombe, S., Milham, M., & Franco, A. R. (2020). Is it Time to Switch Your T1W Sequence? Assessing the Impact of Prospective Motion Correction on the Reliability and Quality of Structural Imaging. *NeuroImage*, 117585. <https://doi.org/10.1016/j.neuroimage.2020.117585>
- Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health*, 86(5), 726–728. <https://doi.org/10.2105/ajph.86.5.726>
- Alexander-Bloch, A. F., Raznahan, A., Vandekar, S. N., Seidlitz, J., Lu, Z., Matthias, S. R., Knowles, E., Mollon, J., Rodrigue, A., Curran, J. E., Göring, H. H. H., Satterthwaite, T. D., Gur, R. E., Bassett, D. S., Hoftman, G. D., Pearlson, G., Shinohara, R. T., Liu, S., Fox, P. T., & Glahn, D. C. (2020). Imaging local genetic influences on cortical folding. *Proceedings of the National Academy of Sciences*, 117(13), 7430–7436. <https://doi.org/10.1073/pnas.1912064117>
- Alexander-Bloch, A., Raznahan, A., Bullmore, E., & Giedd, J. (2013). The Convergence of Maturation Change and Structural Covariance in Human Cortical Networks. *Journal of Neuroscience*, 33(7), 2889–2899. <https://doi.org/10.1523/JNEUROSCI.3554-12.2013>

- Ambrosino, S., de Zeeuw, P., Wierenga, L. M., van Dijk, S., & Durston, S. (2017). What can Cortical Development in Attention-Deficit/Hyperactivity Disorder Teach us About the Early Developmental Mechanisms Involved? *Cerebral Cortex*, 27(9), 4624–4634. <https://doi.org/10.1093/cercor/bhx182>
- Amlien, I. K., Fjell, A. M., Tamnes, C. K., Grydeland, H., Krogsrud, S. K., Chaplin, T. A., Rosa, M. G. P., & Walhovd, K. B. (2016). Organizing Principles of Human Cortical Development—Thickness and Area from 4 to 30 Years: Insights from Comparative Primate Neuroanatomy. *Cerebral Cortex*, 26(1), 257–267. <https://doi.org/10.1093/cercor/bhu214>
- Anderson, N. E., & Kiehl, K. A. (2012). The psychopath magnetized: Insights from brain imaging. *Trends in Cognitive Sciences*, 16(1), 52–60. <https://doi.org/10.1016/j.tics.2011.11.008>
- Ashburner, J. (2009). Computational anatomy with the SPM software. *Magnetic Resonance Imaging*, 27(8), 1163–1174. <https://doi.org/10.1016/j.mri.2009.01.006>
- Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, 11(6), 805–821. <https://doi.org/10.1006/nimg.2000.0582>
- Backhausen, L. L., Herting, M. M., Buse, J., Roessner, V., Smolka, M. N., & Vetter, N. C. (2016). Quality Control of Structural MRI Images Applied Using FreeSurfer—A Hands-On Workflow to Rate Motion Artifacts. *Frontiers in Neuroscience*, 10. <https://doi.org/10.3389/fnins.2016.00558>
- Ball, G., Beare, R., & Seal, M. L. (2019). Charting shared developmental trajectories of cortical thickness and structural connectivity in childhood and adolescence. *Human Brain Mapping*, 40(16), 4630–4644. <https://doi.org/10.1002/hbm.24726>
- Barisano, G., Seppehrband, F., Ma, S., Jann, K., Cabeen, R., Wang, D. J., Toga, A. W., & Law, M. (2018). Clinical 7 T MRI: Are we there yet? A review about magnetic resonance imaging at ultra-high field. *The British Journal of Radiology*, 92(1094), 20180492. <https://doi.org/10.1259/bjr.20180492>
- Barkley, R. A. (2014). *Attention-Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment*. Guilford Publications.
- Barnes, J., Ridgway, G. R., Bartlett, J., Henley, S. M., Lehmann, M., Hobbs, N., Clarkon, M. J., MacManus, D. G., Ourselin, S., & Fox, N. C. (2010). Head size, age and gender adjustment in MRI studies: A necessary nuisance? *Neuroimage*, 53(4), 1244–1255. <https://doi.org/10.1016/j.neuroimage.2010.06.025>
- Bednarz, H. M., & Kana, R. K. (2018). *Neuroscience & Biobehavioral Reviews*, 90, 50–69. <https://doi.org/10.1016/j.neubiorev.2018.03.025>
- Beer, J. C., Tustison, N. J., Cook, P. A., Davatzikos, C., Sheline, Y. I., Shinohara, R. T., Linn, K. A., & Initiative, for the A. D. N. (2020). Longitudinal ComBat: A Method for Harmonizing Longitudinal Multi-scanner Imaging Data. *BioRxiv*, 868810. <https://doi.org/10.1101/868810>
- Belloni, E., Haacke, E., Coleman, P., Sacco, D., Steiger, D., & Gargaro, R. (1986). MR artifacts: A review. *American Journal of Roentgenology*, 147(6), 1271–1281. <https://doi.org/10.2214/ajr.147.6.1271>
- Bender, R., & Lange, S. (2001). Adjusting for multiple testing—When and how? *Journal of Clinical Epidemiology*, 54(4), 343–349. [https://doi.org/10.1016/S0895-4356\(00\)00314-0](https://doi.org/10.1016/S0895-4356(00)00314-0)
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., Cesarini, D., Chambers, C. D., Clyde, M., Cook, T. D., Boeck, P. D., Dienes, Z., Dreber, A., Easwaran, K., Efferson, C., & Johnson, V. E. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10. <https://doi.org/10.1038/s41562-017-0189-z>
- 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>
- Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: The Bonferroni method. *BMJ : British Medical Journal*, 310(6973), 170.
- Blesson, A., & Cohen, J. S. (2020). Genetic Counseling in Neurodevelopmental Disorders. *Cold Spring Harbor Perspectives in Medicine*, 10(4). <https://doi.org/10.1101/cshperspect.a036533>
- Blumenthal, J. D., Zijdenbos, A., Molloy, E., & Giedd, J. N. (2002). Motion artifact in magnetic resonance imaging: Implications for automated analysis. *NeuroImage*, 16(1), 89–92. <https://doi.org/10.1006/nimg.2002.1076>
- Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, 23(2), 724–738. <https://doi.org/10.1016/j.neuroimage.2004.06.018>
- Canino, G., Polanczyk, G., Bauermeister, J. J., Rohde, L. A., & Frick, P. J. (2010). Does the prevalence of CD and ODD vary across cultures? *Social Psychiatry and Psychiatric Epidemiology*, 45(7), 695–704. <https://doi.org/10.1007/s00127-010-0242-y>
- Costa, A. L. F., Appenzeller, S., Yasuda, C.-L., Pereira, F. R., Zanardi, V. A., & Cendes, F. (2009). Artifacts in brain magnetic resonance imaging due to metallic dental objects. *Medicina Oral, Patologia Oral Y Cirugia Bucal*, 14(6), E278-282.
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- Davatzikos, C. (2004). Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. *NeuroImage*, 23(1), 17–20. <https://doi.org/10.1016/j.neuroimage.2004.05.010>
- Davis, T., & Poldrack, R. A. (2013). Measuring neural representations with fMRI: Practices and pitfalls. *Annals of the New York Academy of Sciences*, 1296, 108–134. <https://doi.org/10.1111/nyas.12156>
- De Bellis, M. D., Keshavan, M. S., Beers, S. R., Hall, J., Frustaci, K., Masalehdan, A., Noll, J., & Boring, A. M. (2001). Sex Differences in Brain Maturation during Childhood and Adolescence. *Cerebral Cortex*, 11(6), 552–557. <https://doi.org/10.1093/cercor/11.6.552>
- 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>
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1), 1–15. <https://doi.org/10.1016/j.neuroimage.2010.06.010>
- Dong, H.-M., Castellanos, F. X., Yang, N., Zhang, Z., Zhou, Q., He, Y., Zhang, L., Xu, T., Holmes, A. J., Thomas Yeo, B. T., Chen, F., Wang, B., Beckmann, C., White, T., Sporns, O., Qiu, J., Feng, T., Chen, A., Liu, X., & Zuo, X.-N. (2020). Charting brain growth in tandem with brain templates at school age. *Science Bulletin*, 65(22), 1924–1934. <https://doi.org/10.1016/j.scib.2020.07.027>
- Ducharme, S., Albaugh, M. D., Nguyen, T.-V., Hudziak, J. J., Mateos-Pérez, J. M., Labbe, A., Evans, A. C., & Karama, S. (2016). Trajectories of cortical thickness maturation in normal brain development—The importance of quality control procedures. *NeuroImage*, 125, 267–279. <https://doi.org/10.1016/j.neuroimage.2015.10.010>
- Durston, S., Hulshoff Pol, H. E., Casey, B. J., Giedd, J. N., Buitelaar, J. K., & Van Engeland, H. (2001). Anatomical MRI of the Developing Human Brain: What Have We Learned? *Journal of the*

- American Academy of Child & Adolescent Psychiatry, 40(9), 1012–1020. <https://doi.org/10.1097/00004583-200109000-00009>
- Ecker, C., Bookheimer, S. Y., & Murphy, D. G. M. (2015). Neuroimaging in autism spectrum disorder: Brain structure and function across the lifespan. *The Lancet. Neurology*, 14(11), 1121–1134. [https://doi.org/10.1016/S1474-4422\(15\)00050-2](https://doi.org/10.1016/S1474-4422(15)00050-2)
- Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, 113(28), 7900–7905. <https://doi.org/10.1073/pnas.1602413113>
- Essen, D. C. V. (1997). A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature*, 385(6614), 313–318. <https://doi.org/10.1038/385313a0>
- Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., & Gorgolewski, K. J. (2017). MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLOS ONE*, 12(9), e0184661. <https://doi.org/10.1371/journal.pone.0184661>
- Fairchild, G., Toschi, N., Hagan, C. C., Goodyer, I. M., Calder, A. J., & Passamonti, L. (2015). Cortical thickness, surface area, and folding alterations in male youths with conduct disorder and varying levels of callous–unemotional traits. *NeuroImage: Clinical*, 8, 253–260. <https://doi.org/10.1016/j.nicl.2015.04.018>
- Fischl, B., Liu, A., & Dale, A. M. (2001). Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Transactions on Medical Imaging*, 20(1), 70–80. <https://doi.org/10.1109/42.906426>
- Fischl, B. (2012). *FreeSurfer. NeuroImage*, 62(2), 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021>
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20), 11050–11055. <https://doi.org/10.1073/pnas.200033797>
- 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: Automated labeling of neuroanatomical structures in the human brain. *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)
- Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., & Dale, A. M. (2004). Sequence-independent segmentation of magnetic resonance images. *NeuroImage*, 23, S69–S84. <https://doi.org/10.1016/j.neuroimage.2004.07.016>
- Forde, N. J., Ronan, L., Zwiars, M. P., Alexander-Bloch, A. F., Faraone, S. V., Oosterlaan, J., Heslenfeld, D. J., Hartman, C. A., Buitelaar, J. K., & Hoekstra, P. J. (2017). No Association between Cortical Gyrfication or Intrinsic Curvature and Attention-deficit/Hyperactivity Disorder in Adolescents and Young Adults. *Frontiers in Neuroscience*, 11. <https://doi.org/10.3389/fnins.2017.00218>
- Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, 167, 104–120. <https://doi.org/10.1016/j.neuroimage.2017.11.024>
- Giedd, J. N., Snell, J. W., Lange, N., Rajapakse, J. C., Casey, B. J., Kozuch, P. L., Vaituzis, A. C., Vauss, Y. C., Hamburger, S. D., Kaysen, D., & Rapoport, J. L. (1996). Quantitative Magnetic Resonance Imaging of Human Brain Development: Ages 4–18. *Cerebral Cortex*, 6(4), 551–559. <https://doi.org/10.1093/cercor/6.4.551>
- Gilmore, J. H., Shi, F., Woolson, S. L., Knickmeyer, R. C., Short, S. J., Lin, W., Zhu, H., Hamer, R. M., Styner, M., & Shen, D. (2012). Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cerebral Cortex (New York, N.Y.: 1991)*, 22(11), 2478–2485. <https://doi.org/10.1093/cercor/bhr327>
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80, 105–124. <https://doi.org/10.1016/j.neuroimage.2013.04.127>
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, 101(21), 8174–8179. <https://doi.org/10.1073/pnas.0402680101>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glattard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., & Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- Greene, D. J., Black, K. J., & Schlaggar, B. L. (2016). Considerations for MRI study design and implementation in pediatric and clinical populations. *Developmental Cognitive Neuroscience*, 18, 101–112. <https://doi.org/10.1016/j.dcn.2015.12.005>
- Greve, D. N. (2011). An Absolute Beginner's Guide to Surface- and Voxel-based Morphometric Analysis. *Proceedings of the International Society for Magnetic Resonance in Medicine*, 19(33).
- Greve, D. N., & Fischl, B. (2018). False positive rates in surface-based anatomical analysis. *NeuroImage*, 171, 6–14. <https://doi.org/10.1016/j.neuroimage.2017.12.072>
- Greven, C. U., Bralten, J., Mennes, M., O'Dwyer, L., van Hulzen, K. J. E., Rommelse, N., Schweren, L. J. S., Hoekstra, P. J., Hartman, C. A., Heslenfeld, D., Oosterlaan, J., Faraone, S. V., Franke, B., Zwiars, M. P., Arias-Vasquez, A., & Buitelaar, J. K. (2015). Developmentally Stable Whole-Brain Volume Reductions and Developmentally Sensitive Caudate and Putamen Volume Alterations in Those With Attention-Deficit/Hyperactivity Disorder and Their Unaffected Siblings. *JAMA Psychiatry*, 72(5), 490–499. <https://doi.org/10.1001/jamapsychiatry.2014.3162>
- Heinen, R., Bouvy, W. H., Mendrik, A. M., Viergever, M. A., Biessels, G. J., & de Bresser, J. (2016). Robustness of Automated Methods for Brain Volume Measurements across Different MRI Field Strengths. *PLOS ONE*, 11(10), e0165719. <https://doi.org/10.1371/journal.pone.0165719>
- Herting, M. M., Johnson, C., Mills, K. L., Vijayakumar, N., Dennison, M., Liu, C., Goddings, A.-L., Dahl, R. E., Sowell, E. R., Whittle, S., Allen, N. B., & Tamnes, C. K. (2018). Development of subcortical volumes across adolescence in males and females: A multisample study of longitudinal changes. *NeuroImage*, 172, 194–205. <https://doi.org/10.1016/j.neuroimage.2018.01.020>
- Hoogman, M., Bralten, J., Hibar, D. P., Mennes, M., Zwiars, M. P., Schweren, L. S. J., van Hulzen, K. J. E., Medland, S. E., Shumskaya, E., Jahanshad, N., de Zeeuw, P., Szekely, E., Sudre, G., Wolfers, T., Onnink, A. M. H., Dammers, J. T., Mostert, J. C., Vives-Gilabert, Y., Kohls, G., & Franke, B. (2017). Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis. *The Lancet Psychiatry*, 4(4), 310–319. [https://doi.org/10.1016/S2215-0366\(17\)30049-4](https://doi.org/10.1016/S2215-0366(17)30049-4)
- Hoogman, M., Muetzel, R., Guimaraes, J. P., Shumskaya, E., Mennes, M., Zwiars, M. P., Jahanshad, N., Sudre, G., Mostert, J., Wolfers, T., Earl, E. A., Vila, J. C. S., Vives-Gilabert, Y., Khadka, S., Novotny, S. E., Hartman, C. A., Heslenfeld, D. J., Schweren, L. J. S., Ambrosino, S., & Franke, B. (2019). Brain imaging of the cortex in ADHD: A coordinated analysis of large-scale clinical and population-based samples. *The American Journal of Psychiatry*, 176(7), 531–542. <https://doi.org/10.1176/appi.ajp.2019.18091033>

- Huebner, T., Vloet, T. D., Marx, I., Konrad, K., Fink, G. R., Herpertz, S. C., & Herpertz-Dahlmann, B. (2008). Morphometric brain abnormalities in boys with conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 47(5), 540–547. <https://doi.org/10.1097/CHI.0b013e3181676545>
- Hupé, J.-M. (2015). Statistical inferences under the Null hypothesis: Common mistakes and pitfalls in neuroimaging studies. *Frontiers in Neuroscience*, 9. <https://doi.org/10.3389/fnins.2015.00018>
- Hyatt, C. J., Haney-Caron, E., & Stevens, M. C. (2012). Cortical Thickness and Folding Deficits in Conduct-Disordered Adolescents. *Biological Psychiatry*, 72(3), 207–214. <https://doi.org/10.1016/j.biopsych.2011.11.017>
- Hyman, S. E. (2010). The Diagnosis of Mental Disorders: The Problem of Reification. *Annual Review of Clinical Psychology*, 6(1), 155–179. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091532>
- Ingre, M. (2013). Why small low-powered studies are worse than large high-powered studies and how to protect against “trivial” findings in research: Comment on Friston (2012). *NeuroImage*, 81, 496–498.
- 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. *Human Brain Mapping*, 36(9), 3472–3485. <https://doi.org/10.1002/hbm.22856>
- Jovicich, J., Marizzoni, M., Sala-Llloch, 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>
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602. <https://doi.org/10.1001/archpsyc.62.6.593>
- Klapwijk, E. T., van de Kamp, F., van der Meulen, M., Peters, S., & Wierenga, L. M. (2019a). Qoala-T: A supervised-learning tool for quality control of FreeSurfer segmented MRI data. *NeuroImage*, 189, 116–129. <https://doi.org/10.1016/j.neuroimage.2019.01.014>
- Klapwijk, E. T., van den Bos, W., Tamnes, C. K., Mills, K. L., & Raschle, N. (2019b). Opportunities for increased reproducibility and replicability of developmental cognitive neuroscience. *PsyArXiv*. <https://doi.org/10.31234/osf.io/fxjzt>
- Kraemer, H. C., Yesavage, J. A., Taylor, J. L., & Kupfer, D. (2000). How Can We Learn About Developmental Processes From Cross-Sectional Studies, or Can We? *American Journal of Psychiatry*, 157(2), 163–171. <https://doi.org/10.1176/appi.ajp.157.2.163>
- Lee, H., Nakamura, K., Narayanan, S., Brown, R. A., & Arnold, D. L. (2019). Estimating and accounting for the effect of MRI scanner changes on longitudinal whole-brain volume change measurements. *NeuroImage*, 184, 555–565. <https://doi.org/10.1016/j.neuroimage.2018.09.062>
- Llinares-Benadero, C., & Borrell, V. (2019). Deconstructing cortical folding: Genetic, cellular and mechanical determinants. *Nature Reviews Neuroscience*, 20(3), 161–176. <https://doi.org/10.1038/s41583-018-0112-2>
- Marquand, A. F., Kia, S. M., Zabihi, M., Wolfers, T., Buitelaar, J. K., & Beckmann, C. F. (2019). Conceptualizing mental disorders as deviations from normative functioning. *Molecular Psychiatry*, 24(10), 1415–1424. <https://doi.org/10.1038/s41380-019-0441-1>
- Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: Awareness and response. *Brain Structure and Function*, 214(5), 535–549. <https://doi.org/10.1007/s00429-010-0265-x>
- Mills, K. L., Goddings, A.-L., Herting, M. M., Meuwese, R., Blakemore, S.-J., Crone, E. A., Dahl, R. E., Güroğlu, B., Raznahan, A., Sowell, E. R., & Tamnes, C. K. (2016). Structural brain development between childhood and adulthood: Convergence across four longitudinal samples. *NeuroImage*, 141, 273–281. <https://doi.org/10.1016/j.neuroimage.2016.07.044>
- 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>
- Mutlu, A. K., Schneider, M., Debbané, M., Badoud, D., Eliez, S., & Schaer, M. (2013). Sex differences in thickness, and folding developments throughout the cortex. *NeuroImage*, 82, 200–207. <https://doi.org/10.1016/j.neuroimage.2013.05.076>
- Nakao, T., Radua, J., Rubia, K., & Mataix-Cols, D. (2011). Gray Matter Volume Abnormalities in ADHD: Voxel-Based Meta-Analysis Exploring the Effects of Age and Stimulant Medication. *American Journal of Psychiatry*, 168(11), 1154–1163. <https://doi.org/10.1176/appi.ajp.2011.11020281>
- Nigg, J. T. (2012). Future Directions in ADHD Etiology Research. *Journal of Clinical Child & Adolescent Psychology*, 41(4), 524–533. <https://doi.org/10.1080/15374416.2012.686870>
- Nissim, N. R., O’Shea, A. M., Bryant, V., Porges, E. C., Cohen, R., & Woods, A. J. (2017). Frontal Structural Neural Correlates of Working Memory Performance in Older Adults. *Frontiers in Aging Neuroscience*, 8. <https://doi.org/10.3389/fnagi.2016.00328>
- Noordermeer, S. D., Luman, M., Greven, C. U., Veroude, K., Faraone, S. V., Hartman, C. A., Hoekstra, P. J., Franke, B., Buitelaar, J. K., Heslenfeld, D. J., & Oosterlaan, J. (2017). Structural Brain Abnormalities of Attention-Deficit/Hyperactivity Disorder With Oppositional Defiant Disorder. *Biological Psychiatry*, 82(9), 642–650. <https://doi.org/10.1016/j.biopsych.2017.07.008>
- Noordermeer, S. D., Luman, M., & Oosterlaan, J. (2016). A systematic review and meta-analysis of neuroimaging in oppositional defiant disorder (ODD) and conduct disorder (CD) taking attention-deficit hyperactivity disorder (ADHD) into account. *Neuropsychology Review*, 26(1), 44–72.
- O’Brien, L. M., Ziegler, D. A., Deutsch, C. K., Frazier, J. A., Herbert, M. R., & Locascio, J. J. (2011). Statistical adjustments for brain size in volumetric neuroimaging studies: Some practical implications in methods. *Psychiatry Research: Neuroimaging*, 193(2), 113–122. <https://doi.org/10.1016/j.psychresns.2011.01.007>
- Ooi, M. B., Krueger, S., Thomas, W. J., Swaminathan, S. V., & Brown, T. R. (2009). Prospective real-time correction for arbitrary head motion using active markers. *Magnetic Resonance in Medicine*, 62(4), 943–954. <https://doi.org/10.1002/mrm.22082>
- Opel, N., Goltermann, J., Hermesdorf, M., Berger, K., Baune, B. T., & Dannlowski, U. (2020). Cross-Disorder Analysis of Brain Structural Abnormalities in Six Major Psychiatric Disorders: A Secondary Analysis of Mega- and Meta-analytical Findings From the ENIGMA Consortium. *Biological Psychiatry*, 88(9), 678–686. <https://doi.org/10.1016/j.biopsych.2020.04.027>
- Paus, T., Wong, A.P.-Y., Syme, C., & Pausova, Z. (2017). Sex differences in the adolescent brain and body: Findings from the saguenay youth study. *Journal of Neuroscience Research*, 95(1–2), 362–370. <https://doi.org/10.1002/jnr.23825>
- Peng, X., Lin, P., Zhang, T., & Wang, J. (2013). Extreme Learning Machine-Based Classification of ADHD Using Brain Structural MRI Data. *PLoS ONE*, 8(11). <https://doi.org/10.1371/journal.pone.0079476>
- Popescu, V., Schoonheim, M. M., Versteeg, A., Chaturvedi, N., Jonker, M., de Menezes, R. X., Garre, F. G., Uitdehaag, B. M., Barkhof, F.,

- & Vrenken, H. (2016). Grey Matter Atrophy in Multiple Sclerosis: Clinical Interpretation Depends on Choice of Analysis Method. *PLoS One*, *11*(1), e0143942. <https://doi.org/10.1371/journal.pone.0143942>
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. <https://doi.org/10.1016/j.neuroimage.2013.08.048>
- Pua, E. P. K., Barton, S., Williams, K., Craig, J. M., & Seal, M. L. (2020). Individualised MRI training for paediatric neuroimaging: A child-focused approach. *Developmental Cognitive Neuroscience*, *41*, 100750. <https://doi.org/10.1016/j.dcn.2019.100750>
- Raschle, Lee, M., Buechler, R., Christodoulou, J. A., Chang, M., Vakil, M., Sterling, P. L., & Gaab, N. (2009). Making MR Imaging Child's Play—Pediatric Neuroimaging Protocol, Guidelines and Procedure. *Journal of Visualized Experiments: JoVE*, *29*. <https://doi.org/10.3791/1309>
- Raschle, N., Zuk, J., Ortiz-Mantilla, S., Sliva, D. D., Franceschi, A., Grant, P. E., Benasich, A. A., & Gaab, N. (2012). Pediatric neuroimaging in early childhood and infancy: Challenges and practical guidelines. *Annals of the New York Academy of Sciences*, *1252*, 43–50. <https://doi.org/10.1111/j.1749-6632.2012.06457.x>
- Rauch, S. L. (2005). Neuroimaging and attention-deficit/hyperactivity disorder in the 21st century: What to consider and how to proceed. *Biological Psychiatry*, *57*(11), 1261–1262. <https://doi.org/10.1016/j.biopsych.2005.02.014>
- Raznahan, A., Shaw, P., Lalonde, F., Stockman, M., Wallace, G. L., Greenstein, D., Clasen, L., Gogtay, N., & Giedd, J. N. (2011). How Does Your Cortex Grow? *Journal of Neuroscience*, *31*(19), 7174–7177. <https://doi.org/10.1523/JNEUROSCI.0054-11.2011>
- Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, *53*(4), 1181–1196. <https://doi.org/10.1016/j.neuroimage.2010.07.020>
- Reuter, M., Tisdall, M. D., Qureshi, A., Buckner, R. L., van der Kouwe, A. J. W., & Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage*, *107*, 107–115. <https://doi.org/10.1016/j.neuroimage.2014.12.006>
- Rogers, J. C., & Brito, S. A. D. (2016). Cortical and Subcortical Gray Matter Volume in Youths With Conduct Problems: A Meta-analysis. *JAMA Psychiatry*, *73*(1), 64–72. <https://doi.org/10.1001/jamapsychiatry.2015.2423>
- Rosen, A. F. G., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., Ciric, R., Cook, P. A., Davatzikos, C., Elliott, M. A., Garcia de La Garza, A., Gennatas, E. D., Quarmley, M., Schmitt, J. E., Shinohara, R. T., Tisdall, M. D., Craddock, R. C., Gur, R. E., Gur, R. C., & Satterthwaite, T. D. (2018). Quantitative assessment of structural image quality. *NeuroImage*, *169*, 407–418. <https://doi.org/10.1016/j.neuroimage.2017.12.059>
- Sabuncu, M. R., Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., & Fischl, B. (2014). Event time analysis of longitudinal neuroimaging data. *NeuroImage*, *97*, 9–18. <https://doi.org/10.1016/j.neuroimage.2014.04.015>
- Sanislow, C. A., Ferrante, M., Pacheco, J., Rudorfer, M. V., & Morris, S. E. (2019). Advancing Translational Research Using NIMH Research Domain Criteria and Computational Methods. *Neuron*, *101*(5), 779–782. <https://doi.org/10.1016/j.neuron.2019.02.024>
- Sarkar, S., Daly, E., Feng, Y., Ecker, C., Craig, M. C., Harding, D., Deeley, Q., & Murphy, D. G. M. (2015). Reduced cortical surface area in adolescents with conduct disorder. *European Child & Adolescent Psychiatry*, *24*(8), 909–917. <https://doi.org/10.1007/s00787-014-0639-3>
- Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughhead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. <https://doi.org/10.1016/j.neuroimage.2012.08.052>
- 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>
- Schaer, M., Cuadra, M. B., Tamarit, L., Lazeyras, F., Eliez, S., & Thiran, J.-P. (2008). A Surface-Based Approach to Quantify Local Cortical Gyrfication. *IEEE Transactions on Medical Imaging*, *27*(2), 161–170. <https://doi.org/10.1109/TMI.2007.903576>
- Semrud-Clikeman, M., Pliszka, S. R., Bledsoe, J., & Lancaster, J. (2014). Volumetric MRI Differences in Treatment Naïve and Chronically Treated Adolescents With ADHD-Combined Type. *Journal of Attention Disorders*, *18*(6), 511–520. <https://doi.org/10.1177/1087054712443158>
- Shaw, P., De Rossi, P., Watson, B., Wharton, A., Greenstein, D., Raznahan, A., Sharp, W., Lerch, J. P., & Chakravarty, M. M. (2014). Mapping the Development of the Basal Ganglia in Children With Attention-Deficit/Hyperactivity Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(7), 780–789. <https://doi.org/10.1016/j.jaac.2014.05.003>
- Shaw, P., Eckstrand, K., Sharp, W., Blumenthal, J., Lerch, J. P., Greenstein, D., Clasen, L., Evans, A., Giedd, J., & Rapoport, J. L. (2007). Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. *Proceedings of the National Academy of Sciences*, *104*(49), 19649–19654. <https://doi.org/10.1073/pnas.0707741104>
- Shaw, P., Malek, M., Watson, B., Greenstein, D., de Rossi, P., & Sharp, W. (2013). Trajectories of Cerebral Cortical Development in Childhood and Adolescence and Adult Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *74*(8), 599–606. <https://doi.org/10.1016/j.biopsych.2013.04.007>
- Shaw, P., Malek, M., Watson, B., Sharp, W., Evans, A., & Greenstein, D. (2012). Development of Cortical Surface Area and Gyrfication in Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *72*(3), 191–197. <https://doi.org/10.1016/j.biopsych.2012.11.031>
- Sowell, E. R., Peterson, B. S., Kan, E., Woods, R. P., Yoshii, J., Bansal, R., Xu, D., Zhu, H., Thompson, P. M., & Toga, A. W. (2007). Sex Differences in Cortical Thickness Mapped in 176 Healthy Individuals between 7 and 87 Years of Age. *Cerebral Cortex*, *17*(7), 1550–1560. <https://doi.org/10.1093/cercor/bhl066>
- Sowell, E. R., Thompson, P. M., & Toga, A. W. (2004). Mapping changes in the human cortex throughout the span of life. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *10*(4), 372–392. <https://doi.org/10.1177/1073858404263960>
- Sowell, E. R., Trauner, D. A., Gamst, A., & Jernigan, T. L. (2002). Development of cortical and subcortical brain structures in childhood and adolescence: A structural MRI study. *Developmental Medicine & Child Neurology*, *44*(1), 4–16. <https://doi.org/10.1111/j.1469-8749.2002.tb00253.x>
- Stevens, M. C., & Haney-Caron, E. (2012). Comparison of brain volume abnormalities between ADHD and conduct disorder in adolescence. *Journal of Psychiatry & Neuroscience: JPN*, *37*(6), 389–398. <https://doi.org/10.1503/jpn.110148>
- Storsve, A. B., Fjell, A. M., Tamnes, C. K., Westlye, L. T., Overbye, K., Aasland, H. W., & Walhovd, K. B. (2014). Differential Longitudinal Changes in Cortical Thickness, Surface Area and Volume across the Adult Life Span: Regions of Accelerating and Decelerating Change. *Journal of Neuroscience*, *34*(25), 8488–8498. <https://doi.org/10.1523/JNEUROSCI.0391-14.2014>
- Tamnes, C. K., Hertzog, 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 interrelated longitudinal changes in cortical volume, surface area and thickness. *Journal of Neuroscience*, 3302–3316. <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>
- Thompson, P. M., Jahanshad, N., Ching, C. R. K., Salminen, L. E., Thomopoulos, S. I., Bright, J., Baune, B. T., Bertolin, S., Bralten, J., Bruin, W. B., Bilow, R., Chen, J., Chye, Y., Dannlowski, U., de Kovel, C. G. F., Donohoe, G., Eyster, L. T., Faraone, S. V., Favre, P., & Zelman, V. (2020). ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Translational Psychiatry*, 10(1), 1–28. <https://doi.org/10.1038/s41398-020-0705-1>
- Tijssen, R. H. N., Jansen, J. F. A., & Backes, W. H. (2009). Assessing and minimizing the effects of noise and motion in clinical DTI at 3 T. *Human Brain Mapping*, 30(8), 2641–2655. <https://doi.org/10.1002/hbm.20695>
- Tisdall, M. D., Reuter, M., Qureshi, A., Buckner, R. L., Fischl, B., & van der Kouwe, A. J. W. (2016). Prospective motion correction with volumetric navigators (vNavs) reduces the bias and variance in brain morphometry induced by subject motion. *NeuroImage*, 127, 11–22. <https://doi.org/10.1016/j.neuroimage.2015.11.054>
- 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>
- Vetter, N. C., Backhausen, L. L., Buse, J., Roessner, V., & Smolka, M. N. (2020). Altered brain morphology in boys with attention deficit hyperactivity disorder with and without comorbid conduct disorder/oppositional defiant disorder. *Human Brain Mapping*, 41(4), 973–983. <https://doi.org/10.1002/hbm.24853>
- Vijayakumar, N., Allen, N. B., Youssef, G., Dennison, M., Yücel, M., Simmons, J. G., & Whittle, S. (2016). Brain development during adolescence: A mixed-longitudinal investigation of cortical thickness, surface area, and volume. *Human Brain Mapping*, 37(6), 2027–2038. <https://doi.org/10.1002/hbm.23154>
- 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>
- Walhovd, K. B., Fjell, A. M., Giedd, J., Dale, A. M., & Brown, T. T. (2017). Through Thick and Thin: A Need to Reconcile Contradictory Results on Trajectories in Human Cortical Development. *Cerebral Cortex*, 27(2). <https://doi.org/10.1093/cercor/bhv301>
- Wallace, G. L., White, S. F., Robustelli, B., Sinclair, S., Hwang, S., Martin, A., & Blair, R. J. R. (2014). Cortical and Subcortical Abnormalities in Youths With Conduct Disorder and Elevated Callous-Unemotional Traits. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53(4), 456–465. <https://doi.org/10.1016/j.jaac.2013.12.008>
- Westbrook, C., & Talbot, J. (2018). *MRI in Practice*. John Wiley & Sons.
- Westlye, L. T., Walhovd, K. B., Dale, A. M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., Grydeland, H., Tamnes, C. K., Ostby, Y., & Fjell, A. M. (2010). Life-span changes of the human brain white matter: Diffusion tensor imaging (DTI) and volumetry. *Cerebral Cortex*, 20(9), 2055–2068. <https://doi.org/10.1093/cercor/bhp280>
- White, N., Roddey, C., Shankaranarayanan, A., Han, E., Rettmann, D., Santos, J., Kuperman, J., & Dale, A. (2010). PROMO: Real-time prospective motion correction in MRI using image-based tracking. *Magnetic Resonance in Medicine*, 63(1), 91–105. <https://doi.org/10.1002/mrm.22176>
- Wierenga, L. M., Langen, M., Oranje, B., & Durston, S. (2014). Unique developmental trajectories of cortical thickness and surface area. *NeuroImage*, 87, 120–126. <https://doi.org/10.1016/j.neuroimage.2013.11.010>
- Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, 53(3), 1135–1146. <https://doi.org/10.1016/j.neuroimage.2009.12.028>
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, 92, 381–397. <https://doi.org/10.1016/j.neuroimage.2014.01.060>
- Wood, M. L., & Henkelman, R. M. (1985). MR image artifacts from periodic motion. *Medical Physics*, 12(2), 143–151. <https://doi.org/10.1118/1.595782>
- Yu, Q., McCall, D. M., Homayouni, R., Tang, L., Chen, Z., Schoff, D., Nishimura, M., Raz, S., & Ofen, N. (2018). Age-associated increase in mnemonic strategy use is linked to prefrontal cortex development. *NeuroImage*, 181, 162–169. <https://doi.org/10.1016/j.neuroimage.2018.07.008>
- Yuan, W., Altaye, M., Ret, J., Schmithorst, V., Byars, A. W., Plante, E., & Holland, S. K. (2009). Quantification of head motion in children during various fMRI language tasks. *Human Brain Mapping*, 30(5), 1481–1489. <https://doi.org/10.1002/hbm.20616>
- Zaitsev, M., Dold, C., Sakas, G., Hennig, J., & Speck, O. (2006). Magnetic resonance imaging of freely moving objects: Prospective real-time motion correction using an external optical motion tracking system. *NeuroImage*, 31(3), 1038–1050. <https://doi.org/10.1016/j.neuroimage.2006.01.039>
- Zaitsev, M., Maclaren, J., & Herbst, M. (2015). Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4), 887–901. <https://doi.org/10.1002/jmri.24850>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.