Review article

Advances, challenges, and promises in pediatric neuroimaging of neurodevelopmental disorders

Haley M. Bednarz, Rajesh K. Kana*

Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, USA

A R T I C L E   I N F O

Keywords:
Neurodevelopmental disorders
Autism spectrum disorder
Attention-deficit/hyperactivity disorder
Tourette's syndrome
Neuroimaging
Magnetic resonance imaging

A B S T R A C T

Recent years have witnessed the proliferation of neuroimaging studies of neurodevelopmental disorders (NDDs), particularly of children with autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and Tourette's syndrome (TS). Neuroimaging offers immense potential in understanding the biology of these disorders, and how it relates to clinical symptoms. Neuroimaging techniques, in the long run, may help identify neurobiological markers to assist clinical diagnosis and treatment. However, methodological challenges have affected the progress of clinical neuroimaging. This paper reviews the methodological challenges involved in imaging children with NDDs. Specific topics include correcting for head motion, normalization using pediatric brain templates, accounting for psychotropic medication use, delineating complex developmental trajectories, and overcoming smaller sample sizes. The potential of neuroimaging-based biomarkers and the utility of implementing neuroimaging in a clinical setting are also discussed. Data-sharing approaches, technological advances, and an increase in the number of longitudinal, prospective studies are recommended as future directions. Significant advances have been made already, and future decades will continue to see innovative progress in neuroimaging research endeavors of NDDs.

1. Introduction

Human neuroimaging offers a unique and promising avenue to study neurodevelopmental disorders (NDDs), such as autism spectrum disorder (ASD), attention-deficit hyperactivity disorder (ADHD), and Tourette's syndrome (TS). Over the past decade, there has been a dramatic rise in the number of neuroimaging studies of NDDs. NDDs are characterized by abnormal development of the central nervous system; symptoms of these disorders emerge in early infancy/childhood and often persist throughout adulthood (Bishop and Rutter, 2008; Palframan, 1997). ASD is marked by persistent deficits in social communication and social interaction, as well as the presence of restricted, repetitive patterns of behavior (such as stereotyped motor movements, inflexibility, and sensory fixations) (American Psychiatric Association, 2013). Recent studies suggest a prevalence of approximately 1 in 68 children (Centers for Disease Control and Prevention, 2016). ADHD is characterized by a combination of inattentiveness, hyperactivity, or impulsivity (American Psychiatric Association, 2013) and has a prevalence of approximately 5–10% (Polanczyk and Rohde, 2007; Schall and Schwab-Stone, 2000). TS is marked by multiple motor tics and one or more vocal tics; tics are sudden and recurrent motor movements or vocalizations (American Psychiatric Association, 2013); TS has a prevalence of approximately 0.52% (Scharf et al., 2015). Each of these disorders is more common in males than females (American Psychiatric Association, 2013). NDDs can have detrimental effects on functioning, and it is essential to direct appropriate and targeted treatment as early in development as possible. Neuroimaging can be extremely useful for identifying core neurobiological mechanisms of NDDs, as it allows for real-time visualization of neural changes that are related to the emergence of clinical symptoms. While behavioral and cognitive studies of NDDs continue to offer important information on the symptom profile and manifestation of disorders in children, they, in isolation, are incomplete. Neuroimaging offers valuable information about the underlying neural structure and function, including the morphology of gray and white matter, strength and direction of white matter tracts, brain activation patterns, connectivity between regions and networks, as well as neurochemistry. Importantly, neuroimaging offers a unique opportunity to examine the clinical population in vivo; this allows for focused and targeted study of brain function directly in individuals who need intervention or treatment. This overcomes a limitation of animal studies that rely on indirect neurobiological models; oftentimes these models struggle to truly capture the extent and magnitude of the clinical manifestation of the disorder. Neuroimaging can provide valuable information about the neural origins of NDDs and can clarify complex and...
heterogeneous clinical profiles. Understanding the core neural mechanisms of NDDs can result in the development of novel treatments that are targeted to the pathology of the disorder. Particularly in the context of NDDs, neuroimaging may detect neural changes that precede the emergence of clinical symptoms; this would allow for the earliest implementation of treatment possible. Thus, neuroimaging offers unique advantages to further research and treatment of NDDs.

There are two main goals that are at the heart of neuroimaging research of NDDs. They are: i) Identify neurobiological markers that can lead to faster and improved diagnosis. This goal seeks to find a stable and reliable neural signature that predicts NDDs and improves/assists clinical diagnosis. This biomarker need not target the etiology of the disorder, but it must reliably predict and accurately classify children that have the disorder from children who do not (Yerys and Pennington, 2011). The need for reliable biomarkers stems from the limitations of using behaviorally-based diagnosis of NDDs: A more objective and quantifiable biomarker could limit missed diagnoses, enhance the pace of diagnosis, and predict children at high-risk that would benefit from early preventative treatment; ii) Identify aberrant neural mechanisms of the disorder that can be targeted by treatment. This goal involves assessing how brain-behavior relationships can be established in a targeted and controlled manner. Improved characterization of neural abnormalities may impact the design of targeted pharmacological and therapeutic interventions; additionally, neuroimaging has the potential to measure the effect of intervention on neural circuitry. This goal has translational implications and applies basic neurobiological research findings to treatment.

As neuroimaging studies increase in number, there has been increasing scrutiny and discussion of the methods of neuroimaging research (Bennett et al., 2009; Button et al., 2013; Carp, 2012; David et al., 2013; Eklund et al., 2016; Logothetis, 2008; Poldrack, 2008; Poldrack et al., 2017; Poldrack and Farah, 2015; Schleim and Roiser, 2009; Vul et al., 2009; Yarkoni, 2009) and of the clinical implications of neuroimaging findings (Bullmore, 2012; Linden, 2012; Matthews et al., 2006; Siegle, 2011). The results of neuroimaging studies vary widely, have limited generalizability, and are not specific to particular NDDs. This has affected the consistency, reliability, and sensitivity of neuroimaging-based inferences to specific NDDs. In part, this underscores the complex nature of these disorders at behavioral, cognitive, and biological levels, and researchers are realizing the enormity of the challenges associated with neuroimaging NDDs. This paper aims to review the methodological and conceptual challenges of MRI studies of children with NDDs, as well as the status of current approaches for addressing these problems. The concepts reviewed in this paper will focus primarily on the following neuroimaging techniques: structural MRI, functional MRI (fMRI), and diffusion tensor imaging (DTI). These techniques address neurobiological indices, such as anatomy, function, and white matter connectivity, and each provides unique information about the brain organization in children with NDDs. This paper will especially focus on examining the challenges of pediatric neuroimaging in the following neurodevelopmental populations: ASD, ADHD, and TS, as neuroimaging studies of these populations have increased dramatically over the last decade. Methodology will focus on child participants, as using neuroimaging to study early developmental windows in children can provide valuable insights into the neurobiology of these disorders. This paper discusses several issues that may hinder the replicability and generalizability in neuroimaging studies of NDDs, including: methodological variability and confounds associated with the neuroimaging paradigms used, widespread heterogeneity within the NDD populations, and the complex neurobiology of neurodevelopmental disorders. Advances in addressing these issues will also be presented.

2. Head motion in NDDs

2.1. Impact of head motion

Head motion in the scanner is generally increased in children, particularly children with NDDs; this presents a challenge to collecting high-quality and useable MRI data. The scanning environment can be anxiety-provoking for children and can make it increasingly difficult for them to remain still in the MRI scanner. Furthermore, children have more energy, move more in general, and are less aware of their own movements (Bookheimer, 2000). Previous studies have reported significantly greater motion in children compared to adults during neuroimaging scans (Poldrack et al., 2002; Thomas et al., 1999). In addition, younger age is strongly associated with head motion artifacts (Blumenthal et al., 2002). Further complicating data collection, NDD clinical populations are often associated with unique motion challenges. For example, children with ASD commonly have stereotypic behavioral patterns (Goldman et al., 2009), sensory sensitivities (Benson et al., 2009; Tharpe et al., 2006; Tomchek and Dunn, 2007), and anxiety disorders (Kim et al., 1999), all of which contribute to increased head motion within the scanner. Children with ADHD, characterized by hyperactivity and inattentiveness, also have greater difficulty remaining still within the scanner (Fair et al., 2012). Neuroimaging of children with TS is especially challenging because the tics, whether vocal or motor, can induce movement. This creates a motion artifact that is difficult to dissociate from the real brain activity occurring at the time of the tic generation (Ricksard, 2009). In this way, children with NDDs often present with increased motion, which can result in decreased data quality and unusable data in some cases.

The most worrisome aspect of head motion in NDDs is that it creates spurious effects in the data, even when the motion is relatively low enough for the data to be useable. For instance, greater motion artifact is associated with measures of neurobiological indices, such as smaller gray matter volumes (Blumenthal et al., 2002), biased estimates of cortical thickness and curvature (Alexander-Bloch et al., 2016), decreased long-distance functional connectivity, and increased short-distance connectivity (Power et al., 2012; van Dijk et al., 2012). Some researchers have questioned whether the connectivity hypotheses of ASD and ADHD merely represent an artifact or real effect (Deen and Pelphrey, 2012). Specifically, a pattern of decreased long-distance connectivity and increased local connectivity commonly reported in ASD and ADHD may be influenced by an artifact of in-scanner head motion (Belmonte et al., 2004; Konrad and Eckhoff, 2010; Power et al., 2012; Satterthwaite et al., 2012; van Dijk et al., 2012). Additionally, motion can act as a trait for a subject and be stable across trials within an experiment, confounding a participant’s activation patterns with the degree of motion (Friston et al., 1996; van Dijk et al., 2012). It has been estimated that these artifacts can account for as much as 90% of brain activation in extreme cases (Friston et al., 1996). If motion is not accurately accounted for, the differences between groups may actually represent differences in motion, rather than real differences in neural structure or function, ultimately leading to incorrect inferences from the data with findings difficult or impossible to replicate. Although head motion does not fully account for the variance observed between groups, these artifacts are not fully corrected by typical steps followed in data processing (Deen and Pelphrey, 2012; Fair et al., 2012; Satterthwaite et al., 2012; van Dijk et al., 2012). Further complicating the study of NDDs, not controlling for motion can lead to an overestimation of the relationship between age and within-network connectivity (Satterthwaite et al., 2012). Last, motion may also be correlated with symptoms of interest in NDDs, such as hyperactivity in ADHD. In these situations, it may be impossible to disentangle the real neural relationships from the effects of motion (Satterthwaite et al., 2012). Thus, controlling for head motion is absolutely critical in neuroimaging investigations, especially those involving clinical populations.
2.2. Reducing head motion

Pediatric neuroimaging studies nowadays commonly use several precautionary and preparatory procedures to curb head motion during data collection. For example, the use of mock scanners and behavioral paradigms to improve the quality of data collected can be particularly useful for populations with NDDs. Practice scans in a mock scanner allow children to adapt to the tight space of the scanner and learn to keep their head still. These simulated scans have been shown to decrease anxiety levels in children, increase compliance, and improve data quality (De Amorim E Silva et al., 2006; De Bie et al., 2010; Poldrack et al., 2002; Rosenberg et al., 1997). Well ahead of the scheduled scanning session, video demonstrations and audio recordings of the scanner noise can be sent home with families to allow the child to prepare for the scanning experience (Kana et al., 2011; Nordahl et al., 2016). The MRI scanner may also be decorated with toys, stickers, and colorful sheets to give it a child-friendly appearance (Kana et al., 2011). Behavioral rewards can be combined with the mock scanner paradigm to make children aware of the extent of their own movements and motivate them to remain still (Nordahl et al., 2016; Slifer, 1993). In this type of paradigm, children may be allowed to watch a video in the mock scanner, contingent on motion remaining below a certain threshold. Praise from the research assistant and prizes at the end of a successful session can also motivate the child to remain still within the scanner (Slifer, 1993; Slifer et al., 2002). There has been significant success in using these techniques with children with NDDs (Epstein et al., 2007; Nordahl et al., 2016, 2008). For example, before mock scanning practice, approximately 42% of data acquired from ADHD children would have been excluded due to motion greater than 2 mm (Epstein et al., 2007). However, during the actual scanning session (after mock training), only about 10% of data was unusable. Thus, there is potential of applying operant conditioning techniques to reduce head motion substantially and improve data quality in NDD populations.

A number of additional techniques may be used if participation is not required during the scan (e.g., resting-state fMRI, structural MRI, DTI). For example, the subject may be allowed to watch and listen to a movie throughout the scan using an MR-compatible audiovisual system and associated goggles and headphones (McGuirt, 2016). In children younger than 4 years, feeding and sleeping patterns can be manipulated so that the child falls asleep in the scanner before the session begins (Dean et al., 2014). Additionally, a number of commercially available products can be used to increase patient comfort and decrease head motion in the scanner, including cushions, sand bags, foam positioners, MRI table pads, vacuum immobilization bags, immobilization pads/kits, straps, ear buds, noise-cancelling headphones, and blankets, among other items. In general, judicious and planned use of motion reduction methods is perhaps the best way to improve data quality and yield.

2.3. Motion correction techniques

Over the past several years, a number of post-processing techniques have been developed to limit the impact of head motion on neuroimaging data. In fact, researchers have found themselves confronted with many approaches to address motion-related artifacts, as numerous methods have been developed, each associated with several different algorithms, parameters, and combination of order within a workflow. This has recently been addressed by several comprehensive papers summarizing different fMRI denoising techniques and providing recommendations for creating a pre-processing workflow (Caballero-Gaudes and Reynolds, 2016; Power et al., 2015, 2014; Satterthwaite et al., 2017). Briefly, in-scanner motion is typically measured by the realignment parameters obtained by rigidly aligning each successive volume to a reference volume: images are moved along three translational directions (x, y, z) and rotated along three angles (pitch, roll, yaw). The degree of motion can then be summarized by as the relative volume-to-volume head displacement, commonly termed framewise displacement (FD) (Power et al., 2012; Satterthwaite et al., 2012; van Dijk et al., 2012); while there are several methods for calculating FD, the measures are highly correlated and they target similar features of the data (Caballero-Gaudes and Reynolds, 2016; Power et al., 2015; Satterthwaite et al., 2017). The most common motion correction approach involves regressing motion indices as nuisance variables, especially the six realignment parameters, their temporal derivatives, and their corresponding quadratic terms (Caballero-Gaudes and Reynolds, 2016; Friston et al., 1996, 1995; Satterthwaite et al., 2017). This can significantly reduce motion artifacts, intra-subject variance, as well as inter-subject variance (Land et al., 2005). However, it is important to consider that increasing number of nuisance regressors are associated with a loss of degrees of freedom in data analysis. Temporal censoring, another technique commonly used to remove motion artifacts, involves removing those volumes from the data where significant motion has occurred, by excising the time points from the data (Power et al., 2012) or modeling the spikes as null regressors (Satterthwaite et al., 2013; Yan et al., 2013). Head motion during functional imaging causes a systematic, spurious decrease in long-distance connectivity coupled with an increase in short-distance connectivity (Power et al., 2012; van Dijk et al., 2012). Scrubbing has been found to successfully reverse these spurious correlations in functional connectivity analyses (Power et al., 2012). However, temporal censoring is associated with concerns, including a loss of degrees of freedom that may be more significant in the clinical group, the disruption of the temporal correlation of the dataset, and the introduction of synthetic data where interpolation techniques are used. It is also important to set thresholds for how much data can be censored (there is no established criterion for this) and to ensure that similar amount of censoring is present in both groups, although the groups may still differ in how the missing data is clustered temporally (Power et al., 2015). As the techniques described above rely on motion indices calculated from realignment parameters, these measurements only capture motion between slices and are not sensitive to within-volume (slicewise) motion (Satterthwaite et al., 2017). It is also common to regress the mean signal from tissues assumed to contain physiological noise, such as white matter and cerebrospinal fluid (CSF) (Power et al., 2014; Satterthwaite et al., 2013; Yan et al., 2013). Another method is to use principal component analysis to identify and remove components of white matter and CSF that are most likely to be noise or signal of non-interest (Behzadi et al., 2007). Researchers have also described the use of data-driven approaches, such as independent component analysis (ICA) that can separate noise from neuron-related signal. An additional technique involves using a filter to isolate low-frequency oscillations that are purportedly less susceptible to motion; however, there is also a risk of losing relevant network connectivity information, which spans across the frequency spectrum (Niazy et al., 2011). Finally, regression of the global signal has been shown to effectively reduce motion artifacts (Fox and Zhang, 2009; Power et al., 2014; Satterthwaite et al., 2013; Yan et al., 2013). Selection of an appropriate motion correction pipeline will inevitably depend on the study’s aims, characteristics of the data, MR acquisition parameters, as well as subject or group-specific traits that influence the type of noise observed; there is not a single workflow that will be appropriate or feasible for all studies (Caballero-Gaudes and Reynolds, 2016). Given that many studies of NDDs compare a clinical group to a control group across development, it is essential to select motion correction methods that minimize the residual correlations observed between functional connectivity and head motion (termed “QC-FC” correlation by Satterthwaite et al. (2017)). While no current analysis pipelines fully eliminate this relationship, Satterthwaite et al. (2017) reports that the combination of global signal regression and temporal censoring were most effective at minimizing the relationship.

While great progress has been made in correcting for motion with post-processing, studies continue to indicate that spurious effects remain to some degree (Kim et al., 2013; Power et al., 2012). Thus, there...
has been increased focus on prospective motion correction (PMC) techniques that can limit motion artifacts as the scan occurs rather than controlling for them in post-processing. PMC dynamically realigns the magnet’s field-of-view to match the subject’s head position, while subsequently updating the radiofrequency excitation and gradient waveforms (Zaitsev et al., 2006). The end result is that the magnetic field is kept relatively constant, even when head motion occurs (Maclaren et al., 2013); thus, PMC increases the signal-to-noise ratio and improves scan quality (Speck et al., 2006; Todd et al., 2015; Zaitsev et al., 2006). While not fully implemented in most research laboratories, PMC offers great potential for limiting the effects of motion and allowing for a broader range of children with NDDs to be studied.

3. Normalization and segmentation of pediatric brain images

One of the routine steps in neuroimaging data analysis is spatial normalization of participants’ brains (warping the subject’s brain to a template) to facilitate comparison across subjects and across groups. The two most commonly used brain templates in neuroimaging are the Talairach atlas (Talairach and Tournoux, 1988) and the Montreal Neurological Institute (MINI) (Evans et al., 1999) template, both of which are based on adult brains. However, normalizing pediatric brains to adult templates can be problematic (Hoeksma et al., 2005; Yoon et al., 2009). Specifically, dynamic changes in brain development make children’s brains very different compared to adult brains (Gogtay et al., 2004; Lenroot and Giedd, 2006; Sowell, 2004). For instance, differences in shape, size and composition of cortical matter can create distortions when normalizing pediatric brains to adult templates (Muzik et al., 2000; Wilke et al., 2003). In response to these issues, researchers have attempted to develop registration targets that more accurately represent the characteristics of a pediatric sample. Creating a study-specific brain template is one such strategy; generally, this involves averaging the images obtained from the participants of the study to create tissue maps that are based off the sample itself. However, this approach limits the generalizability across studies, as the template used for normalization is directly dependent on the data itself (Fillmore et al., 2015a,b). Additionally, it may not be feasible to create a study-specific template when the sample size is small, as the average template may not capture enough variance. Finally, for sample sizes spanning broader age ranges, it may be more fitting to create a template from a narrower age range (Fillmore et al., 2015a,b). Thus, creating a template using independent reference data may be optimal. Multiple studies have documented advantages associated with the use of age-appropriate pediatric brain templates. Normalization using pediatric templates entails fewer deformations (Yoon et al., 2009), improved classification of gray matter (Wilke et al., 2003), and reduced error (Fonov et al., 2011) compared to using adult templates. When selecting a template, it is important to consider the number of subjects used to make each age-specific template, as a small sample size may not capture the variability of the specified age group. The age range of the template itself should also closely correspond to the average age range and distribution of your sample. It is also useful to consider whether the template reflects the racial/ethnic makeup of your sample; for example, studies with Chinese participants have found that a template derived from Chinese brains leads to increased spatial precision (Jiao et al., 2009; Tang et al., 2010). Gender has been found to significantly influence the development of brain structure during childhood and adolescence (De Bellis et al., 2001; Good et al., 2001; Lenroot et al., 2007; Reiss et al., 1996; Wilke et al., 2008, 2007), so the gender distribution of the template should also be taken into account. Finally the techniques used to create the templates (e.g., linear versus non-linear registration, iterative versus non-iterative techniques) should be assessed. For example, the Neurodevelopmental MRI database provides age-specific brain templates in narrow age groups spanning from 2 weeks to 89 years of age (Richards et al., 2016; Sanchez et al., 2012, 2011). For this database, templates are provided in 1.5 month, 3-month, or 6-month increments from birth through 4 years of age, and 6-month increments for ages 4.5-19.9. The overall sample for the database (ages 2 weeks to 89 years) is the result of 2762 averaged images, with 9–72 images contributing to each of the age-specific child and adolescent templates. Additionally, the sample reflects the distribution of gender, income, and race/ethnicity according to the United States Census 2000 (Richards et al., 2016; Sanchez et al., 2012, 2011). Fonov et al. (2011) created age-specific templates in increments relevant to pubertal status, resulting in templates for the following age groups: 4.5–8.5 years (82 subjects), 7.0–11.0 years (112 subjects), 7.5–13.5 years (162 subjects), 10.0–14.0 years (105 subjects), 13.0–18.5 years (108 subjects). While these age bands are slightly larger (>4 years increments), they may be optimal depending on the sample and aims of a particular study. The Neurodevelopmental MRI database as well as Fonov et al. (2011) both created templates using data obtained primarily from the NIH MRI Study of Normal Brain Development (NINHPD), and both used nonlinear registration and iterative techniques. Iterative nonlinear registration techniques have the advantage of retaining significant anatomical detail, which can otherwise become blurred in areas of high variability when only linear registration techniques are used (Ashburner and Friston, 1999; Fonov et al., 2011). Wilke et al. (2008) used data from the NIHPD (404 healthy child subjects) to create a toolbox (Template-O-Matic) that allows the user to flexibly create a customized template that matches the age and gender distribution of one’s study. Rather than averaging the images, Template-O-Matic statistically models the effects of age and gender for each voxel in the brain, such that tissue maps can be generated automatically. Template-O-Matic only employed linear registration and opted against using an iterative approach, in which subsequent processing iterations are based on a previous average, to avoid biasing towards adult template priors (Wilke et al., 2008). Template-O-Matic atlas also overcomes the limitations associated with templates in which only a small number of images are available for a given age range. The statistical approaches implemented in Template-O-Matic have recently been updated in the release of CerebroMatic (Wilke, 2017; Wilke et al., 2017). Images were acquired from the NIHPD as well as the Cincinnati MR Imaging of Neurodevelopment Study (C-Mind) (1914 subjects, ages 13 months–75 years) and the effects of age, field strength, gender, and data quality were modeled. Rather than using a general linear model as Template-O-Matic employed, CerebroMatic utilized a more robust, non-parametric, multi-variate approach. Finally, fetal and neonatal brain MRI present with additional challenges, including increased motion artifact, lower signal-to-noise ratio due to the smaller size of the brain, inverted white/gray matter contrast in the predominantly unmyelinated brain, and variability in brain development (Devi et al., 2015; Makropoulos et al., 2017). Thus, several atlases and templates specific to infant brains have also been developed (Altaye et al., 2008; Fillmore et al., 2015a,b; Kazemi et al., 2007; Shi et al., 2011) along with a preprocessing package specific to infant brains (iBEAT) (Dai et al., 2013) and infant-specific tissue segmentation programs (for a recent review, see Makropoulos et al., 2017; Devi et al., 2015). These advances are essential for prospective studies of infants with a high risk of developing NDDs. Thus, the use of age-specific templates for improved precision is especially useful in the context of studying children with NDDs, which require imaging in early childhood and inherently introduce more variability into the data.

4. Psychotropic medication

Prevalence of child psychotropic medication use has increased dramatically in the past several decades (Olsson et al., 2002). For example, a survey in 2003 found that 56.3% of children with ADHD were currently taking medication (Centers for Disease Control and Prevention, 2005). A recent examination of commercially insured children with ASD found that 64% were prescribed psychotropic medications, and 50% of the sample was prescribed more than one
Psychotropic medication class (Spencer et al., 2013a); common medications include antidepressants, antipsychotics, antihypertensives, and stimulants (Aman et al., 2003). Clinical practice commonly includes the prescription of FDA-approved antipsychotic medications for TS, including haloperidol and pimozide, and alpha-2 agonists and atypical neuroleptics are also common (Kurlan, 2014). Consequently, it has become ever more difficult to recruit subject-pools that are medication-naïve, and there is debate within the neuroimaging research community on how to best account for psychotropic medication use.

Participants’ medication history and certain medication on the day of scan can impact the MRI signal and cause alterations in the data. Functional MRI is dependent upon regional cerebral blood flow to localize and quantify brain activity. Several psychiatric medications affect dopamine and serotonin, both of which regulate vasodilation in brain blood vessels and thereby affect cerebral blood flow (Krimer et al., 1998; Sharma et al., 1990). Oftentimes, clinical and control groups differ greatly in their medication use, such that differences in current medication could induce neural differences between groups. For medications with a shorter half-life, such as stimulant medications, it has become a common practice to ask subjects to withhold the medications for a specified time period before the scan (commonly 24–48 h before) (Cherkasova and Hechtman, 2009). This method maximizes sample size and attempts to account for direct effects of medication on the hemodynamic response at the time of the experiment. However, other psychotropic medications, such as anti-depressants, have a longer half-life, so brief withdrawal is not effective; additionally, brief withdrawal could introduce risky side effects or have negative consequences for the treatment of the disorder. It should also be noted that medication might sometimes be necessary to obtain subject compliance and to reduce head motion within the scanner. For children with high anxiety or extreme hyperactivity, withholding medication may make it extremely hard for the child to remain still within the scanner. Similarly, children with NDDs also commonly take psychotropic medications for comorbid disorders (e.g., TS is frequently comorbid with ADHD and OCD), and weaning off of medications could result in an exacerbation of the symptoms of comorbid disorders that are not targeted by the study (Gilbert and Buncher, 2005). Furthermore, the requirement to withhold medication may result in a biased sample, as parents of children with severe symptoms may not be willing to participate (Gilbert and Buncher, 2005; Greene et al., 2016). Thus, while withholding the medications for the day of the scan helps to control for direct effects of psychotropic medication on the BOLD signal, it may also introduce additional confounding factors.

Furthermore, there is evidence that refraining from psychotropic medication for a brief time period does not completely reverse its neurochemical effects. This is especially relevant for NDDs, as individuals often begin taking these drugs during critical periods of neural growth and maturation. Thus, the effects of these drugs may have an even greater long-term impact on children than adults. The effects of psychotropic medication on neuroimaging results have largely been discussed within the ADHD literature. Research indicates that dopaminergic systems can continue to be altered even after discontinuing medication (Moll et al., 2001). Pausing stimulant medication may also result in withdrawal responses, such as increased motor and anterior cingulate cortical activity in ADHD patients. In these cases, it may be difficult to differentiate withdrawal responses from the underlying pathology of ADHD without medication (Langleben et al., 2002). It is also important to consider the possibility that long-term use of psychotropic medications at an early age could have long-lasting effects on neural structure and function. For example, multiple animal studies have associated long-term, high dose stimulant administration to neurotoxicity of dopaminergic neurons in the striatum; however, it is argued whether or not these doses are comparable to human, clinical doses (Berman et al., 2009). Overall, there is evidence that ADHD medications attenuate ADHD neuropathology (Rubia et al., 2014; Shaw et al., 2009b; Spencer et al., 2013b). Few studies have directly examined the effect of psychotropic medication on brain functioning in ASD. A recent study found that current psychotropic medication use was associated with increased cortico-cortical connectivity in children and adolescents with ASD compared to ASD participants with no psychotropic medication history (Linke et al., 2017). One pilot study has examined the effect of propranolol, a beta-adrenergic antagonist, on functional connectivity and cognitive flexibility in ASD (Narayanan et al., 2010). Additionally, several PET studies have examined the direct effects of medication on neural receptor and transporter functioning in ASD (Fernell et al., 1997; Makkonen et al., 2011). DTI studies suggest that psychotropic medication usage does not affect white matter microstructure (Alexander et al., 2007; Lange et al., 2010; Lee et al., 2007). To our knowledge, no studies to date have explicitly examined the neural effects of psychotropic medication in individuals with TS. Additionally, numerous studies have evaluated the effects of psychotropic medications on brain function in other disorders, including bipolar disorder and schizophrenia (Abbott et al., 2013; Hafeman et al., 2012; Phillips et al., 2008), some of these medications which are used in NDDs. Importantly, a recent review which examined the neural effects of psychotropic medications in children and adolescents of a number of clinical disorders – including anorexia, ADHD, ASD, bipolar disorder, depression, OCD, and schizophrenia – found that overall, psychotropic medication appears to have either a null effect or normalizing effect on brain structure and function (Singh and Chang, 2012). While these normalizing effects are clinically promising, it could make it more difficult to detect clinical-control differences from a sample with a history of medication.

Several considerations and recommendations are provided for addressing the issue of psychotropic medication use in neuroimaging of NDDs. First, it is important to consider the purpose of the research study. Specifically, if the ultimate goal is to understand the disorder as a whole and to inform clinical treatment, then it is imperative to make the study sample as representative of the disorder as possible (Greene et al., 2016). Only recruiting medication-naïve participants is likely unrealistic and will not represent the clinical population; even asking participants to withhold medication will likely lead to a biased, healthier sample. There are also ethical considerations associated with temporarily stopping psychotropic medications, such as whether a hiatus will significantly worsen symptoms or negatively affect overall symptom management of the child. Furthermore, it is worth considering to what extent researchers can truly control for the effects of treatment on the brain in clinical populations. Not only is it often not feasible to recruit a medication-naïve sample, many participants with NDDs will likely have been exposed to alternative treatments, such as cognitive and behavior therapies, which also have well-documented effects on the brain (Barsaglini et al., 2014; Frewn et al., 2008; Linden, 2006). Thus, it is often not possible to recruit a medication-naïve sample, and attempting to do so would limit the generalizability and clinical translation of the study. In contrast, if the goal of the study is to target the mechanism of a specific symptom of a disorder (e.g., tics, hyperactivity, social deficits), and it is less relevant that the findings be generalizable to the disorder as a whole, then it may be suitable to restrict the sample to be medication-naïve (Greene et al., 2016). Of course, in these situations, it is necessary to explicitly define whether medication-naïve refers to individuals with no history of psychotropic medication, those that have not taken medication for a significant period, or those that stopped the medication immediately before the scan, as these definitions are not equivalent. In general, it is likely best to recruit large samples that include participants with a variety of medication statuses that are representative of the population. With a large sample, sub-groups can then be examined to detect contributing effects of psychotropic medication use. Important research questions remain regarding the short-term and long-term effects of medications in NDD populations. There remains a gap in the literature for longitudinal as well as placebo-controlled studies to distinguish whether neural differences are a result of medication use or a result of reduction in
found to have greater brain volume and hyper-connectivity at younger ages, which transitions to decreased volume and hypo-connectivity later in development (Courchesne et al., 2003, 2001; Uddin et al., 2013). Depending on the research question, researchers may wish to select an age band during which NDD-TD neurological differences are expected to be most prominent.

5. Developmental trajectories

Childhood and adolescence are dynamic periods for neural growth and development. Significant neurobiological changes occur during the time between childhood and adulthood: whole-brain volume grows exponentially by 25%, gray matter increases by 13%, and white matter increases a noteworthy 74% (Courchesne et al., 2000; Gaillard et al., 2001; Wilke et al., 2003). In addition, children’s skulls are thinner and less muscular, sinuses are not fully formed, and there is less free space containing cerebrospinal fluid (Gaillard et al., 2001). This results in non-negligible differences that exist across children even a few years apart. For example, while only two years apart, 2- and 4-year old children are in different stages of cognitive development in terms of their theory-of-mind and language skills; accordingly, there are significant differences in neural structure, organization, and connectivity. Thus, the age and developmental stages of participants included in neuroimaging studies will have a significant impact on the findings.

5.1. Complex neural changes in NDDs

Previous studies indicate that the developmental trajectories of NDDs vary from TD children. For example, there is brain overgrowth at young ages in ASD, followed by a slowing of growth in childhood (Courchesne et al., 2007). Functional connectivity may also have an alternate developmental trajectory in ASD: Uddin et al. (2013) have proposed that there is hyperconnectivity in ASD during late childhood (7–12 years), but hypocnectivity during adolescence (12–18 years) and adulthood. Thus, studies recruiting participants from different age groups may find different results of hypo- and hyperconnectivity in ASD, and arbitrary age cut-offs can result in dissimilar findings. In ADHD, the pattern of cortical maturation was found to be delayed such that prefrontal regions particularly did not reach peak thickness until significantly later (Shaw et al., 2007). In addition, a meta-analysis of ADHD indicated that hyperactivity of the somatomotor system in children is less pronounced than in adults (Cortese et al., 2012). There is evidence that children with TS have larger prefrontal cortices while adults with TS have smaller prefrontal cortices compared to TD individuals (Peterson et al., 2001). Another longitudinal TS study observed disparate changes over time between children with TS and TD children: While parallel and perpendicular diffusivity increased over time in TD children, these measures decreased over time in most children with TS (Debes et al., 2015). As a result, interactions between age, diagnostic category, and neurobiological measures may mask true group differences when samples encompass broader age ranges. Thus, it may be useful to constrain cross-sectional studies to developmentally relevant age bands. This will differ depending on the construct measured (e.g., executive function, theory of mind) and depending on when a given brain structure and function will mature and develop. Researchers must also consider prior literature on the alternate developmental trajectories in NDDs. For example, there is evidence that ADHD follows a similar pattern of cortical development as typical development, but this process is delayed such that peak brain volumes are reached at a later age (Shaw et al., 2007). In contrast, ASD has been

5.2. Developmental trajectories and symptom severity

In addition, there are likely different developmental trajectories between individuals with severe symptoms versus those whose symptoms remit over time. For example, while measures of white matter integrity increased over time in TD children, these measures decreased over time in most children with TS (Debes et al., 2015). This decrease was most pronounced in children with persisting tics compared to those with remission of tics (Debes et al., 2015). Also, a longitudinal study of children with ADHD found similar cortical thinning trajectories between children with ADHD and TD controls. However, among children with ADHD that had better outcomes, cortical thickness of the right parietal cortex normalized (Shaw et al., 2006). Thus, it is important to consider how patterns of developmental trajectories may differ among subsets of individuals with NDDs. Relatedly, many children with NDDs undergo general delays in development (e.g. delayed motor milestones, use of single words and phrases, adaptive skills) and it should be considered whether it is meaningful to compare them to typically developing children of the same age. In fact, it may be meaningful to also include comparisons to children of equivalent developmental stages to isolate the effects of particular NDD symptoms from global delays. Thus, it is important to consider how patterns of developmental trajectories may differ among subsets of individuals with NDDs.

5.3. Need for longitudinal studies

A critical step in the comprehensive understanding of NDDs is to investigate their developmental progression across the lifetime: this requires the use of longitudinal as well as prospective studies. Many “developmental” studies of NDDs are cross-sectional, which usually attempt to make longitudinal inferences. However, there are several conceptual errors associated with drawing longitudinal conclusions from cross-sectional data (Horga et al., 2014; Kraemer et al., 2000). Essentially, studies utilizing cross-sectional measures commonly report correlations between age and brain structure/function in a way that implies developmental trajectory (e.g. brain activation in the striatum decreases with increasing age). This is misleading, as different ages are likely to be associated with certain characteristics and methodological biases that preclude such conclusions. Specifically, if individuals are recruited from a clinic, older individuals recruited for the study likely have disorders and symptoms that are more persistent, whereas not all of the younger individuals recruited will continue to have such persistent symptoms later in life (Horga et al., 2014). It is likely that those that have persistent NDDs have different brain structure and function (Debes et al., 2015; Plessen et al., 2009), such that age relationships observed from cross-sectional data may be more reflective of the group differences between samples with differing outcomes. Even if subjects are recruited from the community rather than the clinic, it is still possible from symptoms and disorder severity to exert effects and biases on age correlations. For example, a 5-year-old child with ASD may need to be relatively high-functioning for their age group to adjust to the scanner environment, whereas an older child who previously had more severe ASD symptoms may be able to acclimate to the scanner relatively easily due to change in symptom expression over time and exposure to intervention. Thus, developmental trajectories cannot be accurately obtained from cross-sectional data, and longitudinal data are required to accurately ascertain these relationships. While an in-depth discussion of longitudinal research methodology is not in the scope of this paper, several papers have described the methods of acquiring and analyzing
longitudinal neuroimaging data in detail (King et al., 2017; Madhyastha et al., 2017; Mills and Tammes, 2014; Vijayakumar et al., 2017). Of note, accelerated longitudinal design is advantageous to single cohort design as it is less affected by factors such as participant dropout or changes in study design such as scanner upgrade (Vijayakumar et al., 2017). It is necessary to have greater than two time points to adequately assess within-person change, and the number of time points increases for quadratic, cubic, and higher-order trajectories (King et al., 2017). Mixed effects models are commonly used in the context of longitudinal studies as they can flexibly handle missing data as well as data collected at differing ages and time intervals (Vijayakumar et al., 2017).

5.4. Need for prospective studies

It is also important to combine longitudinal studies with prospective investigations of high-risk individuals. While group differences between clinical and control samples are often interpreted as etiological mechanisms or manifestations of behavior, it is very difficult to disentangle the neural differences that may have given rise to the development of the disorder from the neural differences that are compensatory and a result of living with the disorder. Although prospective studies still cannot elucidate cause, they can more clearly reveal the neural differences that distinguish high-risk individuals that do and do not develop the disorder. This has been particularly useful in the context of examining the broader autism phenotype among siblings of children with ASD. Family members of individuals diagnosed with ASD may express qualitative features associated with ASD, but to a milder and not clinically impairing level (Bolton et al., 1994; Piven et al., 1997). Prospective, longitudinal studies have indicated that autistic-like traits (e.g., reduced eye contact, repetitive vocalizations, delayed speech) consistent with the broader autism phenotype emerge by 12 months of age (Georgiades et al., 2013; Ozonoff et al., 2014). Prospective methods show particular promise for identifying predictive markers that can aid in diagnosis and intervention (Homberg et al., 2016).

Over the past ten years, numerous longitudinal studies of ASD and ADHD have been added to the literature (Tables 1 and 2). In particular, several large-scale studies have been conducted that yielded many important findings. For example, the Infant Brain Imaging Study Network is conducting a prospective, longitudinal study that follows siblings of children with ASD from 6 months of age; relevant behavior measures and several forms of neuroimaging are collected at multiple time-points until 2 years of age (Wolff et al., 2012). Another group that is part of the NIH Intramural Study has longitudinally followed children and adolescents with ADHD to characterize neural structure and function across development (Shaw et al., 2006). Of note, the majority of longitudinal neuroimaging studies have been conducted in ASD, with fewer studies in ADHD, and only two studies in TS. There has been one prospective study conducted in TS, while no prospective studies have been conducted in ADHD to date. Additionally, no papers have been published examining diffusion measures of ADHD in a longitudinal design; however, the Neuroimaging of the Children’s Attention Project (NICAP) group has recently released an ongoing study design incorporating DTI measures (Silk et al., 2016). Tables 1–3 outline the notable increase in longitudinal studies that have occurred for ASD and ADHD over the past decade, as well as the remaining gaps in longitudinal and prospective neuroimaging studies of NDDs.

6. Heterogeneity within NDDs

Heterogeneity is characteristic of NDD clinical presentation. For example, ASD is diagnosed and defined by deficits in social communication and restrictive, repetitive patterns of behavior. These two broad categories of symptoms are quite different from each other and each contains a variety of diverse symptoms. There is great variation in the developmental trajectory of language acquisition (e.g., no functional speech to impairment in discourse processing and the social use of language), presence of echolalia, pronoun reversal, reading comprehension, and speech pragmatics in ASD (Rapin and Dunn, 2003; Tager-Flusberg et al., 2005). There is also extreme variation among sensory sensitivities and stereotypies (Bodfish et al., 2000; Leekam et al., 2007). Among children with ADHD, there are inattentive, hyperactive-impulsive, and combined subtypes with varying profiles of severity and weaknesses (Faraone et al., 1998). Similarly, children with TS may present with simple or complex tics along a continuum of severity (Grados and Mathews, 2009). As a result of this heterogeneous manifestation, children with the same diagnosis can have vast differences in ability, level of functioning, profile of symptoms, and developmental trajectory. This presents several challenges for neuroimaging studies, which often attempt to recruit homogenous samples, free of confounding factors, but with findings that are generalizable to the clinical population as a whole.

6.1. NDD subtypes

Interestingly, research suggests that each NDD diagnostic category actually consists of many subtypes. A large number of genes are implicated in these disorders, which do not appear to be converging on a single pathway per disorder (Hu et al., 2014). For example, over 103 genes have been implicated in ASD (Betancur, 2011), and researchers have previously suggested that ASD can no longer be conceptualized as one disorder with 2–3 major domains; rather, it is more likely that ASD reflects a combination of co-occurring disorders, each with potentially independent mechanisms (Happe et al., 2006). Additionally, within the ASD diagnostic category itself, individuals may be further divided into individuals with intellectual disability, those that are minimally verbal, and those with regressive trajectories. Within the ADHD population, multiple subtypes have been identified, including primarily inattentive type and hyperactive type (Carlson and Mann, 2002). While a portion of individuals with ADHD exhibit deficits in executive functioning, this is not pervasive for all of those with the diagnosis (Nigg et al., 2005). Similarly, TS consists of a variety of types and severity of tics (e.g., simple tics, complex tics, palipraxia, echopraxia, coprolalia, copo-praxia) potentially occurring through different neural mechanisms (Eapen and Robertson, 2015). Variations in clinical profile and genetic risks in these disorders may be subsequently associated with neuroanatomical and functional differences. This could potentially explain the variability in neuroimaging findings across research sites, whose clinical samples are very different. Traditional clinical-control group analyses may also obscure the distinct neural features of subgroups within the disorder, and identifying these subgroups can lead to further understanding of the disorder and reveal targets for intervention.

Exploring subgroups within the NDD diagnostic categories may best be done with appropriate choice of analysis. Multivariate approaches can best serve this purpose and have been increasingly employed by researchers in the field. For example, several researchers have used a partial clustering methodology, Topological Data Analysis (TDA), to identify subgroups within neuroimaging data. TDA is an unsupervised machine learning approach; it is a data-driven multivariate pattern analysis approach that does not depend on an a priori hypothesis. For example, Kyeong et al. (2017) used TDA to reveal two unique ADHD subgroups, labeled as mild symptom ADHD and severe symptom ADHD. Another study used TDA to identify subgroups within a young male Fragile X cohort (Bruno et al., 2017). Two clinically and neurally distinct subgroups were identified; one subgroup exhibited increased brain volume across widespread areas of the brain and also exhibited a more severe cognitive profile and ASD symptoms. An additional approach is the use of clustering methods such as community detection, which falls under the overarching branch of graph theory. For example, Gates et al. (2014) employed unified structural equation modeling (uSEM) combined with community detection to identify five subgroups within an ADHD and TD sample. Another study used community detection
### Table 1
Longitudinal neuroimaging studies of ASD (2005–present).

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Ages</th>
<th>Interscan Interval</th>
<th>Type</th>
<th>Main Neuroimaging Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardan et al. (2009)</td>
<td>18 ASD, 16 TD</td>
<td>8–12 years baseline</td>
<td>2 time-points; 30-month interval</td>
<td>Longitudinal</td>
<td>Brain volume; cortical thickness</td>
</tr>
<tr>
<td>Mosconi et al. (2009)</td>
<td>50 ASD, 11 TD</td>
<td>18–35 months baseline</td>
<td>2 time-points; 4 year interval</td>
<td>Longitudinal</td>
<td>Amygdala volume</td>
</tr>
<tr>
<td>Schumann et al. (2010)</td>
<td>41 ASD, 44 TD</td>
<td>12–48 months baseline</td>
<td>6–12 month intervals until exited study at 48–60 months</td>
<td>Prospective, longitudinal</td>
<td>Cerebral gray and white matter volumes</td>
</tr>
<tr>
<td>Hazlett et al. (2011)</td>
<td>51 ASD, 38 TD</td>
<td>2 years</td>
<td>1 time-points; 2 year interval</td>
<td>Longitudinal</td>
<td>Proton magnetic resonance spectroscopic imaging</td>
</tr>
<tr>
<td>Mosconi et al. (2009)</td>
<td>50 ASD, 11 TD</td>
<td>18–35 months baseline</td>
<td>2 time-points; 4 year interval</td>
<td>Longitudinal</td>
<td>Amygdala volume</td>
</tr>
<tr>
<td>Schumann et al. (2010)</td>
<td>41 ASD, 44 TD</td>
<td>12–48 months baseline</td>
<td>6–12 month intervals until exited study at 48–60 months</td>
<td>Prospective, longitudinal</td>
<td>Cerebral gray and white matter volumes</td>
</tr>
<tr>
<td>Hazlett et al. (2011)</td>
<td>51 ASD, 38 TD</td>
<td>2 years</td>
<td>1 time-points; 2 year interval</td>
<td>Longitudinal</td>
<td>Proton magnetic resonance spectroscopic imaging</td>
</tr>
<tr>
<td>Corrigan et al. (2012)</td>
<td>45 ASD, 14 TD</td>
<td>3–4 years baseline</td>
<td>3 time-points; 3 year interval</td>
<td>Longitudinal</td>
<td>Tensor-based morphometry: regional tissue growth rates</td>
</tr>
<tr>
<td>Hazlett et al. (2011)</td>
<td>51 ASD, 38 TD</td>
<td>2 years</td>
<td>1 time-points; 2 year interval</td>
<td>Longitudinal</td>
<td>Brainstem volumes</td>
</tr>
<tr>
<td>Nordahl et al. (2012)</td>
<td>85 ASD, 47 TD</td>
<td>2–4 years baseline</td>
<td>2 time-points; 1 year interval</td>
<td>Longitudinal</td>
<td>Amygdala volumes, total cerebral volumes</td>
</tr>
<tr>
<td>Wolff et al. (2012)</td>
<td>92 high-risk ASD infants</td>
<td>6 months baseline</td>
<td>Imaged at 6 months and 24 months</td>
<td>Longitudinal</td>
<td>DTI: fractional anisotropy of white matter pathways</td>
</tr>
<tr>
<td>Corrigan et al. (2013)</td>
<td>45 ASD, 14 TD</td>
<td>3–4 years baseline</td>
<td>3 time-points; 3 year interval</td>
<td>Longitudinal</td>
<td>Proton magnetic resonance spectroscopic imaging</td>
</tr>
<tr>
<td>Hua et al. (2013)</td>
<td>13 ASD, 7 TD</td>
<td>6–17 years baseline</td>
<td>2 time-points; 2.9 year interval</td>
<td>Longitudinal</td>
<td>Total cerebral volume; extra-axial fluid; lateral ventricle</td>
</tr>
<tr>
<td>Jou et al. (2013)</td>
<td>23 ASD, 23 TD</td>
<td>7–17 years baseline</td>
<td>2 time-points; 2 years</td>
<td>Longitudinal</td>
<td>Amygdala and hippocampus volumes</td>
</tr>
<tr>
<td>Prigge et al. (2013)</td>
<td>40 ASD, 17 TD</td>
<td>3–12 years baseline</td>
<td>Up to 3 time-points; 2.5 year interval</td>
<td>Longitudinal</td>
<td>Volume of Hesch's gyrus</td>
</tr>
<tr>
<td>Shen et al. (2013)</td>
<td>33 high-risk ASD, 22 low-risk ASD</td>
<td>6–9 months</td>
<td>3 time-points; 6–7 month interval</td>
<td>Longitudinal</td>
<td>Total cerebral volume; extra-axial fluid; lateral ventricle</td>
</tr>
<tr>
<td>Barnes-Goraly et al. (2014)</td>
<td>23 ASD, 23 TD</td>
<td>8–12 years baseline</td>
<td>2 time-points; 2 year interval</td>
<td>Longitudinal</td>
<td>Amygdala and hippocampus volumes</td>
</tr>
<tr>
<td>Langen et al. (2014)</td>
<td>49 ASD, 37 TD</td>
<td>9 years baseline</td>
<td>2 time-points; 2.4 year interval</td>
<td>Longitudinal</td>
<td>Volume of striatum</td>
</tr>
<tr>
<td>Traven et al. (2014)</td>
<td>100 ASD, 56 TD</td>
<td>3–41 years across study</td>
<td>1–4 time-points; 2.6 year interval</td>
<td>Accelerated longitudinal design</td>
<td>DTI measures of whole brain white matter</td>
</tr>
<tr>
<td>Zielinski et al. (2014)</td>
<td>97 ASD, 60 TD</td>
<td>3–39 years baseline</td>
<td>3 time-points; 2.7 year interval</td>
<td>Longitudinal</td>
<td>Cortical thickness</td>
</tr>
<tr>
<td>Lange et al. (2015)</td>
<td>100 ASD, 56 TD</td>
<td>3–35 years baseline</td>
<td>3 time-points; 2.7 year interval</td>
<td>Longitudinal</td>
<td>Whole-brain, regional gray matter and white matter volumes</td>
</tr>
<tr>
<td>Murdaugh et al. (2015)</td>
<td>31 ASD, 22 TD</td>
<td>8–13 years</td>
<td>Before and after reading intervention (10 weeks apart)</td>
<td>Longitudinal, reading intervention</td>
<td>Resting-state functional connectivity of language regions</td>
</tr>
<tr>
<td>Nordahl et al. (2015)</td>
<td>139 ASD, 82 TD</td>
<td>3 years baseline</td>
<td>3 timepoints; 1–2 year interval</td>
<td>Longitudinal</td>
<td>Corpus callous organization and diffusion characteristics</td>
</tr>
<tr>
<td>Traven et al. (2015)</td>
<td>100 ASD, 56 TD</td>
<td>3–41 years across study</td>
<td>1–4 time-points; 2.6 year interval</td>
<td>Accelerated longitudinal design</td>
<td>DTI, microstructure of corpus callous</td>
</tr>
<tr>
<td>Wolff et al. (2015)</td>
<td>270 high-risk ASD, 108 low-risk ASD</td>
<td>6 months baseline</td>
<td>6, 12, 24 months of age</td>
<td>Longitudinal, prospective</td>
<td>Cortical thickness and surface area</td>
</tr>
<tr>
<td>Libero et al. (2016)</td>
<td>129 ASD, 49 TD</td>
<td>3–5 years baseline</td>
<td>3–5 years of age</td>
<td>Longitudinal</td>
<td>Total area, length, and thickness of corpus callous</td>
</tr>
<tr>
<td>Mensen et al., 2016</td>
<td>90 ASD, 90 TD</td>
<td>9–20 years baseline</td>
<td>1–3 time-points; variable interval</td>
<td>Accelerated longitudinal design</td>
<td>Head circumference; structural brain volumes</td>
</tr>
<tr>
<td>Murdaugh et al. (2016)</td>
<td>26 ASD, 19 TD</td>
<td>8–13 years</td>
<td>Before and after reading intervention (10 weeks apart)</td>
<td>Longitudinal, reading intervention</td>
<td>Functional activation of language areas</td>
</tr>
<tr>
<td>Qiu et al. (2016)</td>
<td>36 ASD, 18 TD</td>
<td>2–3 years baseline</td>
<td>2 time-points, 2 year interval</td>
<td>Longitudinal</td>
<td>Caudate nucleus volume</td>
</tr>
<tr>
<td>Hazlett et al. (2017)</td>
<td>106 high-risk ASD, 42 low-risk ASD</td>
<td>6 months baseline</td>
<td>6, 12, 24 months of age</td>
<td>Longitudinal, prospective</td>
<td>Cortical volume and surface measurements</td>
</tr>
<tr>
<td>Loth et al. (2017)</td>
<td>437 ASD, 300 TD</td>
<td>ASD age 12–30</td>
<td>6, 12, 24 months of age</td>
<td>Longitudinal</td>
<td>Structural, diffusion, resting-state, task fMRI</td>
</tr>
<tr>
<td>Shen et al. (2017)</td>
<td>221 high-risk ASD, 122 low-risk ASD</td>
<td>6 months baseline</td>
<td>6, 12, 24 months of age</td>
<td>Longitudinal, prospective</td>
<td>DTI measures of cortical, cerebellar, and striatal white matter pathways</td>
</tr>
</tbody>
</table>

---

* Publication part of Infant Brain Imaging Study (IBIS).

b Study methods are published ahead of data analyses.
approaches to identify three novel subgroups of ADHD with distinct neural features and temperament differences (Karalunas et al., 2014). Community detection has also been used to identify ADHD subgroups with distinct reward system network connectivity with related impulsive behavior characteristics (Costa Dias et al., 2015). Thus, a useful approach to account for the heterogeneity within NDD diagnostic categories is to apply data-driven techniques to large, heterogeneous samples to identify subgroups that differ in neural structure and also have relation to clinical presentation.

6.2. Comorbidities

Children with NDDs often present with comorbid pathologies and symptoms that span multiple diagnoses. For instance, high rates of anxiety disorders, phobias, and ADHD have been observed in ASD (Salazar et al., 2015). There is evidence that ADHD is six times as common in children with ASD (Taurines et al., 2012), and autism traits are significantly more common among individuals with ADHD (Mulligan et al., 2009). In addition, ADHD is one of the most common comorbid conditions with TS (Robertson, 2000), and presents in as many as 21–90% of TS cases (Robertson and Eapen, 1992). TS also commonly occurs with obsessive-compulsive disorders, self-injurious behaviors, and anxiety (Robertson, 2000). Thus, it is usually difficult to find children with ASD, ADHD, or TS who have no comorbid conditions. Additionally, not only is it incredibly difficult to collect a “pure sample”, doing so could limit the validity and generalizability of the study. If the ultimate aim of the study is to improve characterization and treatment of the clinical population, then the sample should reflect the heterogeneity observed in the overall population. Additionally, comorbidity tends to be more typical than atypical in the clinical population, and excluding individuals with comorbidities will likely result in a sample that is not representative of the overall clinical population seeking treatment. Furthermore, while comorbidities are often referred to as “confounds”, researchers have questioned whether this is an accurate characterization (Greene et al., 2016). Specifically, these comorbidities cannot simply be “controlled for” with analyses such as regression and covariance, as comorbidities interact in complex ways across development, and those interactions are often of clinical and research interest. Attempting to remove these complex interactions could in fact remove core characteristics of the clinical disorders. Thus, it is often appropriate and best to recruit large samples that include individuals with and without comorbidities, as well as a wide range of the types of comorbidities. This will best capture the heterogeneity within the sample and allow the contribution of comorbidities to be investigated through the analyses.

7. Sample composition

7.1. Sample size and statistical power

Neuroimaging studies have historically been hindered by smaller sample sizes, perhaps due to the difficulty in acquiring high quality data from participants and due to higher costs of MRI scanning. Smaller sample size in neuroimaging studies is an issue that has been consistently raised in recent years: In addition to the failure to detect true effects, low statistical power can exaggerate measured effect sizes and enhance false report probability (Berger and Sellke, 1987; Button et al., 2013; Ioannidis, 2011; Pollard and Richardson, 1987). In fact, a recent study found that the median power across 461 studies of brain volume was only 8% (Button et al., 2013), indicating that the probability of detecting a true effect was exceptionally low and the effects detected were likely spurious. Of note, recent meta-analyses of ASD, ADHD, and TS indicate relatively smaller sample sizes. For example, a meta-analysis of fMRI studies of language in ASD indicated sample sizes ranging from 10 to 26 (Herringshaw et al., 2016). A meta-analysis of 55 task-based fMRI studies in ADHD included studies with sample sizes ranging...
from 7 to 30 (Cortese et al., 2012). The results obtained from these small sample studies do not always remain when tested in larger samples. For example, one relatively consistent finding within the ASD literature is abnormal volume and structure of the cerebellum (Carper and Courchesne, 2006; Courchesne et al., 2003, 1994, 1988; Fatemi et al., 2012, 2002). However, a recent study investigated this finding using over 600 subjects from an ASD neuroimaging database; power was estimated to be 85% to detect an effect (Traut et al., 2017). Unfortunately, this analysis did not observe any significant relationship between ASD diagnosis and cerebellar volume; in fact, age, sex, and IQ contributed greatly to cerebellar volume variability independent of ASD diagnosis. The researchers speculate that this discrepancy in the literature is likely due to heterogeneity within the ASD population, small sample size, and publication bias. In recent years, there has been a concerted effort in neuroimaging research to undertake power calculations beforehand in order to determine adequate sample size. Several convenient neuroimaging power calculation tools for brain activation studies have been developed, such as fmripower (Mumford, 2012), PowerMap (Joyce and Hayasaka, 2012), and neuropower (Durnez et al., 2016). Nevertheless, it should be noted that power calculation tools for other analyses, especially some types of functional connectivity, are still not available. Overall, smaller sample sizes in neuroimaging studies greatly hinder researchers’ ability to find true and consistent findings for NDDs; it is important to undertake power calculations when possible and invest in larger sample sizes.

### 7.2. High versus low-functioning participants

Neuroimaging study samples are also inherently limited to high-functioning participants that may not represent the broader clinical population. As the MRI scanning environment can be challenging, only NDD children that can adjust to the scanning environment, follow instructions, and remain extremely still can participate in neuroimaging studies. Thus, including relatively high-functioning children makes the sample less heterogeneous, which may not be representative of the larger NDD population. For example, symptoms such as severe sensory sensitivities, high levels of anxiety, hyperactivity, and full-body tics could preclude children from successfully completing the scan. While scanning only children that are high-functioning is obviously preferred to not scanning any children with NDDs at all, it greatly hinders the generalizability of findings. First, it is possible that children that are low-functioning represent a different subtype of the disorder with distinct neural etiologies (Happé et al., 2006). As a result, the neural signatures of high-functioning children with NDDs may not be applicable to the broader clinical sample. Second, the ability to detect group differences would be greatly increased by studying children with the most severe symptoms, while group differences may only be subtle when recruiting high-functioning clinical populations.

Therefore, it is important to continue developing improved and faster scanning methods to allow for the inclusion of participants that present with severe symptoms. This involves application of technological advancements to control for motion artifact and spurious effects, as well as improved behavior techniques to reduce the occurrence of motion. Previous studies indicate promise in the use of behavior techniques to obtain compliance in children with NDDs (Epstein et al., 2007; Kana et al., 2011; Nordahl et al., 2016, 2008), and PMC offers an exciting new development in this direction (Maclaren et al., 2013). Additionally, in clinical studies for which participant sedation is possible, several forms of data can be acquired which do not require participation, including resting-state fMRI, structural MRI, and DTI. Several studies outline successful sedation protocols prior to MRI in the ASD population (Ahmed et al., 2014; Lubisch et al., 2009) and ADHD (Kitt et al., 2015). Thus, improved methods to obtain compliance and reduce head motion can allow for the inclusion of a broader range of participants with NDDs.

Recently, several research groups have described successful paradigms for acquiring neuroimaging data from children with ASD that are low-functioning. One study used behaviorally-based approaches to obtain higher quality structural and DTI scans from children with ASD that also had intellectual disability (Mean IQ = 54.1, SD = 12.1) (Nordahl et al., 2016). First, an in-depth parent interview was conducted to evaluate the child’s capacity to follow scanner procedures, as well as to assess potential motivating rewards to use during the scan. A video model of the scanning procedure was provided to the child and parents to practice at home. During the mock scanning session, a behaviorally trained research assistant broke down the entire scanning procedure into simple steps, such as entering the scanning room and approaching the scanner. During this session, a preferred video reinforcer was played contingent on low movement thresholds. Of the thirteen low-IQ subjects with ASD imaged, 100% of T1-weighted images met quality assurance standards for head motion artifacts within five attempts. For the DTI scans, 71% of these subjects had 0 volumes with slice dropout; the remaining participants had slice dropout on one to four volumes. This study indicates that while low-functioning children with ASD present unique challenges, appropriate training with careful planning and the use of structured behavioral procedures can create a scanning environment that is conducive, predictable and amenable to these subjects.

### 7.3. Data-sharing approaches

In order to recruit sample sizes that are adequately powered and diverse, multisite studies and data-sharing approaches are encouraged. Investing in larger studies that have adequate power may help definitively answer research questions and reflect the heterogeneity of the population. Recruiting participants across multiple sites will have the added effect of gaining access to samples from greater number of clinics and centers; this may more accurately represent the population. Researchers should also take advantage of, and engage in, data-sharing platforms to efficiently use previously collected data and encourage replication. Data-sharing has been called for by many neuroimaging researchers (Milham, 2012; Poldrack, 2012; Van Horn and Gazzaniga, 2002) as it allows investigations to achieve impressively large sample sizes from previously collected data. The National Institutes of Health (NIH) has also focused on data-sharing needs, recently in the context of its Big Data to Knowledge (BD2K) Initiative (Bourne et al., 2015; Margolis et al., 2014). Additionally, the National Institute of Mental Health Data Archive (NDA) provides an infrastructure for aggregating and sharing research data; it makes data available from several data-sharing platforms spanning diverse scientific domains. For example, one of these platforms is the Research Domain Criteria Database (RDoC-DB), which shares clinical, EEG, eye-tracking, and MRI data in line with

### Table 3

Longitudinal neuroimaging studies of TS (2005–present).

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Ages</th>
<th>Interscan Interval</th>
<th>Type</th>
<th>Main Neuroimaging Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bloch et al. (2005)</td>
<td>43 TS</td>
<td>8.5-13.9 years at baseline</td>
<td>2 time-points, 7.5 year interval</td>
<td>Prospective; longitudinal</td>
<td>Caudate nucleus volume</td>
</tr>
<tr>
<td>Debes et al. (2015)</td>
<td>22 TS, 21 TD</td>
<td>11.4-19.3 years at baseline</td>
<td>2 time-points, 4.4 year interval</td>
<td>Longitudinal; comparison between persisting/ remitting</td>
<td>Voxel-based morphometry; DTI</td>
</tr>
</tbody>
</table>

**Notes:**

- TS = Tourette Syndrome
- DB = Database
- RDoC = Research Domain Criteria
- Caudate nucleus volume
- Voxel-based morphometry
- DTI = Diffusion Tensor Imaging
the overarching goals of RDoC. Several data-sharing platforms are particularly relevant for the study of NDDs: The Autism Brain Imaging Data Exchange (ABIDE-I (Di Martino et al., 2014) is a dataset of 1112 structural MRI and resting-state functional MRI images of participants with autism and healthy controls, ages 7–64 years, from 17 international sites. ABIDE-II has recently been released, and contains 1044 additional datasets; in addition, a small subset of these individuals has been scanned at two time points. In addition, a number of relevant behavioral measures are included in ABIDE-II, such as scores from the Autism Diagnostic Observation Schedule, 2nd edition (ADOS-2), Autism Diagnostic Interview-Revised (ADI-R), as well as measures of IQ, social functioning, adaptive behavior, and executive functioning. ABIDE-II also indicates the medication status of some participants, specifically whether they were taking certain medications within 3 months of the scan. Another data-sharing platform specific to ASD is the National Database for Autism Research (NDAR), which includes data from multiple modalities (i.e., MRI, genetics, behavioral, cognitive) acquired from individuals with ASD as well as typically developing individuals (Hall et al., 2012). The ADHD-200 database (The ADHD-200 Consortium, 2012) includes 776 anatomical and resting-state fMRI images of typically developing individuals (491) and individuals with ADHD (285), ages 7–21 years. This database also includes information on diagnostic status, ADHD symptom measures, Full-scale IQ, and lifetime medication status. Another relevant database for understanding typical neurodevelopment is available from the NIH Study of Normal Brain Development (Evans, 2006), which contains longitudinal images and relevant demographic data from 554 psychiatrically normal children, adolescents, and young adults. Together, these specific databases offer a repertoire of neuroimaging and neuropsychological data to understand the cognitive and neurobiological mechanisms of NDDs using larger and diverse samples.

While data-sharing approaches offer viable solutions for expanding sample size and increasing power, combining neuroimaging data from multiple sites needs several methodological considerations. Specifically, sites differ in scanner manufacturer/type, field strength, pulse sequence, signal-to-noise ratios, computer software updates, and general scanning protocols. There has been increased discussion on the potential differences introduced by multi-site studies, and findings are mixed regarding scanner effects on MRI studies. For example, there is evidence that inter-scanner differences exist even when using similar scanners and protocols (Focke et al., 2011; Takao et al., 2013), and differences have been found to result from systematic effects of head coils (Focke et al., 2011), site-specific non-linearities in the imaging gradients (Jovicich et al., 2006), scanner upgrades (Takao et al., 2013), and having different ratios of cases to controls between scanners (Takao et al., 2014). On the other hand, studies have also suggested that scanner differences are substantially less than the effect of interest (Stonnington et al., 2008), or are not always problematic, such as when there is similar ratio of cases to controls between sites (Pardoe et al., 2008; Takao et al., 2014). Previous studies have also found some success in accounting for scanner effects by including site as a covariate in statistical analyses (Pardoe et al., 2008; Takao et al., 2014), modeling scanning site as a random effect in a mixed-effects model (Fennema-Notestine et al., 2007), and using independent component analysis to identify components that are associated with scanning parameters (Chen et al., 2014). While most research has focused on volumetric measures, a recent study investigating multisite reliability of resting-state functional connectivity found limited impact of scanner effects (Noble et al., 2017). Importantly, little research on multisite methods has been conducted with child or developmental populations; almost all studies have included adult samples. This is an important gap in the literature, given that developmental effects (e.g., age) may interact with scanner effects and present an additional confounding factor. One recent study used data acquired from the ABIDE database to examine the influence of scanner factors on cortical thickness measurements in children and adults with ASD (Auzias et al., 2016). Notably, there was a significant effect of scanner in most cortical regions of the brain, and scanner effects frequently exceeded diagnostic effects. There were particularly large effects in the frontal cortex, which is found to be, although not reliably, a site of abnormality in ASD. For most brain regions, the decrease of cortical thickness with age was reliable across sites. However, for the insula, there was a significant interaction between age and scanner factors. The findings of this study suggest significant scanner-specific effects on findings of NDD studies, which may contribute to inconsistent findings as well as pose challenges for multi-site studies. Further research is needed to replicate these findings and explore effects in other imaging modalities (e.g., VBM, rs-fMRI) and additional NDDs. For researchers utilizing multi-site neuroimaging data, several methods have recently shown promise for accounting for scanner effects, including: source based morphometry (SBM) (Chen et al., 2014) and the ComBat algorithm (Fortin et al., 2018). SBM uses independent component analysis to identify components of the data that are significantly associated with scanning parameters; it is a flexible data-driven approach that has been shown to effectively detect and remove scanner effects from multi-site structural MRI studies (Chen et al., 2014). ComBat adopts an approach previously applied in genomics to correct for batch effects; imaging features are modeled as a linear combination of biological variables and site effects, whose error term is influenced by additional site-specific scaling factors (Fortin et al., 2018). ComBat has been successfully applied to DTI data (Fortin et al., 2017) and cortical thickness measurements (Fortin et al., 2018).

Neuroimaging data-sharing platforms offer great promise for investigating the neurobiology of NDDs and will continue to evolve with the changing research landscape. For example, these databases can be improved by adding imaging data acquired from low-functioning clinical populations, specifying more behavioral measures, including measures that are on a trait-based scale (research domain criteria, RDoC, more described below), and including more detailed medication history and comorbidity status. Behavioral measures should also be expanded to not just include measures specific to that diagnosis, but also measures that extend across diagnostic categories (such as trait levels of attention, executive control, anxiety; or an RDoC scale). While ABIDE-II (Di Martino et al., 2014) has begun to incorporate many of these features, a database with ADHD samples has not been updated, nor is there a database publicly available with data collected from individuals with TS. In the future, data-sharing platforms may prove to be especially useful for NDDs with extremely low incidence (e.g. genetic disorders such as Rett’s Syndrome or Pitt Hopkins) for which recruitment and obtaining large samples are particularly difficult. Last, databases should also aim to include multimodal information, including genetic phenotypic information so that the relationship between genes and neurobiology can be better understood in these populations (Belmonte et al., 2008).

8. Neurobiological markers

The quest for identifying stable neural biomarkers has become a focus area in the field of neuroimaging of NDDs. Broadly, NIH has defined a diagnostic biomarker as “a biomarker used to detect or confirm presence of a disease or condition of interest or to identify individuals with a subtype of the disease” (Group, 2016). However, care should be taken when using and defining the term “biomarker” in the context of neuroimaging research. According to Yerys and Pennington (2011), a valid biomarker should have three characteristics: 1) it should be present prior to the onset of symptoms, 2) be specific to the disorder, and 3) define the disorder independently from the behavioral symptoms used to make the diagnosis. McPartland (2016) further delineates distinct types of biomarkers, including diagnostic biomarkers (provide objective identification of the diagnosis), screening biomarkers (predict risk of a diagnosis prior to the emergence of behavioral symptoms), stratification biomarkers (divide a clinical population into treatment-relevant subgroups), target engagement biomarkers (indicate that treatment is...
targeting identified process), and early efficacy biomarkers (quickly indicate effectiveness of a treatment on symptoms or underlying processes). While there has been some success in the identification of diagnostic biomarkers for neurological disorders, such as Alzheimer’s Disease (Dubois et al., 2007; Jack et al., 2010), researchers continue to struggle to identify firm diagnostic or screening biomarkers of NDDs (Pu and Costafreda, 2013; Thome et al., 2012; Voineagu and Yoo, 2013; Walsh et al., 2011). Numerous studies have been published using machine learning techniques to distinguish ASD or ADHD from TD individuals; however, the results remain inconsistent and do not yet have implications for diagnosis (Uddin et al., 2017). These challenges persist amidst great improvements in developing intelligent classifiers and multimodal imaging techniques: to date, neuroimaging cannot reliably distinguish if a child has or is likely to develop an NDD, neither can it distinguish among disorders. In light of this, a valid question is whether the identification of biomarkers is feasible or probable for these populations.

8.1. Challenges to identifying biomarkers in NDDs

Several challenges are associated with identifying biomarkers for ASD, ADHD, and TS, including the high degree of comorbid disorders, obtaining specificity in classification, and accounting for clinical subtypes of NDDs. Regarding comorbidities, it is unknown what a neural marker would look like for a child that is at high-risk to develop ASD, ADHD, as well as anxiety. For example, would this biomarker be a unique neural representation of developing all three of these disorders, or would it represent the simultaneous presence of three biomarkers for three distinct disorders? This, indeed, is a complicated question as there is evidence for both distinct and overlapping genetic risk factors among these disorders (Lichtenstein et al., 2010; Ronald et al., 2006) and neurocognitive models also struggle to adequately distinguish between NDDs (Craig et al., 2016). The issue of comorbidity has great implication for the identification of biomarkers for NDDs, given that the research and clinical samples for these populations rarely consist of individuals with “pure” disorders. Another challenge is identifying biomarkers that can differentiate one NDD from another. Specifically, most neuroimaging classifier studies only compare a single NDD to TD individuals (e.g. ASD versus TD, ADHD versus TD) and do not differentiate the NDD from related disorders. Of note, Lim et al. (2013) created a classifier to differentiate ADHD from typically developing participants and participants with ASD using structural MRI features. This classifier exhibited a sensitivity of 86.2% and a specificity of 84.2% when identifying ADHD participants as compared to ASD participants. Thus, there is evidence that such specificity between NDDs is possible, and future studies should continue to include subjects from multiple populations. Another issue concerning the feasibility of identifying a biomarker is the likelihood that there may be multiple subtypes of these disorders with many combinations of etiology. The involvement of hundreds of genes, with no single identified cause of ASD (Happe et al., 2006), indicates that a single marker is unlikely. Additionally, biomarkers may change across the lifespan, given that symptoms of NDDs emerge early in life and influence subsequent development (McPartland, 2016). Finally, NDD diagnostic categories are based on clinical delineations of behavior patterns, but it is likely that these clusters of behaviors are the result of different etiologies. Thus, there is a need for fresh and diverse directions in the search for NDD biomarkers.

8.2. Trait-based analyses

In psychiatric research, there has been a call for trait-based analyses that are free from the limitations of diagnostic categories (Morris and Cuthbert, 2012; Casey et al., 2014; Cuthbert, 2014). Most prominently, RDoC has been presented as a method for trait-based, diagnostic-free, psychiatric research (Casey et al., 2014; Cuthbert, 2014; Morris and Cuthbert, 2012). RDoC is in response to Strategy 1.4 of the National Institute of Mental Health (NIMH) Strategic Plan, to “Develop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures” (Morris and Cuthbert, 2012). The RDoC matrix defines five domains of constructs for the study of mental illness: negative valence systems (e.g. active threat/“fear”), positive valence systems (e.g. reward learning), cognitive systems (e.g. attention), systems for social processes (e.g. theory of mind), and arousal/regulatory systems (e.g. arousal and regulation). It also presents eight units of analysis to study these constructs: genes, molecules, cells, circuits, physiology, behavior, self-report, and paradigms. Neuroimaging techniques fall under “circuits”. Thus, instead of merely studying group differences defined by diagnostic categories, trait-based analyses would allow for mechanistic investigations of constructs that are related to pathology. This frees researchers from the constraints of categorization that are based on rendering a binary diagnosis present-absent decision. Instead, it uses symptoms as the units of analysis. For neuroimaging in particular, RDoC is meant to examine how phenotypic heterogeneity is the result of varying severity and degree of neural circuit dysfunction (Morris and Cuthbert, 2012).

Within the neurodevelopmental research community specifically, there have been calls in line with RDoC to study traits that extend across diagnostic categories to account for the comorbidity among disorders (Dougherty et al., 2016; Homberg et al., 2016). In instances where NDDs are heterogeneous and share characteristics, statistical t-tests between them may not yield statistically significant results; however, correlating the behavioral phenotype with a brain response among all subjects may yield significant and meaningful relationships (Dougherty et al., 2016). For example, ASD traits, such as social deficits, have been found to be elevated in children with ADHD (Grzadzinski et al., 2011; O’Dwyer et al., 2016; Reiersen et al., 2007), and this exists even when controlling for the severity of ADHD (Grzadzinski et al., 2011; O’Dwyer et al., 2016). Similarly, there has been overlap of deficits in attention and reward processing in ASD and ADHD as well (Taurines et al., 2012). Furthermore, neuroimaging can be an excellent tool to identify specific neural regions and circuits that contribute to a trait that extends across diagnostic categories. Specifically, O’Dwyer et al. (2016) examined gray matter volumes among children with ADHD, unaffected siblings, and TD children. Among all subjects, the volume of the left caudate nucleus was negatively correlated with symptoms of ASD; this indicated that the caudate contributes to executive functioning within populations that express sub-clinical levels of ASD.

Relating neural function to behavior is associated with several challenges and considerations. First, while we can correlate brain-behavior information, it does not imply causation. As a result, it is often not possible to infer that a neural process directly leads to the development of a clinical symptom. Thus, while researchers commonly use brain-behavior relationships to identify potential targets of intervention, it is often unknown whether affecting the neural target will indeed have an effect on behavior. Inferring these processes is also further complicated by potentially changing brain-behavior relationships across development. Additionally, trait-based analyses are only as valid and informative as the measures used to quantify the traits themselves. Consequently, it is important for researchers to consider the specificity of a measure, its method of measurement, and its relevance to the overall research goals. Researchers should also use measurements that are validated in the clinical sample and the age range of the sample. Commonly, a measure of a single domain actually encompasses multiple sub-processes that may be overlooked by an overall index score. For example, an index score that represents social functioning broadly may not capture the diverse mechanisms that contribute to social skills deficits, such as impulsivity, inattention, theory of mind, anxiety, slowed processing speed, etc. Such a broad measure may create false equivalents between individuals that have vastly different types of
social impairment, and these mechanisms are likely attributable to distinct neural processes. This may be especially relevant when comparing NDDs that have deficits within the same overarching domain, but have distinct profiles within that domain. Thus, measures that tap specific functions will likely have greater ability to identify specific brain networks and patterns underlying that behavior. However, in some cases, it may be beneficial to include a broad measurement to represent overall clinical severity or relate neural function to general diagnostic criteria. Researchers should also consider the advantages and limitations of using rating scales versus neuropsychological testing. Rating scales inherently contain the biases of the specific rater, which may result in different ratings across self, parent, and teacher informant. Additionally, several studies have found that individuals with ASD (Lerner et al., 2012; Johnson et al., 2009) and ADHD (Hoa et al., 2004; Steward et al., 2017; Owens et al., 2007) tend to endorse fewer deficits compared to other informant reports, which may be representative of a decreased awareness of deficits. Thus, researchers must ascertain which type of informant rating most reliably represents the function they are trying to capture. An advantage of rating scales is that they provide information on real-life, clinical function that is difficult to capture in a neuropsychological battery. In contrast, a neuropsychological assessment avoids the bias of a specific rater and may tap more specifically into a domain of interest. Likely, it is useful to include a combination of raters and neuropsychological assessments, capturing both specific and broad domains of function. Thus, although there are several limitations to inferring brain-behavior relationships, careful study design and measurement selection can lead to more accurate characterization of neural networks that relate to symptoms of NDDs.

9. Neuroimaging and genetics

Recently, there have been efforts to integrate genetic and neuroimaging investigations of NDDs. Genetic information provides neuroimaging researchers an index to differentiate between biologically and environmentally mediated variables in normal development and disease processes. One way that genetics and neuroimaging may be integrated is by investigating genetic disorders that concur increased risk for an NDD. In the case of ASD, disorders such as Fragile X and Tuberosus Sclerosis Complex (TSC) can be examined for neural similarities to the broader ASD population (Ameis and Szatmari, 2012). Neuroimaging can also be used to examine the impact of specific and related genetic variants on the neurobiology, such as genes related to dopamine transmission (DAT1 and DRD4) in ADHD (Durston, 2010). There is great potential to pair genetics and neuroimaging findings to identify subgroups of NDDs that have unique biological underpinnings; such endeavors can reveal endophenotypes (the developmental manifestation of a gene or set of genes) (Mahajan and Mostofsky, 2015). For example, neuroimaging can be used to compare individuals with syndromic ASD that do and do not possess a particular genetic variant.

Several initiatives have been taken to further the integration of genetics and neuroimaging in NDDs. The NeuroIMAGE study is a multi-site prospective study that collects behavioral and cognitive measures, structural and functional neuroimaging, and genome-wide genetic information. This database will allow for detailed mapping of gene to neural expression across adolescence and adulthood in individuals with ADHD (von Rhein et al., 2015). The NeuroDevNet ASD Demonstration Project aims to identify genetic variants that are risk factors for ASD, and then compare ASD probands with and without the genetic variants using measurements of cortical thickness, volume, and area (Zwaigenbaum et al., 2011). Further efforts to integrate genetics and neuroimaging in larger samples will greatly improve our understanding of how genes that concur risk for NDDs impact neural structure and function.

10. Neuroimaging in the clinical setting

There is great hope for the integration of brain imaging into the clinic to aid diagnosis, clinical recommendation, and treatment implementation. Specifically for NDDs, using neuroimaging to identify high-risk infants before symptoms emerge could provide better outcomes for early intervention. Nevertheless, challenges remain in integrating neuroimaging, clinical diagnosis, and treatment in the field of NDDs. One challenge involves the identification of reliable predictive markers that extend from the group level to the individual level. Neuroimaging studies investigating biomarkers of NDDs usually follow a machine-assisted classification analysis to predict group membership. Specifically, a computer uses algorithms to learn a set of patterns and rules that accurately sort subjects into groups based on relevant features of the data. In the case of neuroimaging, these features may include measures of functional connectivity strength, gray and white matter volumes, surface area of a region, regional homogeneity, along with other measures. The methodology used in these studies involves training the classifier on one dataset and testing it on multiple larger and diverse datasets in order to determine the validity and reliability of the classifier. The neuroimaging field has seen great advancements in the use of machine learning classifiers, particularly for ASD and ADHD; specifically, the salience and default mode networks have emerged as potential discriminating features for ASD, while frontal and cerebellar regions appear to be important to classifying ADHD (Uddin et al., 2017). However, these markers are not yet reliable or reproducible, and these neural markers do not provide diagnostic information beyond what current behavioral measures provide. Additionally, there has been growing awareness that there is no single neurobiological feature that distinguishes these disorders and there is a call to break apart these larger diagnostic categories to examine subgroups and individual variation (Muller and Amaral, 2017; Waterhouse and Gilberg, 2014). As a result of this neural similarity, most classifiers can only differentiate a single NDD from the TD population; few have effectively discriminated among NDDs, such as ASD and ADHD (Lim et al., 2013; Uddin et al., 2017). Thus, most progress to date has been in differentiating brain function that is “not typical” from “typical”, and the clinical utility of this is likely limited. On the one hand, the ability to detect atypical brain function during early development, such as in infancy before symptoms emerge, could allow for the direction of general early intervention services that could have broad positive effects for a range of clinical disorders. This becomes less beneficial as symptoms emerge and targeted interventions become necessary. There are also additional ethical implications of notifying a parent that their young child has atypical brain function without providing additional explanation, as this could refer to a range of disorders of varying severity and onset, such as ASD, anxiety, schizophrenia, personality disorders, etc. Finally, while classifiers may be successful at classifying a portion of subjects, they are not yet reliable and accurate at the individual level. If neuroimaging methods can accurately identify those at risk of developing the disorder, or can provide additional objective information to aid diagnosis, it will be an integral part of the clinical diagnosis of NDDs.

Neuroimaging is currently used effectively in other clinical settings, including neurosurgical mapping (Bullmore, 2012; Matthews et al., 2006), identification of pharmacological drug targets (Matthews et al., 2006), tracking intervention (Bullmore, 2012), and guiding neuropsychological practice (Bigler, 2015). Several studies have indicated that deep brain stimulation has promising clinical impacts for adults with TS (Schrock et al., 2015; Servello et al., 2008), and presurgical mapping using structural or fMRI could be of use. Neuroimaging also presents the opportunity for tracking treatment paradigms in NDDs. For example, a behavioral intervention for ADHD could be monitored throughout implementation for changes in the executive control network during rs-fMRI. Otherwise, transcranial magnetic stimulation (TMS) paradigms (including TMS, transcranial direct current stimulation, repetitive TMS) could be tracked using fMRI for evidence that the
network of choice is being targeted. For example, several studies have indicated the efficacy of rTMS in affecting social functioning in ASD (Enticott et al., 2014, 2011), attention in ADHD (Bloch et al., 2010), and tics in TS (Kwon et al., 2011). Also, fMRI could be used as a substrate for neurofeedback for training the brain in behaviorally relevant ways. Research has begun to show the impact of neurofeedback paradigms in ASD (Cohen et al., 2010), ADHD (Lofthouse et al., 2012), as well as TS (Farkas et al., 2015); the combination of monitoring the neural effects of these treatments with neuroimaging could prove to be clinically significant in the future. In traumatic brain injury clinical setting, neuropsychologists have begun to integrate information about neural structure into case conceptualization and treatment (Bigler, 2015). Technologies such as NeuroQuant can be used to compare patient’s structural volumes to age-matched norms, providing quantitative statistics on neural structure and composition that can explain symptoms (Ross et al., 2012). It is possible that neuroimaging could similarly aid the diagnosis and treatment of NDDs by providing important information on structural variations that are related to neuropsychological profiles. Thus, in several ways, neuroimaging has already begun to be integrated into clinical psychological settings that may become relevant for NDDs.

There are several barriers to the general application of neuroimaging in the clinic for NDDs. Siegle (2011) offers three things that must take place before the implementation of neuroimaging into the clinic as a diagnostic tool: First, there must be standardized databases that define “normal” ranges of neural functioning for a given age and demographic profile. This will entail reporting MRI data in an easy-to-understand way that is applicable across many sites, controlling for potential cross-scanner noise, and standardizing the implementation of protocol (Matthews et al., 2006). Second, there should be standardized procedures for interpretation of neuroimaging results that can easily be interpreted by physicians. Third, these technologies must be efficient and affordable for effective implementation. Thus, while there is great potential for neuroimaging to positively impact the diagnosis and treatment of NDDs in the clinical setting, there is still much progress that must be made. Neuroimaging will likely first be clinically used to monitor treatment before aiding diagnosis, as there are not yet clear biomarkers for these disorders.

11. Conclusions

Neuroimaging advances over the last several decades have led to significant progress in our understanding and conceptualization of NDDs. From identification of abnormal brain regions and networks, conceptualization of NDDs as disorders of aberrant connectivity, to examinations of the neural developmental trajectories associated with these disorders, neuroimaging has offered effective theoretical explanations for these disorders that would not have been reached by behavioral techniques alone. Amidst such measurable gains and progress, neuroimaging struggles to fully elucidate the neural mechanisms of NDDs and to have clinical impact. Instead, the field is troubled by inconsistent results and poor generalizability of findings. This is likely related to the unique methodological and conceptual challenges associated with neuroimaging of children with NDDs. Specifically, there are difficulties associated with correcting for head motion, normalization to pediatric brain templates, accounting for psychotropic medication use, delineating complex developmental trajectories, and overcoming smaller sample sizes. Additionally, complex genetic etiology, heterogeneous clinical presentation, and high rates of comorbid disorders complicate the picture. This paper discussed how these issues directly impact the study of ASD, ADHD, and TS and offered recommendations for addressing these issues and moving the field forward. Namely, research samples should be increasingly heterogeneous and include variety of medication status, comorbidities, and degree of impairment to best reflect the clinical population. There should also be continued investment in prospective, longitudinal studies that make use of data-sharing platforms and integrate trait-based analyses approaches. Before neuroimaging can offer diagnostic utility of NDDs, it must effectively distinguish between multiple clinical diagnoses and offer information beyond what a behavioral diagnosis can provide, such as further objectivity, earlier identification, and understanding of the affected neural mechanism. Thus, while neuroimaging is an appropriate and invaluable tool to study NDDs, there is still a long way to go to overcome these methodological obstacles, to use neuroimaging to fully understand the nature of these disorders, and to demonstrate its clinical utility.

References


H.M. Bednarz, R.K. Kana

Neuroscience and Biobehavioral Reviews 90 (2018) 50-69


