Published in final edited form as:
J Child Psychol Psychiatry. 2014 June ; 55(6): 659–680. doi:10.1111/jcpp.12185.

Annual Research Review: Current limitations and future directions in MRI studies of child- and adult-onset developmental psychopathologies

Guillermo Horga, **Tejal Kaur**, and **Bradley S. Peterson**
Department of Psychiatry, New York State Psychiatric Institute and College of Physicians and Surgeons, Columbia University, New York, NY, USA.

Abstract

The widespread use of magnetic resonance imaging (MRI) in the study of child- and adult-onset developmental psychopathologies has generated many investigations that have measured brain structure and function in vivo throughout development, often generating great excitement over our ability to visualize the living, developing brain using the attractive, even seductive images that these studies produce. Often lost in this excitement is the recognition that brain imaging generally, and MRI in particular, is simply a technology, one that does not fundamentally differ from any other technology, be it a blood test, a genotyping assay, a biochemical assay, or behavioral test. No technology alone can generate valid scientific findings. Rather, it is only technology coupled with a strong experimental design that can generate valid and reproducible findings that lead to new insights into the mechanisms of disease and therapeutic response.

In this review we discuss selected studies to illustrate the most common and important limitations of MRI study designs as most commonly implemented thus far, as well as the misunderstanding that the interpretations of findings from those studies can create for our theories of developmental psychopathologies. Those limitations are in large part responsible thus far for the generally poor reproducibility of findings across studies, poor generalizability to the larger population, failure to identify developmental trajectories, inability to distinguish causes from effects of illness, and poor ability to infer causal mechanisms in most MRI studies of developmental psychopathologies. For each of these limitations in study design and the difficulties they entail for the interpretation of findings, we discuss various approaches that numerous laboratories are now taking to address those difficulties, which have in common the yoking of brain imaging technologies to studies with inherently stronger designs that permit more valid and more powerful causal inferences. Those study designs include epidemiological, longitudinal, high-risk, clinical trials, and multimodal imaging studies. We highlight several studies that have yoked brain imaging technologies to these stronger designs to illustrate how doing so can aid our understanding of disease mechanisms and in the foreseeable future can improve clinical diagnosis, prevention, and treatment planning for developmental psychopathologies.

Correspondence Bradley S. Peterson, 1051 Riverside Drive, Unit 74, New York, NY 10032, USA; PetersoB@nyspi.columbia.edu.
Conflicts of interest statement: No conflicts declared.
The authors have declared that they have no competing or potential conflicts of interest.

Keywords

Brain imaging; magnetic resonance imaging; study design; psychopathology; development

Introduction

Magnetic resonance imaging (MRI) is one of very few non-invasive techniques that permit *in vivo* assessment of brain structure and function without exposure to radiation. Given the ethical constraints of radiation-based imaging in children, MRI has been used almost exclusively in imaging studies of normal brain development and developmental psychopathologies. Similarly, MRI has become the most common technique in imaging studies of adult psychopathologies, largely because of its widespread availability and non-invasiveness, and particularly because of its ability to quantify neurobiological features throughout the brain across numerous modalities, including brain structure, the concentrations of chemical metabolites and neurotransmitters, the integrity of nerve fiber tracts, and activity-dependent changes in blood flow.

By virtue of its ability to provide a window into the living, developing brain, MRI as a technology has the unique potential to illuminate the core pathophysiological mechanisms of mental illness in children and adults and to help identify novel therapeutic targets that will improve their treatments. Despite its unquestionable and extraordinary value as a technological tool, however, the use of MRI in studies of applied neuroscience and developmental psychopathologies thus far, in some important ways, has fallen short of its promises in what it has taught us about the mechanisms of psychiatric disease and its treatments. Some of the most prominent concerns for the limited progress made in MRI studies thus far include:

1. Most MRI studies have been designed as case-control studies in which participants have been drawn from samples of convenience, usually from local clinics of chronically ill and already treated patients. The differing sample characteristics across studies have contributed to the inconsistent findings across imaging studies. Because reproducibility is a precondition of validity, the generally poor reproducibility of MRI findings across studies (Valera et al., 2007, Hart et al., 2013, Ahmed et al., 2012, Hulvershorn et al., 2011, Stanfield et al., 2008, Hirai and Jones, 1989, Arnone et al., 2009) has generated a body of work that is, to a considerable extent, of questionable validity and value. Moreover, biased sampling has produced MRI findings that are unlikely to generalize to the larger population of patients with that illness. The poor generalizability associated with sampling bias and the poor reproducibility of positive findings have both been compounded by the poor statistical power associated with small sample sizes of most imaging studies, which is generally a consequence of the relative expense and technological complexity of conducting MRI studies. These problems of reliability, validity, and generalizability associated with small sample sizes are only worsened by the frequent use of liberal statistical thresholds and poorly specified a priori hypotheses, as noted elsewhere (Button et al., 2013, Philip et al., 2012).

2. Very few imaging studies have thus far identified valid trajectories of brain development in health and illness. Many cross-sectional studies have reported age correlates of MRI measures that have been interpreted as developmental trajectories, but most of those inferred trajectories are likely to be incorrect because, as is well known and amply documented, valid developmental trajectories cannot be inferred from data acquired only in temporal cross-section (Kraemer et al., 2000).

3. The vast majority of findings reported in studies of psychiatric illnesses have been unable to distinguish the causes of an illness from its effects or its treatments. Many of the findings reported in the literature, for example, have been interpreted as likely representing a central pathogenic mechanism that causes the illness, when in fact subsequent findings suggest that the initial findings represent instead either the downstream, epiphenomenal effects of illness (such as the effects of stress that chronic illness inevitably bring) or responses of the brain that help to compensate for the presence of illness.

4. Imaging studies thus far have largely failed to identify the causal mechanisms that produce developmental psychopathologies or of the interventions that have been used to treat them.

5. Imaging studies have not yet integrated information across imaging modalities to understand how anatomical, functional, neurochemical, and connectivity characteristics of the brain relate to one another in any given disorder, or which of these characteristics is the more fundamental of the various disturbances identified in an illness.

6. Imaging studies have not proved useful for improving clinical diagnosis or for planning individualized treatments.

Herein we review in detail these various limitations of MRI research thus far in improving our understanding of the neural bases of developmental psychopathologies in children and adults. We suggest experimental approaches to address each limitation and highlight studies that have successfully employed them. We emphasize that MRI is simply an experimental tool that provides an interesting set of measures of the brain and that no tool in and of itself can produce valid or reproducible findings. No tool or measure, in other words, can substitute for sound study design. When MRI is used without regard to the strengths and limitations of the design of the study in which it is used, this interesting tool fails to achieve its intended purpose and cannot realize its fullest potential. In contrast, when MRI techniques are integrated into sound study designs that are developmentally and pathophysiologically informative, the great potential of MRI to reveal the core mechanisms of neuropsychiatric illness and their treatments can be more fully realized.

The fundamental challenge of imaging studies, and the challenge for any tool in studying the pathogenesis of human illness, for that matter, is the impossibility of conducting a true “experiment” in which a variable is experimentally controlled and manipulated to understand its pathogenic role in the illness. To state the crux of the problem rather crudely, human beings cannot be assigned randomly to have an illness or not. Because the illness

variable cannot be experimentally controlled and manipulated, we must rely both on naturalistic associations of imaging measures with the illness variable and on study designs other than experimental manipulations to infer the underlying causal mechanisms of developmental disorders. Imaging studies have increasingly invoked study designs that permit stronger causal inferences than most study designs have permitted previously.

Given our current focus on the influences of study design on the findings of imaging studies, we will not discuss in depth the technical limitations of MRI studies. Technical limitations that are independent of study design—such as inadequate statistical analyses, small sample sizes that are underpowered to detect hypothesized effects, the use of unreliable imaging measures, or the use of unsuitable control groups (Button et al., 2013, Schwartz and Susser, 2011)—clearly can compromise the validity and reproducibility of findings from MRI studies. Our review, however, will concentrate on the limitations of the design of imaging studies because they have received relatively less attention thus far in the literature, and because they are generally independent of technical or statistical problems in imaging studies.

Limitations of case-control designs with samples of convenience: the importance of ascertainment bias

Participants in case-control studies are selected on the basis of the presence or absence of an illness in order to determine the relative risk that a particular “exposure” or past attribute likely was involved in causing the illness. When both groups of individuals are drawn from representative and demographically matched populations, and when those groups are demographically matched to one another, the “differences” that emerge between the ill and well groups likely reflect an underlying causal risk. In fact, the appropriate use of a case-control design *requires* that both groups derive from representative populations. Sampling representative populations ensures that the differences detected between groups validly associate with the disease process. Nevertheless, in the absence of additional elements of a developmentally informed study design, the use of representative populations alone cannot generate findings that unambiguously differentiate the effects of an illness from its causes.

In contrast, when individuals who are neither representative of all people in the population who have the illness nor demographically representative of the general community are recruited into a case-control study, the “differences” detected across the cases and controls likely represent their differing demographic and socioeconomic characteristics, not the relative risk that the measured difference carries for the illness. Thus when case-control imaging studies recruit participants from samples of convenience, such as local clinics, patient advocacy organizations, or particular socioeconomic groups, the differences detected in the brains of the cases and controls likely represent features that derive from these ascertainment biases rather than from the features of the brain that ultimately cause the illness. Moreover, studies that recruit predominantly from local clinics or inpatient units of a hospital may recruit selectively for particular subtypes of illness or certain social-demographic characteristics. Different clinics have different demographics of patient populations and different reasons for their patients coming to clinical attention and seeking care, and the various features of their illnesses and comorbidities, and the social forces that

bring those patients to clinical attention, vary dramatically across sites. Therefore, case-control imaging studies that recruit from different clinical populations will differ in the imaging findings that those clinical samples yield.

Take, for example, the fact that although 60% of individuals with autism have an IQ less than 70, most imaging studies of autism recruit predominantly “high functioning” individuals with IQs above 70 (Chakrabarti and Fombonne, 2005, Stanfield et al., 2008). Findings in those samples cannot be assumed to accurately represent the brain features that would be found in the general population of all persons with the illness. Meta-analyses provide additional evidence that the significant heterogeneity in findings across study sites may derive from differences in sampling of populations with autism. A reduction in the cross-sectional area of cerebellar vermal lobules VII, for example, was one of the first anatomical abnormalities reported in persons with autism (Courchesne et al., 1988) and was thought to contribute to cognitive abnormalities associated with the illness. This finding, however, has been detected inconsistently in subsequent studies (Kleiman et al., 1992, Hardan et al., 2001, Hashimoto et al., 1995). Meta-analyses suggest that a smaller vermis may associate with a younger age and lower IQ in persons who have autism, and therefore this finding may not be present in either older or higher functioning persons in the population (Stanfield et al., 2008). Although longitudinal studies of representative populations are necessary to clarify the sources for inconsistent findings across sites, we posit that much of the inconsistency derives from differing sampling or recruitment procedures that produce clinical and demographic differences in study populations across sites.

Often when critiquing the generalizability of the population recruited into a study, we assess how well the ill participants match the distribution of demographic and clinical characteristics expected in the general population, and we attend less to the representativeness of the healthy control sample. Most studies include a healthy control group of participants who have no axis I disorders, even those disorders that are present in many individuals within the patient group (Schwartz and Susser, 2011). The use of these “well controls” may actually decrease the validity of findings from a study because differences detected between the ill and well groups may not reflect the disorder of interest, but instead the differences associated with comorbid illnesses (Schwartz and Susser, 2011). Furthermore, the inclusion of “well controls” can bias the detection of group differences. A critical review of neuroimaging findings of lateral ventricle size in schizophrenia concluded that although absolute volumes appeared consistent across studies of persons with schizophrenia, the ability to detect a difference between patient and control groups varied depending on the use of “well” controls or matched controls (Chua and McKenna, 1995). On the other hand, control participants recruited as samples of convenience can also include individuals who are less healthy than expected. For example, control participants who respond to advertisements often yield children whose parents may have had long-standing concerns about their child's health and who view the study as one means of acquiring a free assessment of their child (Harth and Thong, 1990). Samples of convenience also tend to recruit either unusually altruistic families who want to help society, or impoverished families who view study participation as a means to supplement income (Harth et al., 1992, Hoberman et al., 2013, Harth and Thong, 1990). These ascertainment biases may be

represented in the brains of these children in some unknown way. Those brain features, whether a product of ascertainment bias in the ill or well groups, may obscure the true distinguishing features of the illness being studied. Finally, the problematic effects of differing ascertainment bias in imaging findings across differing recruitment sites are only compounded by the use of differing imaging methods across sites, such as use of differing scanners and field strengths, different pulse sequences, and different methods for processing the images.

Suggested solutions—Imaging designs should increasingly minimize selection bias by recruiting representative samples of study participants. One possible option, which can require considerable resources, is to recruit and scan a representative, epidemiologically ascertained sample (Waber et al., 2007, Sowell et al., 2004). Another option involves yoking imaging studies to already ascertained epidemiological samples, or at least a representative subset of such a sample. This latter approach is more feasible to accomplish than creating an epidemiological sample de novo, because it recruits from a population whose members have established a relationship with the primary study team and have demonstrated a willingness to participate in research, making high rates of participation, and therefore representative sampling, more likely. In addition and most importantly, however, the imaging of participants in these representative samples minimizes confounding from ascertainment biases, and therefore the findings from these samples should generalize to the larger population.

An added benefit to recruitment from epidemiologically ascertained populations comes when the epidemiological study has prospectively collected biological or environmental data at one time point, and imaging is conducted at a second time point, allowing inferences to be made concerning how prior biological or environmental risk factors may affect subsequent structural and functional brain phenotypes (Paus, 2010). One epidemiological study, for example, followed from pregnancy a non-clinical, community based sample of children who had varying levels of prenatal exposure to the pesticide chlorpyrifos (CPF) and showed that prenatal CPF level was associated with lower full-scale IQ in children at age seven (Rauh et al., 2011). When imaging was yoked to the study, forty of the original 369 children were stratified as either low- or high-exposure to CPF (Rauh et al., 2012). Because only a subset of the original sample was scanned, it is important to note that nearly all eligible participants in the low- and high-exposure groups participated in the imaging study, and that neither the low nor high-exposure children differed from each other in demographic or birth characteristics, nor did the scanned sample differ from the original sample, all of which suggests that the imaging study retained the representative sampling present in the original epidemiologically ascertained cohort. The study found that CPF exposure correlated positively, and in a dose-dependent manner, with regional enlargements of the cerebral surface and with thinning of frontal and parietal cortices. The dose-response relationship between CPF levels and brain structure, combined with the fact that exposure to the neurotoxic agent temporally preceded the imaging measurements and cognitive assessments, supports the strong causal inference that CPF exposure produced the imaging findings and cognitive abnormalities in this sample. These causal inferences would not have been possible in the absence of a community-based, representative sampling of participants.

Moreover, the availability and use of the imaging technology allowed identification of the brain regions that CPF affects and a preliminary understanding of the biological pathways through which CPF disrupts cognition.

Although recruiting from epidemiological samples increases the generalizability and validity of findings in studies that employ case-control designs, generalizability may still be compromised when the individuals who consent to participate in the imaging component differ from the subset who decline participation. The latter subset often excludes the more severely affected participants who are unable to comply with the demands of the scanning procedures. Further, if scanning requires travel to the study site, scanned individuals may be those who have greater functional capacity and thereby can negotiate the complexities of travel. When assessing the generalizability of findings, even of an imaging study yoked to an epidemiologically ascertained sample, it is essential to note whether the imaged sample differed statistically in demographic or clinical measures from the original sample. Of course the ideal way to ensure that the imaging subsample is representative of the larger sample from which it is drawn is to ensure that nearly all members of the larger sample, or some randomly selected subsample, participate in the imaging study.

Limitations of cross-sectional studies: inability to identify developmental trajectories

Cross-sectional imaging studies typically compare one brain imaging measure across two or more diagnostic groups, and the group differences are usually interpreted as representing an abnormality in the patient group that has contributed causally to the disease process. Even when groups are demographically matched to one another, a cross-sectional study cannot distinguish the effects of an illness from its causes. Cross-sectional studies also often report age correlations that are interpreted as representing the developmental trajectory of the imaging measure in one or both groups. Those findings from cross-sectional studies, however, often confuse and mislead us in our understanding of the true natural history and developmental correlates of an illness (Kraemer et al., 2000). Investigators of cross-sectional studies, or at least readers of them, commonly misinterpret age correlations by assuming that younger and older members of the sample who nominally share the same diagnosis belong to a single larger population of participants who have the same illness, so that younger subject will, with time, be identical to their older counterparts in the pathogenesis of their illness as well as in its phenomenology, familial risk, patterns of comorbidities, natural history, and treatment responsiveness. These assumptions, however, are nearly always invalid for developmental psychopathologies, as essentially all of these features differ between illnesses that begin in childhood or adulthood. Moreover, these features usually differ between younger and older participants with childhood onset illness because most illnesses with a childhood onset improve in some of the children over time, so that those who are ascertained in samples of convenience from treatment clinics will recruit older participants who have more persistent illness than the total sample of younger participants, because the older participants who have improved have preferentially left the clinic. Therefore, when recruiting participants of differing ages into a cross-sectional study, ascertainment biases lead to differing subtypes of illness being preferentially recruited in childhood and adulthood. Because children ascertained in temporal cross-section almost certainly will not grow into their adult counterparts in the study, the statistically significant

differences between children, adolescents, and adults, and the correlations of imaging measures with age within a cross-sectional study, cannot be equated with the true developmental trajectory of the imaging measure in persons who have the illness. Instead, those findings much more likely represent the differences between age groups that derive from powerful ascertainment biases that determine who is recruited and who is not recruited into the imaging study.

In our own studies of Tourette syndrome, for example, we reported larger prefrontal cortices in affected children but smaller prefrontal cortices in affected adults compared with age-matched, healthy controls (Peterson et al., 2001). Further, we found that cortical volume correlated inversely with symptom severity. Rather than suggesting that cortical volumes may decline over the life course in Tourette syndrome, we interpreted the findings from this study as representing a selection bias in which the adult participants, who were patients ascertained from a local clinic, represented the smaller subset of persons with a lifetime history of Tourette syndrome who continue to display symptoms past childhood, who have more severe illness, and who likely are unable to compensate for symptoms through prefrontally mediated pathways. Subsequent studies have shown beyond any doubt that adults who have persistent illness differ from children and adolescent patients in measures of brain structure and function (Plessen et al., 2009), and together those studies suggest strongly that an ascertainment bias that derives from recruiting still-symptomatic adults is responsible for these age-specific differences in imaging measures across diagnostic groups.

Studies of persons with autism also suggest the presence of significant age effects in multiple brain regions. Enlargement of the cerebral hemispheres is generally, but not consistently, detected in persons with autism, and studies suggest that although cerebral enlargement may characterize individuals in early childhood, cerebral enlargement may either “normalize” or reverse in adulthood (Duerden et al., 2012). Relative to controls, children with autism were more likely than adults with autism to have increased gray matter in the fusiform gyrus, cingulate gyrus, and insula, regions thought to support the processing of social context (Duerden et al., 2012). Despite studies suggesting that an enlarged amygdala in individuals with autism associates with the core deficits of the disorder (Mosconi et al., 2009, Schumann et al., 2009, Kim et al., 2010), other studies suggest that volumes of the amygdala are either small (Pierce et al., 2001, Nacewicz et al., 2006) or similar to volumes in typically developing controls (Haznedar et al., 2000, Corbett et al., 2009, Palmen et al., 2006, Nacewicz et al., 2006). Meta-analyses suggest that this inconsistency may derive from enlarged amygdala volumes in young children but normal volumes in older groups (Dickstein et al., 2013). Although many of these studies suggest that brain abnormalities may “normalize” over time, this inference cannot be made from cross-sectional data alone.

Cross-sectional studies of attention deficit hyperactivity disorder (ADHD) have also produced variable findings across multiple brain regions, despite the presence of reasonably consistent findings of prefrontal hypoactivation across both adults and children (Cortese et al., 2012). For example, a meta-analysis of cross-sectional data suggests that hypoactivation in the somatomotor system is less prominent in adults than in children with ADHD, whereas adults with ADHD have greater activity in visual and dorsal attention systems (Cortese et

al., 2012). These findings may suggest that abnormal functional activity in prefrontal regions characterizes illness in both children and adults, whereas hypoactivation of somatomotor regions in children may attenuate and "normalize" as motoric hyperactivity decreases with age. Conversely, the increased activity within the dorsal and ventral attentional systems of adults could represent neural mechanisms that compensate for the presence of illness to support the improvement in clinical symptoms over adolescence and young adulthood. However, in the absence of longitudinal studies, we cannot validly interpret from these cross-sectional studies alone how brain changes may associate with symptom measures throughout development.

Cross-sectional studies of psychotic populations similarly yield findings that may misinform interpretations of the developmental trajectories of the brain that relate to the illness. Cross-sectional studies that recruit on the basis of differing phase of illness attempt to generate the same findings as those that would be obtained from longitudinal studies. These studies may recruit, for example, adolescents and young adults who are at high-risk for schizophrenia and individuals who are either in their first episode or a chronic phase of illness to assess how an imaging measure that varies with age is associated with progression of the illness (Velakoulis et al., 2006). Yet without longitudinal follow-up of the participants in the earlier phases of illness in these samples, we cannot know whether participants recruited as a high-risk group will develop a full-blown psychotic disorder, and if they do develop full-blown psychosis, whether an age-related change in imaging measure would accompany this transition. Further, individuals in different stages of schizophrenic illness differ greatly in their phenotypes, degrees of familial risk, associated comorbidities, life stressors, and treatment histories. The oft-made assertion that an imaging finding, such as hippocampal age-related "atrophy," causes the illness or contributed to its developmental progression, can never be proved in a study of cross-sectional design. In fact, that developmental interpretation seems highly unlikely at face value, in light of the numerous ascertainment biases that create systematic differences across participants of differing ages or phases of illness in the study sample.

Suggested solutions—Longitudinal imaging studies have clarified, extended, and often corrected interpretations of the initial findings generated by cross-sectional imaging studies. For example, one cross-sectional study in children and adolescents with ADHD demonstrated the presence of reductions bilaterally in inferior prefrontal and anterior temporal cortices (Sowell et al., 2003). The absence of differing age correlates across the ADHD and control groups suggested that these were likely to be static deficits from young childhood through later adolescence. A larger, longitudinal study subsequently demonstrated, however, that these reductions in frontal and temporal cortices were likely the consequence of a developmental delay in which prefrontal and anterior temporal regions are the last to mature, lagging behind in the ADHD group by an average of 3 or more years relative to the typically developing control participants (Shaw et al., 2007). This maturational delay likely would never have been detected, and it could not have been proved, in a cross-sectional study alone.

Two longitudinal imaging studies of young children between 1 and 5 years of age have reported altered trajectories of brain development in Autism Spectrum Disorders (ASD)

(Schumann et al., 2010, Hazlett et al., 2011). Both studies reported that brain volumes in infants with ASD were larger than brain volumes in typically developing controls at multiple times points, beginning prior to age two. The findings of the studies are thus highly consistent with one another and inform us immensely about the abnormal developmental trajectories of early brain development in ASD. As the studies only scanned children through age 5, however, we cannot yet know how developmental trajectories in these volumes continue to change through later childhood and adolescence. In addition, these studies document that brain enlargement occurs early in the course of illness and continues into early childhood, but the participants recruited into these samples were already ill; therefore, we cannot know from these studies whether brain enlargement precedes or follows onset of illness, nor whether enlargement is a potential cause or consequence of the illness. Although longitudinal studies better inform us than do cross-sectional studies about the developmental trajectories of brain measures in health and illness, even longitudinal studies of already-ill individuals cannot by themselves disentangle the causes of an illness from its consequences.

Despite the strengths of longitudinal study designs in identifying true developmental trajectories, we note that the validity of findings from longitudinal studies can be undermined substantially by the effects of participant dropout over time, particularly in studies that use an “accelerated longitudinal design”, in which longitudinal data are collected in multiple age cohorts over time. Accelerated longitudinal designs are intended to provide a way to assess developmental trajectories over a broad age range in a study of relatively short duration (Harezlak et al., 2005). One imaging study that employed such a design is the NIMH Childhood Onset Schizophrenia study, which longitudinally followed multiple age cohorts spanning early childhood to young adulthood. Participants were scanned approximately every two years. The majority of participants were scanned at only two time points, and some participants were scanned at only one time point. Thus the majority of participants were not scanned longitudinally from childhood through adulthood. The findings from that study suggested that exaggerated cortical thinning may be an endophenotype of schizophrenia, as it is more prominent in young people with schizophrenia and in their unaffected siblings than in healthy controls (Gogtay, 2008). The repeated measures within the study also suggested that cortical thinning in the sibling cohort may “normalize” between the ages of 17 and 20 (Gogtay et al., 2007, Mattai et al., 2011). The study, however, suffers from a substantial “cohort” effect, wherein the participants who were scanned at a younger age do not necessarily correspond longitudinally with the participants scanned at an older age, in that participants were recruited into the study at various ages, and then the cohort suffered substantial dropout. Thus we cannot know whether the age-related normalization of cortical thickness would be true for all participants if they had all been recruited into the study at roughly the same age and had all been retained in the sample as they were followed and imaged longitudinally. It is possible, for example, that a more rapid thinning of the cortical mantle produces greater debility and thereby predisposes to dropout from the study, in which case older participants retained in the study would have less prominent cortical thinning, and cortical thickness in the retained cohort would erroneously appear to normalize with increasing age. Accelerated longitudinal designs can provide important preliminary data in disorders of low prevalence in which

recruiting representative samples proves difficult. However, cohort effects can significantly undermine the validity of the developmental interpretations from the findings of these studies.

Another challenge associated with MRI studies of development, in particular longitudinal studies, derives from to the marked qualitative and quantitative anatomical changes that the brain undergoes across development. These changes render desirable the use of age-appropriate brain templates and normalization procedures needed to transform individual images of the brain into a common stereotactic space (Almli et al., 2007), so as to minimize the introduction of bias within image processing procedures that can influence findings, at least for a given age group (Fonov et al., 2011).

Lastly, most longitudinal studies using fMRI to study task-related brain activations have assumed that age-related changes in blood oxygenation level dependent (BOLD) responses to specific tasks can be interpreted as representing corresponding age-related changes in underlying neural activity. However, changes in the cerebrovascular system with age can alter the BOLD signal, complicating this interpretation. These limitations in interpreting longitudinal fMRI data can be overcome to some extent by adding to the imaging measures arterial spin labeling (ASL) combined with a hyperoxic challenge, known as “calibrated fMRI”. This technique provides measures of cerebral blood flow (CBF) and BOLD signal that can help to quantify the various components of the BOLD signal, including changes in cerebral metabolic rate of oxygen. Calibrated fMRI can therefore help to determine whether age-related changes in BOLD signal represent true age-related changes in underlying neural activity and responsivity (Mohtasib et al., 2012). The validity of longitudinal studies using fMRI can be further compromised by use of fMRI measurements with low test-retest reliability, a key characteristic of fMRI studies that only recently has begun to receive the attention it deserves (van den Bulk et al., 2013). Similarly, the effects of habituation or learning on fMRI measurements over time should be considered because they can bias longitudinal measurements. Thus, although longitudinal study designs can improve our ability to identify true developmental trajectories in imaging measures, fMRI in particular has additional limitations when used to study age correlates in longitudinal studies, limitations that can be overcome only by technological adjustments to the imaging measures themselves (for example, using calibrated fMRI) or by using appropriate experimental designs that minimize random or systematic errors in the acquisition of measurements longitudinally.

The limitations of studying already-affected individuals: inability to distinguish causes from consequences of illness

Perhaps the major current challenge common to nearly all brain imaging studies is distinguishing which findings represent core pathophysiological processes, which findings are simply epiphenomena, and which represent compensatory or adaptive changes of the nervous system to the presence of illness (Peterson, 2003). Mental illnesses, especially those that manifest early in life, produce chronic, debilitating symptoms that have numerous physical, emotional, cognitive, and social consequences for the individuals who have those illnesses, and those consequences are likely to be represented prominently in the brain as

epiphenomenal effects. The symptoms associated with mental illness in already-affected individuals are thus likely to induce secondary changes in neural systems distributed throughout the brain that coexist with the primary brain disturbances that produced and maintained symptoms of the illness in the first place. These primary brain abnormalities that caused the illness can be exceedingly difficult to disentangle from the secondary ones that are its consequences. Long-lasting brain changes induced by treatment further complicate attempts to disentangle the causes of an illness from its effects.

Among these secondary abnormalities are changes that represent compensatory responses to the presence of illness that originates elsewhere in the brain. Compensatory responses are likely represented commonly in the findings of brain imaging studies, given that the brain is quintessentially a homeostatic organ (Karatsoreos and McEwen, 2013). Despite the likely presence of compensatory changes in the brain, imaging studies of already-affected individuals almost invariably interpret group differences in an imaging measure as representing a cause of the illness rather than as a possible compensatory response. Undoubtedly, identifying compensatory responses of the brain can itself be of great scientific importance, as it can help to identify targets for the development of new treatments. But the fact remains that MRI studies as thus far designed and executed have generally been unable to discriminate the causes of an illness from its compensatory responses, or from the epiphenomenal effects of the illness, or from the consequences of prior treatments. These failings in many respects have left the field of brain imaging adrift in a sea of findings that are difficult to interpret in terms of their true relevance to pathogenesis and treatment.

Imaging findings from already-affected patients thus inherently preclude us from discerning the causes of an illness from its consequences, an interpretive problem that can be amply illustrated with findings from cross-sectional studies in patients with obsessive-compulsive disorder (OCD). Resting hypermetabolism and elevated blood flow in OCD patients in fronto-striatal circuits, including the anterior cingulate cortex, inferior prefrontal cortex, and striatum, has long been interpreted as evidence favoring that increased activity in these regions produces OCD symptoms such as obsessions and compulsions (Baxter et al., 1987, Baxter et al., 1988, Breiter et al., 1996, Rauch et al., 1994, Swedo et al., 1989). This interpretation was subsequently reinforced by findings that activity in fronto-striatal regions correlates positively with severity of OCD symptoms and that successful treatments for OCD normalize this activity (Baxter et al., 1992, Benkelfat et al., 1990, Rubin et al., 1995, Schwartz et al., 1996, Swedo et al., 1992, Lazaro et al., 2008). But whether elevated activity in fronto-striatal circuits is a cause or a consequence of OCD symptoms is unclear. Indeed, findings from studies of Tourette syndrome and other conditions have called into question the interpretation that elevated fronto-striatal activity is a cause of OCD, for several reasons.

First, fronto-striatal circuits increase in activity during the voluntary suppression of motor tics in Tourette syndrome in regions similar to those where activity was reported to be increased in OCD (Peterson et al., 1998). The similar patterns of fronto-striatal activity during tic suppression in Tourette syndrome and during rest in OCD suggest that the increased activity in OCD could be a consequence of the general requirement that patients with OCD suppress their compulsions in the scanner or during the period of radioligand

uptake before the scan. The suppression of compulsions and tics may therefore engage similar regulatory circuits which, rather than causing these symptoms, help to regulate or control them under certain conditions. Support for this alternative interpretation comes from a variety of studies in which the task or scanning environment requires suppression of some kind of automatic response. For example, activation of frontostriatal circuits has been reported during performance of tasks that require inhibitory control, such as the Stroop, Simon, and go/no-go tasks (Bush et al., 1998, Carter et al., 1995, Leung et al., 2000, Pardo et al., 1990, Peterson et al., 2002, Peterson et al., 1999, Horga et al., 2011, Kerns et al., 2004), during pain studies in which participants must inhibit withdrawal of the painful extremity (Craig et al., 1996, Derbyshire and Jones, 1998, Derbyshire et al., 1997, Talbot et al., 1991), and in studies of induced itch in which participants are not allowed to scratch (Hsieh et al., 1994).

Second, neither the correlations of neural activity with symptom severity nor treatment-induced normalization of activity in OCD can definitively distinguish cause from effect of the illness. More severe urges to execute tics or compulsive behaviors will typically require stronger activation of regulatory mechanisms to suppress those behaviors. Therefore, correlations between the strength of activation in fronto-striatal circuits and ratings of symptom severity can simply indicate that patients with more severe symptoms have stronger demands for regulatory control in the experimental setting. Consequently, improvements in symptom severity following treatment will also reduce the demands for engaging regulatory control and, thus, will also reduce neural activation of regulatory circuits. Correlations between symptom severity and brain measures that represent secondary adaptations to symptoms are thus expected, as are correlations between changes in symptom severity over time and changes in brain measures secondary to those symptoms. For this reason, the cause-or-consequence dilemma affects cross-sectional and longitudinal designs equally. Brain changes that are detected longitudinally in close temporal proximity to the onset or exacerbation of symptoms can still represent either the core process underlying the progression of symptoms or a compensatory reaction to them.

Suggested solutions—To disentangle cause and effect, imaging studies may be extended to high-risk cohorts to identify “trait” or vulnerability markers that predispose individuals to mental illness. These studies compare individuals who are not ill, but who are nevertheless at high risk for developing an illness, with individuals who are at low risk for developing that illness. High-risk cohorts, such as the unaffected siblings or offspring of individuals affected with an illness, offer a unique opportunity to investigate vulnerability markers involved in the illness in the absence of overt clinical symptoms, and thus also without significant contamination from compensatory changes induced by the illness, the so-called “state” or illness markers. This approach can be further enhanced by prospectively following high-risk samples through time to identify pathophysiological processes associated with conversion from vulnerability to overt illness and to identify individuals who are resilient to the illness despite being at increased risk for it.

High-risk studies of developmental psychopathologies are now relatively common. For example, an MRI study investigated nonpsychotic, full siblings of an epidemiologically ascertained sample of probands with schizophrenia. These high-risk individuals had smaller

hippocampal volumes relative to healthy individuals without a family history of schizophrenia (Van Erp et al., 2002). Smaller hippocampal volumes and volume abnormalities in other cortical regions have been established as vulnerability markers for schizophrenia spectrum disorders in other high-risk samples defined on the basis of family history or in clinical samples based on the presence of prodromal symptoms (Fusar-Poli et al., 2011). Similar designs have been employed to study brain-based vulnerabilities for addiction. Healthy individuals at high familial risk for stimulant addiction, for example, had enlarged volumes of the amygdala and putamen relative to healthy individuals at low familial risk, a trait that was also present in addicted individuals (Ersche et al., 2012). These findings suggest that enlarged volumes of the amygdala and putamen are true markers of vulnerability to addiction and not simply the brain-based consequences of prior addiction.

A similar approach was used to disentangle brain-based vulnerability markers associated with schizophrenia from those associated with bipolar disorder (McDonald et al., 2004). The markers of vulnerability were sought in unaffected, first-degree relatives of patients in both diagnostic groups. Correlations were assessed between the degree of familial loading for illness, defined as a continuous measure of liability based on the affected status and genetic relatedness of family members, and measures of brain morphology in family members from each of the two diagnostic groups. The degree of familial loading for schizophrenia was associated with abnormalities in gray matter volume in bilateral frontal, striatal, thalamic, and lateral temporal regions, whereas familial loading for bipolar disorder was associated specifically with abnormalities in volumes of the anterior cingulate gyrus and ventral striatum.

These studies illustrate the potential value of high-risk designs in identifying brain-based markers of vulnerability for developing an illness, and therefore for disentangling a likely causal risk factor for the illness from the consequences of the illness. The designs of these prior studies, however, generally did not permit the further identification of resilience and illness markers in the same sample. An exemplary study design that permits distinction of these different markers has been implemented in a large, multi-generation cohort of families with and without major depressive disorder (Weissman et al., 2006). Offspring with and without familial history for depression, some of whom had a prior history of depressive illness and many who did not, were scanned to identify vulnerability endophenotypes that represented familial risk for depression independent of any personal history of depression. Morphometric analyses of the brain in this cohort demonstrated large expanses of cortical thinning across the lateral surface of the right cerebral hemisphere in persons at high familial risk compared to those at low familial risk (Peterson et al., 2009b), along with reduced white matter volumes in fronto-parietal regions (Dubin et al., 2012). Importantly, these risk effects were present in individuals who had never been ill and therefore could not be the consequences of prior illness or its treatment. The morphological abnormalities therefore represented a true endophenotype for familial depressive illness.

A subsequent fMRI study aimed to isolate further the functional markers of risk, resilience, and illness by capitalizing on the four subpopulations in this cohort defined by risk group and prior illness history: (1) those at low familial risk who were never ill, (2) those at high risk who were never ill, (3) those at low risk who had been ill, and (4) those at high risk who

had been ill (Peterson et al., in press). Vulnerability markers were defined as brain activations present significantly more in the high-risk group compared with the low-risk group. Resilience markers were defined as brain activations present in high-risk individuals who had never been ill but not in high-risk persons with lifetime illness, and that were more prominent in the high-risk than in the low-risk group. Markers of lifetime illness were defined as manifesting significantly more in persons with a lifetime history of depression than in persons without a lifetime history of illness. Based on these definitions, the risk marker included increased activation of insular, lateral prefrontal, and posterior cingulate cortices as participants performed a self-regulatory task. A resilience marker comprised increased activation of the dorsal anterior cingulate and dorsolateral prefrontal cortices. Markers of lifetime illness included exaggerated deactivation of default-mode circuits. Although the predictive value of these vulnerability and resilience markers remains to be tested in prospective follow-up of this cohort, the design of this study provides a novel framework to differentiate the neurobiological correlates of vulnerability, resilience, and illness effects.

Limitations of correlational designs: inability to identify causal mechanisms of illnesses and their treatments

MRI by itself cannot provide information about the causal relationship between brain imaging measures and behavior. Nearly all studies of psychopathologies and their treatments have yoked MRI measures to naturalistic, correlational designs, an understandable decision given that randomizing an individual to have a given illness or not is impossible. Nevertheless, because an imaging measure can only be correlated with a particular diagnosis, and because correlation does not prove causality, determining whether an imaging abnormality caused the disease, or whether the disease caused the imaging abnormality, or whether an unidentified third underlying variable caused both the MRI abnormality and the disease to yield the correlation of the imaging variable with the disease label, is unfortunately impossible using naturalistic designs. Yoking imaging to high risk designs can help to identify vulnerability markers, as we have already discussed. But high risk studies cannot definitively prove that the imaging abnormality caused the illness. This limited capacity for causal inference that naturalistic studies afford, however, is not unique to MRI; rather, it is common to all phenomenological and neurophysiological measures employed in naturalistic study designs.

Possible solutions—The gold standard method used to infer the causal effects of an independent variable on a second, dependent variable is by experimentally controlling and systematically varying the independent variable. A direct, systematic manipulation of the values of the independent variable in conjunction with an unbiased assessment of the effects that experimental manipulation of the independent variable has on the dependent variable allows investigators to interpret any observed relationships between the variables as causal. In medical research, this experimental control in the service of causal inference is best exemplified by the Randomized Controlled Trial (RCT), a study design that manipulates an independent variable, treatment, and then assesses the effects of that manipulation on a second variable, illness severity. The RCT randomly assigns treatment, such as active medication or placebo, to each participant in the study, thereby ensuring the absence of

ascertainment bias in determining which participant receives which intervention. The effects of that intervention on the outcome variable are then assessed blindly, with the participant and investigator both unaware of the treatment that the participant is receiving. At the completion of the trial, the random assignment and blind assessments allow the strong inference to be made that the medication *caused* the observed change in the outcome variable.

A similar strength of causal inference can be achieved when yoking imaging technologies to an RCT. Instead of the traditional outcome measure of change in symptom severity, however, the outcome measure in the yoked imaging study is a brain imaging measure. Then the randomization of treatment and the blind assessments of outcome allow the inference that the treatment caused the observed change in imaging measure from the beginning of the trial to its completion. We note, however, that even though the RCT supports the inference that the treatment caused the observed change in imaging measure, we cannot conclude that somehow the change in imaging measure tells us something about the cause of the illness being treated. We can conclude from an RCT, for example, that penicillin causes recovery from a streptococcal infection, but we cannot validly conclude that the illness was in any way caused be a deficiency in penicillin. The logic and truth of this statement may seem obvious, but psychiatric research has a long history of succumbing to this logical fallacy by assuming, for example, that because a medication with an affinity for a particular neurotransmitter receptor can reduce the severity of an illness, the illness is somehow a consequence of abnormal activity of that neurotransmitter or receptor.

Neuroimaging studies have increasingly adopted RCT designs to study the causal effects of therapeutic interventions on brain structure and function. An early example comes from an fMRI study in youth with Attention-Deficit/Hyperactivity Disorder (ADHD) who were scanned twice, on and off stimulation medication, in a randomized and counterbalanced order. Owing to this randomization of treatment status, the study demonstrated that stimulant medication improved the ability of patients to suppress activation in the so-called “default-mode” network, a brain system involved in self-referential processing and mind-wandering (Buckner et al., 2008), while they performed the Stroop task (Peterson et al., 2009a). Similarly, an fMRI study yoked to a randomized clinical trial of duloxetine versus placebo for chronic depression also demonstrated effects of medication on functioning of the default-mode network. A clinically effective regimen with the antidepressant diminished the abnormally strong density of connections in the default-mode network that patients with chronic depression displayed at baseline (Posner et al., 2013). Because dysfunction in the default-mode network may represent a marker of prior lifetime illness for major depressive disorder (Peterson et al., in press), normalization of that dysfunction suggests that duloxetine may reduce a functional brain abnormality that is the consequence of the illness. Some fMRI studies yoked to an RCT design additionally incorporate an arm of healthy controls receiving with the same treatment as the patients. For instance, an fMRI study assessed the effects of methylphenidate in both children with ADHD and healthy children using a crossover, counterbalanced design (Vaidya et al., 1998). A direct comparison of methylphenidate effects on patients versus controls revealed that the medication increased activation in the striatum in patients but reduced it in control children. A potential

shortcoming of this kind of direct comparison of patients versus healthy controls is that it cannot disentangle whether the differential effects of treatment on brain measures across groups represent either (1) true differences in treatment effects, (2) effects that derive from differing baseline characteristics across the groups (such as long-term effects of prior treatment on brain function in patients), or (3) effects associated with short-term changes in symptoms and the reduced need for compensatory responses in the patient but not in the control group. Recent studies of methylphenidate effects using fMRI have partially mitigated some of these confounds in interpretation by including only patients who are medication-naïve (Rubia et al., 2011a, Rubia et al., 2011b).

Although the yoking of MRI technologies to an RCT design provides powerful evidence that the intervention caused the changes in imaging measures over the course of a trial, the changes in imaging measures alone cannot identify what cognitive, emotional, or behavioral effects of the intervention the imaging changes represent. Pre-post changes in imaging measures over the course of an RCT lasting weeks, for example, can represent either the proximal effects of the intervention on the imaging measures or more distal, indirect influences on those measures that are mediated by treatment-induced changes in thoughts, emotions, or behaviors. Thus, assessing the short-term effects of randomized interventions on imaging measures, presumably before the intervention has fully induced all of its overt behavioral or clinical changes, may add valuable insight into the more direct and immediate effects of intervention. Some neuroimaging studies have therefore evaluated the effects of short-term interventions using RCT designs (Ersche et al., 2010). For example, two fMRI studies using a crossover RCT design demonstrated that a single dose of methylphenidate enhanced activations in inferior frontal cortex and striatum relative to placebo in medication-naïve children with ADHD when performing an inhibitory control task (Rubia et al., 2011a, Rubia et al., 2011b). Another randomized, crossover study showed that a single dose of methylphenidate induced in persons with cocaine addiction an immediate normalization of connections within the striatum (Konova et al., 2013) that are regarded as central to pathogenic addictive processes in animal models. Analyzing short-term changes in imaging measures in relation to longer-term clinical effects within RCT designs may afford path analyses using various consecutive time points (baseline, intermediate points, and outcome) to determine whether intermediate changes in imaging measures mediate the effects of an intervention on long-term clinical outcome, or whether longer-term clinical changes mediate the effects of the intervention on changes in imaging measures.

A second, related way potentially to sort out cause and effect in imaging studies may be the use of parallel designs that yoke imaging measures to RCTs of identical design in separate studies of humans and animals. Although examples of this parallel approach in animals and humans are rare, notable attempts in this direction have been made. For example, researchers used a parallel design with high-resolution, gadolinium-enhanced fMRI in mice and humans to assess the potential confounding effects of antipsychotic medications on their fMRI measures in persons with schizophrenia (Schobel et al., 2009). By imaging mice while they received either an antipsychotic medication or placebo, the investigators showed that antipsychotic medication was unlikely to induce the imaging abnormalities detected in the human patients. These same investigators used a similar parallel design coupled with measures of cerebral blood volume and morphology of the hippocampus to assess predictors

of psychosis in individuals who had prodromal symptoms of psychosis and in mice treated with ketamine, a pharmacological animal model for acute psychosis (Schobel et al., 2013). Elevated blood volume in specific subregions of the hippocampus predicted which individuals later developed schizophrenia, and it predicted a subsequent reduction of hippocampal volumes. Similar findings were detected in the ketamine-treated mice. The mouse arm permitted additional controlled manipulations to establish increased glutamate as the likely source of the increased perfusion and subsequent reduction in volumes of the hippocampus.

A third approach to studying causal mechanisms in the brain with MRI uses experimental manipulations to induce transient, reversible symptoms of disease in otherwise healthy adult humans and thereby show that the manipulated variable could plausibly produce the illness. One study, for example, capitalized on the well-established ketamine model of psychosis. By administering ketamine to healthy adults within a placebo-controlled RCT design, researchers were able to recapitulate a deficit in error-related signals in prefrontal cortex that had been observed previously in psychotic patients. Furthermore, the magnitude of the prefrontal deficit induced by ketamine correlated with the severity of psychotic-like experiences that the drug produced (Corlett et al., 2006). The use of this human model of psychosis was thus instrumental in establishing that deficient error signals are a plausible cause of psychotic symptoms. A similar approach to disentangling cause and effect can be implemented using the various non-invasive techniques of focal brain stimulation, such as transcranial magnetic stimulation (TMS), instead of pharmacological manipulations, to induce brain changes with experimental control. Sham-controlled TMS in combination with brain imaging, for example, has helped to establish a causal role for the prefrontal cortex in modulating visual processing and working memory (Zanto et al., 2011). TMS-based investigations implemented in the context of RCT designs could be used in analogous ways to induce focal activations of the brain and to establish their causal role in producing behavioral phenotypes or symptoms that resemble those seen in patients.

Failure to relate structure, function, connectivity, and metabolism

A main advantage of MRI over other imaging techniques is that it can provide information about multiple features of the brain, including morphological measures of brain structure, neurovascular measures of brain function, measures of structural connectivity, and measures of metabolite concentrations in brain tissue. Because each MRI modality offers unique information about brain structure or function, the incorporation of multiple informative modalities can provide more information about the neural bases of normal and abnormal behavior than can any single modality alone (Hao et al., 2013). Each MRI modality also has its own unique technological shortcomings. For example, diffusion tensor imaging at present is generally unable to determine the true orientation of nerve fiber bundles at points where fiber bundles cross, owing to a much lower spatial resolution of the imaging modality than the resolution of the crossing fibers (Jones et al., 2013). In addition, the spatial resolution of most magnetic resonance spectroscopy techniques as most commonly implemented is low, requiring measurement of metabolites in volumes of 8 cubic centimeters or more within a single voxel at a time (Capizzano et al., 2007), although recent advances in the acquisition of spectroscopy data permit the mapping of metabolites simultaneously throughout the brain

in voxels of 1 cubic millimeter or less (Hao et al., 2013), offering great promise for the more widespread use of spectroscopy in the future.

Structure and function are known to influence each other mutually both at the level of single neurons and at the level of neural systems (Lee et al., 2012). Structure can determine function, and vice versa. Imaging studies should examine the interrelations between brain structure and function to deepen our understanding of normal and pathological brain processes. Although few researchers would dispute the importance or validity of this imperative, only rarely have any studies directly assessed structure-function interrelations in the brain, and even fewer have assessed how those interrelations may be disrupted in pathological conditions.

In addition to identifying biologically meaningful interactions between brain structure and function, additional technical and interpretive considerations underscore the importance and value of using and integrating multiple imaging modalities in brain imaging studies. Structural anomalies, for instance, can introduce artifacts or systematic confounds into functional imaging measures (Meltzer et al., 1990). Systematic differences in brain structure between patients and controls, for example, can introduce errors into the spatial co-registration of functional images that produces a disproportionate loss of functional signals in one of the groups, which in turn can produce artificial differences in functional activations across groups. From this perspective, structure-function analyses are important to enhance the quality and interpretability of functional data. Another strength of multimodal imaging studies is their ability to help constrain the interpretations of findings in one modality based on findings in other modalities (for example, by showing that reduced volumes in a given cortical region accompanies lower concentrations of a specific metabolite in the same region), thereby improving the neurobiological validity of those findings and their interpretation.

In the past, different imaging modalities were typically collected in different samples of participants, which precluded investigations that integrated functional and structural information within individual participants. The acquisition of different modalities in different samples also made it difficult to relate structural and functional features at each point of the brain. Currently, however, most imaging studies acquire both structural and functional scans. The continuing reluctance to integrate information from these modalities into a database that permits statistical analyses across modalities likely relates to the technical difficulties of spatially co-registering both datasets into a single template and extracting information from each modality in a form that is amenable to statistical analysis at each point of the brain. In addition, reporting findings separately for each imaging modality is easier, more convenient, and more traditional than reporting analyses in one modality that account for the influences of measures from the other modality. Nonetheless, studies that combine structural and functional datasets and integrate their findings are underway and will become increasingly common in the future.

Suggested solutions—Addressing the failures to integrate information across imaging modalities and to assess the interrelations of structure and function is conceptually straightforward: collect data from multiple imaging modalities for each participant within

the same study and then integrate information from those modalities into a single template brain so as to permit the point-by-point correlation of measures from one modality with those from another modality. This approach is exemplified in a report showing that inter-individual differences in functional activation to simple sensory and motor tasks correlated with underlying variation in local volumes of brain tissue and with underlying concentrations of *N*-acetyl aspartate, a metabolite that indexes that density of viable neurons at that location (Hao et al., 2013). Substantial variability in local brain volumes and functional activation derived from the underlying variability in density of neurons in those regions. Many of these imaging measures in turn correlated strongly with measures of intellectual ability.

A handful of studies in adult clinical populations have reported structural and functional findings within a single study sample. A notable example identified anomalies in white-matter tracts, morphological measures, and functional activation during a working memory task, all of which converged in the medial prefrontal cortex of patients with schizophrenia (Pomarol-Clotet et al., 2010). Another study investigated the differences in brain structure and function between depressed and non-depressed individuals among sufferers of mild traumatic brain injury. Among other findings, the depressed individuals with brain injury showed greater activation of the amygdala and lower activation of the dorsolateral prefrontal cortex during an emotional processing task, in conjunction with lower measures of integrity of several white matter tracts (Matthews et al., 2011). Another study identified disturbances in white matter integrity in the presence of normal measures of gray matter morphology and functional connectivity in persons with late-onset major depressive disorder, suggesting a specific role for white-matter pathology in the pathogenesis of late-life depression (Sexton et al., 2012). Finally, a multimodal examination of alcohol-dependent patients demonstrated extensive brain atrophy in the frontal cortices of patients who subsequently relapsed compared with both those who remained abstinent and healthy controls (Beck et al., 2012), and in addition adjusted fMRI measures of activation to drug cues for these local differences in brain volume, thereby eliminating the volume abnormalities as a potential confound for the functional findings.

These examples attest to the value of multimodal imaging in establishing both the convergence of findings across modalities and the specificity of findings within each modality. These prior studies, however, were unable to realize the full scientific potential that multimodal techniques afford because their interrogation of data from the multiple modalities used relatively coarse imaging measures (such as the overall volume of brain regions) rather than point-wise correlations of measures across modalities throughout the brain. The studies therefore were unable to perform detailed, mechanistic investigation into the relationships of information acquired across imaging modalities. Cross-modal analyses of multiple MRI modalities, especially point-wise analyses of data from multiple modalities that are integrated into a common stereotaxic space, is now feasible and can provide complementary and mutually informative data about tissue organization that exceeds the sum of analyses within all modalities (Hao et al., 2013). We anticipate that application of this type of cross-modal analysis will soon prove useful in the study of mental illness, provided that it is accompanied by the use of suitable study designs.

Limitations of prior imaging approaches to diagnostic classification

Perhaps the most exciting and longest standing promise of neuroimaging in the clinical arena is that of improving diagnostic classifications in psychiatry. Neuroimaging has already helped to improve the diagnostic classification of neurological illnesses, such as including imaging markers as diagnostic criteria for Alzheimer's disease (Dubois et al., 2007). The use of imaging to aid diagnostic classification in psychiatry, however, remains elusive. One explanation for this disparity across disciplines is that neurological patients often have focal brain pathologies that are sufficiently apparent to the naked eye to make diagnosis possible by radiologists reading clinical scans, which is certainly not the case for psychiatric illnesses. Psychiatric pathologies are generally subtle and likely affect multiple, distributed neural systems throughout the brain. The identification of brain markers for such subtle, distributed anomalies is thus technologically challenging.

Many imaging studies have compared patients and controls using conventional, univariate statistics, such as the familiar two-sample t-test. This and similar tests compare the distributions of a summary measure indexing some brain feature across diagnostic groups to produce an estimate of the difference between the two distributions. Associated with this group difference is a probability value that represents how likely this difference will be observed under the null hypothesis that no real differences exists between the groups. Even when the difference estimate is sizeable and clearly beyond chance, the group distributions in imaging values usually overlap substantially across diagnostic groups. In other words, conventional comparisons assess how likely it is that the group distributions are truly different, not how much overlap exists between the distributions. However, having non-overlapping, or minimally overlapping, distributions is essential for accurate classification. This requirement means that even when a marked group difference in an imaging measure is identified, in most cases the identified “marker” will not be useful by itself in classifying persons into one of the groups on an individual basis; the sensitivity and specificity of this univariate measure will be low, and so it will have little clinical value.

Statistical analysis of brain imaging data has become increasingly sophisticated, moving from univariate analysis of single brain regions to mass-univariate analyses covering hundreds of thousands of voxels in the brain. More recently, following a similar trend in genetics, multivariate pattern classification analyses involving machine-learning and clustering techniques have been applied increasingly as a means to exploit the massive amounts of correlated data in neuroimaging datasets (Haubold et al., 2012). Rather than evading the complexities and abundance of information contained in imaging data, these methods embrace multiple features of the dataset in the service of enhancing classification accuracy. Because these powerful machine-learning techniques are capable of identifying subtle, distributed patterns of brain anomalies, their application to psychiatry is spurring a renewed interest in and hope for identifying imaging biomarkers for psychiatric classification. Some of these techniques additionally permit automated detection of biological subgroupings among individuals within a diagnostic category, and presumably those subtypes will provide improved prediction of the future course of illness and treatment outcome, thereby supporting the individualized selection of more effective treatments.

Multivariate classification techniques, however, are subject to the same design considerations that apply to any other imaging tool. These sophisticated technologies can produce classification algorithms that improve patient care only in combination with appropriate study designs and when applied to clinically relevant questions. If classification algorithms are to generalize to the larger population of persons with the classified illness, investigators need to ensure that these classification techniques are not simply detecting the brain-based, epiphenomenal consequences of prior illness or treatment, and that instead the classification algorithms are operating on brain markers that are central to the pathogenesis of illness. If, for instance, a multivariate approach is highly accurate in classifying patients with schizophrenia and healthy controls by detecting the neuroplastic changes induced by the antipsychotic medications used universally to treat patients with schizophrenia, the classification will likely have limited application in clinical practice (other than perhaps determining the degree of compliance with the medication), because clinicians would be unable to apply this algorithm in diagnosing an individual who has an unclear diagnosis and who has never been treated with antipsychotic medication. They would also likely have difficulty differentiating a schizophrenic individual treated with antipsychotic medication from a person with bipolar disorder or psychotic depression who has also been treated with antipsychotic medication. Indeed, some evidence suggests that multivariate classification algorithms are sensitive to the distributed changes in brain function that specific medications induce (Marquand et al., 2012). In addition, the multivariate prediction of future clinical course or response to treatment would be particularly useful in informing clinical decisions directly and would potentially prevent delays in the initiation of effective interventions. A multivariate approach that predicts, for instance, which treatment-naïve, depressed adolescents will develop a bipolar versus a unipolar illness could have immediate and direct clinical utility by aiding clinicians in the selection of the most appropriate treatment to prevent a manic switch.

Suggested solutions—The statistical and machine-learning tools that make neuroimaging-based diagnosis possible are already at our disposal. The combination of these tools with suitable study designs may be able to produce diagnostic and prognostic classification algorithms that potentially can revolutionize clinical practice.

Although the technical details of the various approaches to multivariate pattern classification are beyond the scope of this review, the general principles of these approaches and their applications to clinical classification in developmental and adult psychopathologies are well described elsewhere (Bray et al., 2009, Haubold et al., 2012, Kloppel et al., 2012, Norman et al., 2006, Orru et al., 2012, Pereira et al., 2009). Suffice to say that multivariate pattern classification techniques, such as support vector machines, use multiple features of the data (such as signal intensity at multiple voxels) to generate a high-dimensional feature space wherein measures from two or more diagnostic groups of participants minimally overlap and are thus readily separable. In general, these methods require partitioning of the data into a training set, which develops and defines the classification rules to discriminate the diagnostic groups, and a test set, which assesses the accuracy of diagnosis in a set of brains that were not involved in generating the classification algorithm. The sensitivity and specificity of the classifier for diagnosis can be obtained easily using the test set.

Numerous published studies demonstrate the promise of multivariate pattern analysis for psychiatric diagnosis. Not only do these techniques allow for a distinction between patients and controls, but in many instances they do so with impressive accuracy. An early example is an fMRI study of brain connectivity during an auditory detection task (Calhoun et al., 2008). Based on images corresponding to connectivity within the auditory and default-mode networks for each participant, an algorithm was developed to differentiate between healthy controls and two groups of treated patients who had either schizophrenia or bipolar disorder. The algorithm not only accurately separated each group of patients from controls, but also discriminated the two groups of patients from one another with sensitivities and specificities near 90%. Linear discriminant analysis of resting-state brain function was similarly successful in differentiating a small sample of youth with ADHD from an age-matched control group (Zhu et al., 2008).

Although these latter two studies used fMRI data for diagnostic classification, multivariate classification techniques can also be implemented using anatomical MR images. Automated morphometry of anatomical MRIs in combination with machine-learning classified treated patients with major depressive disorder from matched controls with accuracies of approximately 90% (Mwangi et al., 2012). Furthermore, the multivariate patterns identified by the classification algorithm correlated strongly with ratings of illness severity, confirming that these abnormal patterns related meaningfully to clinical features of the illness. Another study used a clustering algorithm based on the spatial patterns of variation across the morphological surfaces of numerous cortical and subcortical brain regions to diagnose, with astounding accuracy, large numbers of youth and adults who were affected by a wide range of illnesses, including ADHD, Tourette Syndrome, Schizophrenia, Bipolar Disorder, and familial Major Depressive Disorder (Bansal et al., 2012). This approach attained nearly perfect values (95-100%) for sensitivity and specificity in discriminating not only the various groups of patients from matched controls, but also in discriminating between schizophrenia and bipolar disorder in adults and between Tourette Syndrome and ADHD in children. It was even able to discriminate individuals at high familial risk for developing MDD from persons at low familial risk, including persons who had never been ill previously with depression. That it could do so shows that these techniques were not simply classifying individuals based on the epiphenomenal consequences of prior illness or treatment, and it demonstrates the feasibility of identifying people at high risk for developing illness who can be targeted for preventive interventions.

An attractive feature of multivariate classifiers is their ability to combine various features of an imaging dataset, even when those features do not derive from a single imaging modality. Although previous classification studies using multiple features from a single imaging modality (only structural or only functional MRI, for example) can achieve remarkable diagnostic accuracies, the combination of several imaging modalities may further improve diagnostic accuracy in situations where group distinctions are particularly challenging. Evidence thus far suggests that multivariate classifiers combining functional data and measures of white matter integrity can achieve better results than classifiers based on a single modality (Colby et al., 2012, Sui et al., 2011). One study showed that a decision-tree algorithm that combined multivariate information from brain activations while performing three different cognitive tasks yielded superior diagnostic accuracy for depression compared

with algorithms based solely on activation from one of these cognitive processes (Hahn et al., 2011). These early demonstrations speak to the immense clinical potential of multivariate pattern classifications and their flexibility in combining diverse datasets to inform diagnostic classification.

Although the above studies highlight the successes of multivariate approaches to diagnostic classification, not all efforts in this area have been as successful. The unsatisfying results from a competition launched by the ADHD-200 Consortium highlight the challenges of multivariate classification for psychiatric diagnosis, in particular fMRI studies for the diagnostic classification of youth with ADHD (Consortium, 2012, Eloyan et al., 2012, Lim et al., 2013). Based on large, multi-site resting-state fMRI and anatomical datasets, the various approaches proposed by different teams achieved accuracies ranging from 55% to 78%, and even lower accuracies (up to 61%) in a second validation using an external dataset, results that were attributed in part to the lack of standardization in data collection across sites. In any case, these unsatisfying results underscore the fact that imaging-based diagnosis is an emerging technology and that the promising results reported in many studies await confirmation in future studies.

We have presented examples of imaging biomarkers that can yield comparable or even better diagnostic accuracy than tests that are commonly prescribed in medical settings, such as fasting plasma glucose for diagnosis of type 2 diabetes. Why then have these imaging biomarkers not yet begun to make their way into clinical practice? Certainly more research is needed to document the improved performance and robustness of the algorithms and to establish which disorders can be considered suitable candidates for imaging-based diagnosis in real-world clinical settings. Most of these studies, for example, have thus far focused on cases of chronic, treated patients, in which diagnosis is not commonly questioned. The methods must be shown to work well for newly presenting cases and in persons for whom the clinical diagnosis is uncertain. Moreover, some studies have discriminated clinical groups that are rarely considered in a differential diagnosis, such as distinguishing persons with schizophrenia from those with Tourette syndrome. The techniques must be shown to produce accurate diagnoses in persons for whom the diagnosis is clinically relevant, such as when discriminating schizophrenia from schizoaffective or bipolar illnesses. These methods should also be shown to improve diagnostic classifications or predictions of clinical course or treatment response and therefore to enhance current standards of care compared with other, already available, clinical tools (Brown et al., 2012). Nevertheless, the fact that these highly accurate diagnostic discriminations were performed in a completely automated manner, without any human assistance, their application to these diagnostic categories are still highly informative of the clinical promise that these approaches hold.

Several studies have already used multivariate classification to address questions that are directly relevant for clinical practice, questions such as which individuals with prodromal symptoms of psychosis will develop a full-fledged psychotic disorder. One longitudinal study of persons at risk for psychosis, for example, used a machine-learning algorithm to identify morphological markers that predicted the transition from a prodromal state to overt psychosis (Koutsouleris et al., 2009). It showed that an algorithm, using a baseline anatomical MRI scan, predicted with 88% accuracy which individuals would transition to

psychosis after 4 years of clinical follow-up. The prediction in this study of a future diagnosis using a baseline scan approximates the anticipated future clinical use of brain imaging for diagnosis in which a brain image in a situation of clinical uncertainty would assign a diagnosis that is predicted to emerge as the clinical picture becomes clearer over time. Another study addressed the similarly relevant issue of differential diagnosis between unipolar and bipolar depression. A multivariate classifier based on fMRI responses to facial emotions reached 90% accuracy in discriminating bipolar and unipolar patients in an acute depressive episode of comparable severity (Grotegerd et al., 2013).

Some studies have also suggested that multivariate classifiers are capable of differentiating among two or more childhood-onset developmental disorders. One study developed a machine-learning classifier based on volumetric and geometric features across the cortical surface that was able to discriminate accurately (with 90% sensitivity and 80% specificity) between adults with autism spectrum disorder and healthy controls (Ecker et al., 2010). The same classifier further identified ill controls with ADHD as not belonging to the autism group with an accuracy of nearly 80%, demonstrating disease-specificity in the classification. Another study also demonstrated disease-specificity by showing high classification accuracy (85%) in discriminating between children with autistic spectrum disorders and ADHD based on volumetric measures of gray matter (Lim et al., 2013). Finally, multivariate classification may be useful in predicting future clinical course. A multivariate classifier applied to anatomical MRI scans at baseline was able to predict, with accuracies up to 70%, subsequent response to antidepressant medication in medication-naïve patients with depression (Gong et al., 2011).

One additional reason for the delays in clinical implementation and dissemination of multivariate classifiers is that most studies using classification algorithms are still lacking independent replications that attest to the robustness of classification accuracy on independent, representative samples of patients and healthy controls across different sites and different demographic populations. Large-scale validations such as these are needed to confirm that the sensitivity and specificity of an algorithm, originally established in one population, are comparable to those when the algorithm is applied to persons in independent populations, thereby justifying the widespread use of the algorithm in clinical settings. These confirmatory studies should additionally examine the predictive value of the algorithm—i.e., the probability in a given population that a positive classification indicates the presence of an illness or that a negative classification indicates its absence. Predictive value depends not only on the sensitivity and specificity of the test, but also on the prevalence of the illness in the population to which the test is applied. Therefore, confirmatory studies should include samples in which the percentage of patients relative to healthy individuals, or the percentage of patients in one diagnostic category relative to patients in a second diagnostic category, reflects the prevalence of the illness or illnesses in the population in which the algorithm will be applied (although stepwise algorithms may be capable of adjusting for prior probability of illness depending on the setting) (Hahn et al., 2011). Note that populations of interest in confirmatory studies will likely comprise individuals at high risk for illness or symptomatic individuals who have an uncertain diagnosis (i.e., cases in which prognostic information or diagnosis clarification is most valuable), rather than in the general population, given that applications of imaging-based

diagnostic classification for population screening are unlikely in the foreseeable future. Only after these confirmatory studies produce reliable information about the predictive value of imaging-based diagnostic algorithms for a target population, will clinicians be able to apply these tools safely and to interpret their results with the greatest possible confidence.

Conclusions

Despite the many challenges for and limitations of MRI studies thus far in identifying causal processes and compensatory responses in developmental psychopathologies, neuroimaging is proving an invaluable tool that has in many ways already advanced our understanding of brain development in health and illness. An earlier era of predominantly cross-sectional, case-control designs that recruited nonrepresentative participant samples of convenience is beginning to give way to a new wave of methodologically more rigorous studies that include longitudinal, high-risk, epidemiological, and randomized controlled trial designs. Yoking imaging to these study designs will produce more valid conclusions about diagnosis and the mechanisms of pathogenesis and treatment. Likewise, the integration of imaging findings across imaging modalities is beginning to clarify mechanisms of illness by bringing together the disparate pieces of information that imaging modalities in isolation have thus far generated. Multivariate techniques of classification, if coupled with clinically meaningful designs, may soon produce classification algorithms with sufficient diagnostic and prognostic value for use in clinical settings.

We have discussed common limitations that prevent MRI studies of developmental psychopathologies from attaining their ultimate aim, that of identifying abnormalities in brain mechanisms and neurodevelopmental trajectories that are centrally involved in the pathogenesis of mental illnesses in youth. Deviations from the normal trajectories of brain development may be a relevant feature of many or all mental illnesses and, in our opinion, perhaps the defining feature of most. Prospective, longitudinal studies that acquire MRI measures repeatedly through time in reasonably representative samples of participants are absolutely necessary to identify these developmental trajectories correctly. Naturalistic, longitudinal studies of already-affected participants, however, are not alone sufficient to infer the causes of developmental psychopathologies, because those study designs provide only statistical associations between changes in imaging measures and changes in some feature of the illness—and associations do not prove causation. Causal inferences are particularly problematic in studies of already-ill people, in whom imaging abnormalities may represent either the cause of the illness or its consequences. Those consequences include brain-based compensatory responses, down-stream epiphenomenal effects of the illness, or the effects of prior treatment.

To establish stronger causal relationships between imaging measures and developmental psychopathologies, imaging studies should image participants either in the prodromal phase of illness or, ideally, in high-risk samples even before symptoms begin. Yoking imaging technologies to high risk study designs can help to disentangle brain markers for vulnerability from those that confer resilience and from others that represent the consequences of prior illness. Identifying vulnerability markers will have important therapeutic implications, as it will help to identify vulnerable individuals who could benefit

from preventive interventions. Likewise, identifying resilience markers will reveal new treatment targets for enhancing neural resilience.

Combining the use of imaging technologies with paradigms for symptom provocation can also help to identify causal pathways to illness, as well as help to disentangle the causes of an illness from its consequences. Conducting parallel imaging studies in patients and in animal models, including RCTs of pharmacological interventions, will increasingly help to identify the cellular and molecular determinants of the imaging correlates of the disease or its treatments. Yoking imaging studies to RCTs combines the controlled manipulation of an experimental variable—randomized assignment to treatment—with a blind assessment of outcome, which in this instance includes a change in imaging measures, to provide the strongest possible basis for causal inferences regarding the effects of the therapeutic intervention, whether that intervention is pharmacological, psychological, or behavioral. We have provided examples of studies that have yoked MRI measures to randomized controlled trials to begin identifying causal mechanisms for the effects of treatments on the brain. This approach to studying treatment effects will be used increasingly in the future to identify the brain-based mediators of therapeutic change in the symptoms of a disease.

Above all, imaging findings have little value if they cannot be extrapolated to the general population of ill and healthy people that participants in the study are supposed to represent. Thus, future imaging studies of child- and adult-onset developmental psychopathologies should and will move towards recruiting more representative samples of participants, including populations that are ascertained epidemiologically. The recruitment of more representative samples will benefit imaging studies of any design, whether they are case-control, prospective longitudinal, high risk, or treatment studies, by improving the reproducibility and generalizability of their findings.

One of the major breakthroughs for brain imaging studies of psychiatric illnesses in the near future will be the clinical application and dissemination of multivariate classification algorithms that produce readouts of MRI scans that accurately classify patients into valid diagnostic groups. Objective indicators and predictors of mental illness are thus emerging that may be incorporated into real-world clinical practice in the near future, particularly when ongoing research demonstrates that these approaches can address clinically relevant questions in participant samples that are representative of the clinical populations from which they are drawn. Although conducting these large-scale studies undoubtedly will be challenging, they have the invaluable potential to transform diagnosis and treatment planning in clinical practice.

Despite the substantial limitations of the findings from imaging studies of developmental psychopathologies thus far, imaging studies have made and will continue to make important contributions to our understanding of the causes and consequences of psychopathologies in the developing and the adult brain. A more complete understanding of the mechanisms that underlie these illnesses and their treatments will come from the integration of multiple research disciplines, study designs, and investigative tools. Among those tools, MRI holds a privileged position in that it provides safe, *in vivo* visualization of the brain. When used in combination with sound study designs, imaging technologies will ultimately produce more

valid and more generalizable findings that will inform us about the brain-based mechanisms that cause mental illnesses and that produce meaningful therapeutic responses.

Acknowledgments

This work was supported in part by: NIMH grants 1P50MH090966, MH36197, MH093677, 2T32 MH16434, K01-MH077652; NIEHS grant ES015579; NIDA grant DA027100; a grant from the National Alliance for Research on Schizophrenia and Depression (NARSAD); and a fellowship from the Italian Academy for Advanced Studies in America, Columbia University. This review article was invited by the journal, for which the authors have been offered a small honorarium payment towards personal expenses; it has undergone full, external peer review.

REFERENCES

Ahmed F, Ras J, Seedat S. Volumetric structural magnetic resonance imaging findings in pediatric posttraumatic stress disorder and obsessive compulsive disorder: a systematic review. Front Psychol. 2012; 3:568. [PubMed: 23272001]

Almli CR, Rivkin MJ, McKinstry RC, Brain Development Cooperative, G. The NIH MRI study of normal brain development (Objective-2): newborns, infants, toddlers, and preschoolers. Neuroimage. 2007; 35:308–325. [PubMed: 17239623]

Arnone D, Cavanagh J, Gerber D, Lawrie SM, Ebmeier KP, McIntosh AM. Magnetic resonance imaging studies in bipolar disorder and schizophrenia: meta-analysis. Br J Psychiatry. 2009; 195:194–201. [PubMed: 19721106]

Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J, Weissman M, Peterson BS. Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. PLoS One. 2012; 7:e50698. [PubMed: 23236384]

Baxter LR Jr. Phelps ME, Mazziotta JC, Guze BH, Schwartz JM, Selin CE. Local cerebral glucose metabolic rates in obsessive-compulsive disorder. A comparison with rates in unipolar depression and in normal controls. Arch Gen Psychiatry. 1987; 44:211–218. [PubMed: 3493749]

Baxter LR Jr. Schwartz JM, Bergman KS, Szuba MP, Guze BH, Mazziotta JC, Alazraki A, Selin CE, Ferng HK, Munford P, et al. Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. Arch Gen Psychiatry. 1992; 49:681–689. [PubMed: 1514872]

Baxter LR Jr. Schwartz JM, Mazziotta JC, Phelps ME, Pahl JJ, Guze BH, Fairbanks L. Cerebral glucose metabolic rates in nondepressed patients with obsessive-compulsive disorder. Am J Psychiatry. 1988; 145:1560–1563. [PubMed: 3264118]

Beck A, Wustenberg T, Genauck A, Wrase J, Schlagenhauf F, Smolka MN, Mann K, Heinz A. Effect of brain structure, brain function, and brain connectivity on relapse in alcohol-dependent patients. Arch Gen Psychiatry. 2012; 69:842–852. [PubMed: 22868938]

Benkelfat C, Nordahl TE, Semple WE, King AC, Murphy DL, Cohen RM. Local cerebral glucose metabolic rates in obsessive-compulsive disorder. Patients treated with clomipramine. Arch Gen Psychiatry. 1990; 47:840–848. [PubMed: 2393342]

Bray S, Chang C, Hoeft F. Applications of multivariate pattern classification analyses in developmental neuroimaging of healthy and clinical populations. Front Hum Neurosci. 2009; 3:32. [PubMed: 19893761]

Breiter HC, Rauch SL, Kwong KK, Baker JR, Weisskoff RM, Kennedy DN, Kendrick AD, Davis TL, Jiang A, Cohen MS, Stern CE, Belliveau JW, Baer L, O'Sullivan RL, Savage CR, Jenike MA, Rosen BR. Functional magnetic resonance imaging of symptom provocation in obsessive-compulsive disorder. Arch Gen Psychiatry. 1996; 53:595–606. [PubMed: 8660126]

Brown MR, Sidhu GS, Greiner R, Asgarian N, Bastani M, Silverstone PH, Greenshaw AJ, Dursun SM. ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. Front Syst Neurosci. 2012; 6:69. [PubMed: 23060754]

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network: anatomy, function, and relevance to disease. Ann N Y Acad Sci. 2008; 1124:1–38. [PubMed: 18400922]

Bush G, Whalen PJ, Rosen BR, Jenike MA, McInerney SC, Rauch SL. The counting Stroop: an interference task specialized for functional neuroimaging--validation study with functional MRI. Hum Brain Mapp. 1998; 6:270–282. [PubMed: 9704265]

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafo MR. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci. 2013; 14:365–376. [PubMed: 23571845]

Calhoun VD, Maciejewski PK, Pearlson GD, Kiehl KA. Temporal lobe and “default” hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. Hum Brain Mapp. 2008; 29:1265–1275. [PubMed: 17894392]

Capizzano AA, Jorge RE, Acion LC, Robinson RG. In vivo proton magnetic resonance spectroscopy in patients with mood disorders: a technically oriented review. J Magn Reson Imaging. 2007; 26:1378–1389. [PubMed: 17968885]

Carter CS, Mintun M, Cohen JD. Interference and facilitation effects during selective attention: an H215O PET study of Stroop task performance. Neuroimage. 1995; 2:264–272. [PubMed: 9343611]

Chakrabarti S, Fombonne E. Pervasive developmental disorders in preschool children: confirmation of high prevalence. Am J Psychiatry. 2005; 162:1133–1141. [PubMed: 15930062]

Chua SE, McKenna PJ. Schizophrenia--a brain disease? A critical review of structural and functional cerebral abnormality in the disorder. Br J Psychiatry. 1995; 166:563–582. [PubMed: 7620741]

Colby JB, Rudie JD, Brown JA, Douglas PK, Cohen MS, Shehzad Z. Insights into multimodal imaging classification of ADHD. Front Syst Neurosci. 2012; 6:59. [PubMed: 22912605]

Consortium HD. The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. Front Syst Neurosci. 2012; 6:62. [PubMed: 22973200]

Corbett BA, Carmean V, Ravizza S, Wendelken C, Henry ML, Carter C, Rivera SM. A functional and structural study of emotion and face processing in children with autism. Psychiatry Res. 2009; 173:196–205. [PubMed: 19665877]

Corlett PR, Honey GD, Aitken MR, Dickinson A, Shanks DR, Absalom AR, Lee M, Pomarol-Clotet E, Murray GK, McKenna PJ, Robbins TW, Bullmore ET, Fletcher PC. Frontal responses during learning predict vulnerability to the psychotogenic effects of ketamine: linking cognition, brain activity, and psychosis. Arch Gen Psychiatry. 2006; 63:611–621. [PubMed: 16754834]

Cortese S, Kelly C, Chabernaud C, Proal E, Di Martino A, Milham MP, Castellanos FX. Toward systems neuroscience of ADHD: a meta-analysis of 55 fMRI studies. Am J Psychiatry. 2012; 169:1038–1055. [PubMed: 22983386]

Courchesne E, Yeung-Courchesne R, Press GA, Hesselink JR, Jernigan TL. Hypoplasia of cerebellar vermal lobules VI and VII in autism. N Engl J Med. 1988; 318:1349–1354. [PubMed: 3367935]

Craig AD, Reiman EM, Evans A, Bushnell MC. Functional imaging of an illusion of pain. Nature. 1996; 384:258–260. [PubMed: 8918874]

Derbyshire SW, Jones AK. Cerebral responses to a continual tonic pain stimulus measured using positron emission tomography. Pain. 1998; 76:127–135. [PubMed: 9696465]

Derbyshire SW, Jones AK, Gyulai F, Clark S, Townsend D, Firestone LL. Pain processing during three levels of noxious stimulation produces differential patterns of central activity. Pain. 1997; 73:431–445. [PubMed: 9469535]

Dickstein DP, Pescosolido MF, Reidy BL, Galvan T, Kim KL, Seymour KE, Laird AR, Di Martino A, Barrett RP. Developmental meta-analysis of the functional neural correlates of autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2013; 52:279–289. e216. [PubMed: 23452684]

Dubin M, Weissman M, Xu D, Bansal R, Hao X, Liu J, Warner V, Peterson B. Identification of a circuit-based endophenotype for familial depression. Psychiatry Res. 2012; 201:175–181. [PubMed: 22516664]

Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol. 2007; 6:734–746. [PubMed: 17616482]

Duerden EG, Mak-Fan KM, Taylor MJ, Roberts SW. Regional differences in grey and white matter in children and adults with autism spectrum disorders: an activation likelihood estimate (ALE) meta-analysis. Autism Res. 2012; 5:49–66. [PubMed: 22139976]

Ecker C, Marquand A, Mourao-Miranda J, Johnston P, Daly EM, Brammer MJ, Maltezos S, Murphy CM, Robertson D, Williams SC, Murphy DG. Describing the brain in autism in five dimensions--magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach. J Neurosci. 2010; 30:10612–10623. [PubMed: 20702694]

Eloyan A, Muschelli J, Nebel MB, Liu H, Han F, Zhao T, Barber AD, Joel S, Pekar JJ, Mostofsky SH, Caffo B. Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. Front Syst Neurosci. 2012; 6:61. [PubMed: 22969709]

Ersche KD, Bullmore ET, Craig KJ, Shabbir SS, Abbott S, Muller U, Ooi C, Suckling J, Barnes A, Sahakian BJ, Merlo-Pich EV, Robbins TW. Influence of compulsivity of drug abuse on dopaminergic modulation of attentional bias in stimulant dependence. Arch Gen Psychiatry. 2010; 67:632–644. [PubMed: 20530013]

Ersche KD, Jones PS, Williams GB, Turton AJ, Robbins TW, Bullmore ET. Abnormal brain structure implicated in stimulant drug addiction. Science. 2012; 335:601–604. [PubMed: 22301321]

Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL, Brain Development Cooperative, G. Unbiased average age-appropriate atlases for pediatric studies. Neuroimage. 2011; 54:313–327. [PubMed: 20656036]

Fusar-Poli P, Borgwardt S, Crescini A, Deste G, Kempton MJ, Lawrie S, Mc Guire P, Sacchetti E. Neuroanatomy of vulnerability to psychosis: a voxel-based meta-analysis. Neurosci Biobehav Rev. 2011; 35:1175–1185. [PubMed: 21168439]

Gogtay N. Cortical brain development in schizophrenia: insights from neuroimaging studies in childhood-onset schizophrenia. Schizophr Bull. 2008; 34:30–36. [PubMed: 17906336]

Gogtay N, Greenstein D, Lenane M, Clasen L, Sharp W, Gochman P, Butler P, Evans A, Rapoport J. Cortical brain development in nonpsychotic siblings of patients with childhood-onset schizophrenia. Arch Gen Psychiatry. 2007; 64:772–780. [PubMed: 17606811]

Gong Q, Wu Q, Scarpazza C, Lui S, Jia Z, Marquand A, Huang X, McGuire P, Mechelli A. Prognostic prediction of therapeutic response in depression using high-field MR imaging. Neuroimage. 2011; 55:1497–1503. [PubMed: 21134472]

Grotegerd D, Suslow T, Bauer J, Ohrmann P, Arolt V, Stuhrmann A, Heindel W, Kugel H, Dannlowski U. Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study. Eur Arch Psychiatry Clin Neurosci. 2013; 263:119–131. [PubMed: 22639242]

Hahn T, Marquand AF, Ehlis AC, Dresler T, Kittel-Schneider S, Jarczok TA, Lesch KP, Jakob PM, Mourao-Miranda J, Brammer MJ, Fallgatter AJ. Integrating neurobiological markers of depression. Arch Gen Psychiatry. 2011; 68:361–368. [PubMed: 21135315]

Hao X, Xu D, Bansal R, Dong Z, Liu J, Wang Z, Kangarlu A, Liu F, Duan Y, Shova S, Gerber AJ, Peterson BS. Multimodal magnetic resonance imaging: The coordinated use of multiple, mutually informative probes to understand brain structure and function. Hum Brain Mapp. 2013; 34:253–271. [PubMed: 22076792]

Hardan AY, Minshew NJ, Harenski K, Keshavan MS. Posterior fossa magnetic resonance imaging in autism. J Am Acad Child Adolesc Psychiatry. 2001; 40:666–672. [PubMed: 11392344]

Harezlak J, Ryan LM, Giedd JN, Lange N. Individual and population penalized regression splines for accelerated longitudinal designs. Biometrics. 2005; 61:1037–1048. [PubMed: 16401277]

Hart H, Radua J, Nakao T, Mataix-Cols D, Rubia K. Meta-analysis of functional magnetic resonance imaging studies of inhibition and attention in attention-deficit/hyperactivity disorder: exploring task-specific, stimulant medication, and age effects. JAMA Psychiatry. 2013; 70:185–198. [PubMed: 23247506]

Harth SC, Johnstone RR, Thong YH. The psychological profile of parents who volunteer their children for clinical research: a controlled study. J Med Ethics. 1992; 18:86–93. [PubMed: 1619628]

Harth SC, Thong YH. Sociodemographic and motivational characteristics of parents who volunteer their children for clinical research: a controlled study. BMJ. 1990; 300:1372–1375. [PubMed: 2372582]

Hashimoto T, Tayama M, Murakawa K, Yoshimoto T, Miyazaki M, Harada M, Kuroda Y. Development of the brainstem and cerebellum in autistic patients. J Autism Dev Disord. 1995; 25:1–18. [PubMed: 7608030]

Haubold A, Peterson BS, Bansal R. Annual Research Review: Progress in using brain morphometry as a clinical tool for diagnosing psychiatric disorders. Journal of Child Psychology and Psychiatry. 2012; 53:519–535. [PubMed: 22394424]

Hazlett HC, Poe MD, Gerig G, Styner M, Chappell C, Smith RG, Vachet C, Piven J. Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. Arch Gen Psychiatry. 2011; 68:467–476. [PubMed: 21536976]

Haznedar MM, Buchsbaum MS, Wei TC, Hof PR, Cartwright C, Bienstock CA, Hollander E. Limbic circuitry in patients with autism spectrum disorders studied with positron emission tomography and magnetic resonance imaging. Am J Psychiatry. 2000; 157:1994–2001. [PubMed: 11097966]

Hirai T, Jones EG. A new parcellation of the human thalamus on the basis of histochemical staining. Brain Res Brain Res Rev. 1989; 14:1–34. [PubMed: 2720229]

Hoberman A, Shaikh N, Bhatnagar S, Haralam MA, Kearney DH, Colborn DK, Kienholz ML, Wang L, Bunker CH, Keren R, Carpenter MA, Greenfield SP, Pohl HG, Mathews R, Moxey-Mims M, Chesney RW. Factors That Influence Parental Decisions to Participate in Clinical Research: Consenters vs Nonconsenters. JAMA Pediatr. 2013:1–6. [PubMed: 23460048]

Horga G, Maia TV, Wang P, Wang Z, Marsh R, Peterson BS. Adaptation to conflict via context-driven anticipatory signals in the dorsomedial prefrontal cortex. J Neurosci. 2011; 31:16208–16216. [PubMed: 22072672]

Hsieh JC, Hagermark O, Stahle-Backdahl M, Ericson K, Eriksson L, Stone-Elander S, Ingvar M. Urge to scratch represented in the human cerebral cortex during itch. J Neurophysiol. 1994; 72:3004–3008. [PubMed: 7897505]

Hulvershorn LA, Cullen K, Anand A. Toward dysfunctional connectivity: a review of neuroimaging findings in pediatric major depressive disorder. Brain Imaging Behav. 2011; 5:307–328. [PubMed: 21901425]

Jones DK, Knosche TR, Turner R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. Neuroimage. 2013; 73:239–254. [PubMed: 22846632]

Karatsoreos IN, McEwen BS. Annual Research Review: The neurobiology and physiology of resilience and adaptation across the life course. J Child Psychol Psychiatry. 2013; 54:337–347. [PubMed: 23517425]

Kerns JG, Cohen JD, MacDonald AW 3rd, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. Science. 2004; 303:1023–1026. [PubMed: 14963333]

Kim JE, Lyoo IK, Estes AM, Renshaw PF, Shaw DW, Friedman SD, Kim DJ, Yoon SJ, Hwang J, Dager SR. Laterobasal amygdalar enlargement in 6- to 7-year-old children with autism spectrum disorder. Arch Gen Psychiatry. 2010; 67:1187–1197. [PubMed: 21041620]

Kleiman MD, Neff S, Rosman NP. The brain in infantile autism: are posterior fossa structures abnormal? Neurology. 1992; 42:753–760. [PubMed: 1565227]

Kloppel S, Abdulkadir A, Jack CR Jr. Koutsouleris N, Mourao-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. Neuroimage. 2012; 61:457–463. [PubMed: 22094642]

Konova AB, Moeller SJ, Tomasi D, Volkow ND, Goldstein RZ. Effects of methylphenidate on resting-state functional connectivity of the mesocorticolimbic dopamine pathways in cocaine addiction. JAMA Psychiatry. 2013; 70:857–868. [PubMed: 23803700]

Koutsouleris N, Meisenzahl EM, Davatzikos C, Bottlender R, Frodl T, Scheuerecker J, Schmitt G, Zetzsche T, Decker P, Reiser M, Moller HJ, Gaser C. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch Gen Psychiatry. 2009; 66:700–712. [PubMed: 19581561]

Kraemer HC, Yesavage JA, Taylor JL, Kupfer D. How can we learn about developmental processes from cross-sectional studies, or can we? Am J Psychiatry. 2000; 157:163–171. [PubMed: 10671382]

Lazaro L, Caldu X, Junque C, Bargallo N, Andres S, Morer A, Castro-Fornieles J. Cerebral activation in children and adolescents with obsessive-compulsive disorder before and after treatment: a functional MRI study. J Psychiatr Res. 2008; 42:1051–1059. [PubMed: 18261744]

Lee KF, Soares C, Beique JC. Examining form and function of dendritic spines. Neural Plast. 2012; 2012:704103. [PubMed: 22577585]

Leung HC, Skudlarski P, Gatenby JC, Peterson BS, Gore JC. An event-related functional MRI study of the stroop color word interference task. Cereb Cortex. 2000; 10:552–560. [PubMed: 10859133]

Lim L, Marquand A, Cubillo AA, Smith AB, Chantiluke K, Simmons A, Mehta M, Rubia K. Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (ADHD) relative to autism using structural magnetic resonance imaging. PLoS One. 2013; 8:e63660. [PubMed: 23696841]

Marquand AF, O'Daly OG, De Simoni S, Alsop DC, Maguire RP, Williams SC, Zelaya FO, Mehta MA. Dissociable effects of methylphenidate, atomoxetine and placebo on regional cerebral blood flow in healthy volunteers at rest: a multi-class pattern recognition approach. Neuroimage. 2012; 60:1015–1024. [PubMed: 22266414]

Mattai AA, Weisinger B, Greenstein D, Stidd R, Clasen L, Miller R, Tossell JW, Rapoport JL, Gogtay N. Normalization of cortical gray matter deficits in nonpsychotic siblings of patients with childhood-onset schizophrenia. J Am Acad Child Adolesc Psychiatry. 2011; 50:697–704. [PubMed: 21703497]

Matthews SC, Strigo IA, Simmons AN, O'Connell RM, Reinhardt LE, Moseley SA. A multimodal imaging study in U.S. veterans of Operations Iraqi and Enduring Freedom with and without major depression after blast-related concussion. Neuroimage. 2011; 54(Suppl 1):S69–75. [PubMed: 20451622]

McDonald C, Bullmore ET, Sham PC, Chitnis X, Wickham H, Bramon E, Murray RM. Association of genetic risks for schizophrenia and bipolar disorder with specific and generic brain structural endophenotypes. Arch Gen Psychiatry. 2004; 61:974–984. [PubMed: 15466670]

Meltzer CC, Leal JP, Mayberg HS, Wagner HN Jr. Frost JJ. Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. J Comput Assist Tomogr. 1990; 14:561–570. [PubMed: 2370355]

Mohtasib RS, Lumley G, Goodwin JA, Emsley HC, Sluming V, Parkes LM. Calibrated fMRI during a cognitive Stroop task reveals reduced metabolic response with increasing age. Neuroimage. 2012; 59:1143–1151. [PubMed: 21843646]

Mosconi MW, Cody-Hazlett H, Poe MD, Gerig G, Gimpel-Smith R, Piven J. Longitudinal study of amygdala volume and joint attention in 2- to 4-year-old children with autism. Arch Gen Psychiatry. 2009; 66:509–516. [PubMed: 19414710]

Mwangi B, Ebmeier KP, Matthews K, Steele JD. Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. Brain. 2012; 135:1508–1521. [PubMed: 22544901]

Nacewicz BM, Dalton KM, Johnstone T, Long MT, McAuliff EM, Oakes TR, Alexander AL, Davidson RJ. Amygdala volume and nonverbal social impairment in adolescent and adult males with autism. Arch Gen Psychiatry. 2006; 63:1417–1428. [PubMed: 17146016]

Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci. 2006; 10:424–430. [PubMed: 16899397]

Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev. 2012; 36:1140–1152. [PubMed: 22305994]

Palmen SJ, Durston S, Nederveen H, Van Engeland H. No evidence for preferential involvement of medial temporal lobe structures in high-functioning autism. Psychol Med. 2006; 36:827–834. [PubMed: 16512972]

Pardo JV, Pardo PJ, Janer KW, Raichle ME. The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proc Natl Acad Sci U S A. 1990; 87:256–259. [PubMed: 2296583]

Paus T. Population neuroscience: why and how. Hum Brain Mapp. 2010; 31:891–903. [PubMed: 20496380]

Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. Neuroimage. 2009; 45:S199–209. [PubMed: 19070668]

Peterson BS. Conceptual, methodological, and statistical challenges in brain imaging studies of developmentally based psychopathologies. Dev Psychopathol. 2003; 15:811–832. [PubMed: 14582941]

Peterson BS, Kane MJ, Alexander GM, Lacadie C, Skudlarski P, Leung HC, May J, Gore JC. An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. Brain Res Cogn Brain Res. 2002; 13:427–440. [PubMed: 11919006]

Peterson BS, Potenza MN, Wang Z, Zhu H, Martin A, Marsh R, Plessen KJ, Yu S. An FMRI study of the effects of psychostimulants on default-mode processing during Stroop task performance in youths with ADHD. Am J Psychiatry. 2009a; 166:1286–1294. [PubMed: 19755575]

Peterson BS, Skudlarski P, Anderson AW, Zhang H, Gatenby JC, Lacadie CM, Leckman JF, Gore JC. A functional magnetic resonance imaging study of tic suppression in Tourette syndrome. Arch Gen Psychiatry. 1998; 55:326–333. [PubMed: 9554428]

Peterson BS, Skudlarski P, Gatenby JC, Zhang H, Anderson AW, Gore JC. An fMRI study of Stroop word-color interference: evidence for cingulate subregions subserving multiple distributed attentional systems. Biol Psychiatry. 1999; 45:1237–1258. [PubMed: 10349031]

Peterson BS, Staib L, Scahill L, Zhang H, Anderson C, Leckman JF, Cohen DJ, Gore JC, Albert J, Webster R. Regional brain and ventricular volumes in Tourette syndrome. Arch Gen Psychiatry. 2001; 58:427–440. [PubMed: 11343521]

Peterson BS, Wang Z, Horga G, Warner V, Rutherford B, Klahr K, Graniello B, Wickramaratne P, Garcia F, Yu S, Hao X, Adams P, Qian M, Liu J, Gerber A, Weissman M. Brain activations discriminate risk and resilience endophenotypes from markers of lifetime illness in familial Major Depressive Disorder. JAMA Psychiatry. in press.

Peterson BS, Warner V, Bansal R, Zhu H, Hao X, Liu J, Durkin K, Adams PB, Wickramaratne P, Weissman MM. Cortical thinning in persons at increased familial risk for major depression. Proc Natl Acad Sci U S A. 2009b; 106:6273–6278. [PubMed: 19329490]

Philip RC, Dauvermann MR, Whalley HC, Baynham K, Lawrie SM, Stanfield AC. A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. Neurosci Biobehav Rev. 2012; 36:901–942. [PubMed: 22101112]

Pierce K, Muller RA, Ambrose J, Allen G, Courchesne E. Face processing occurs outside the fusiform ‘face area’ in autism: evidence from functional MRI. Brain. 2001; 124:2059–2073. [PubMed: 11571222]

Plessen KJ, Bansal R, Peterson BS. Imaging evidence for anatomical disturbances and neuroplastic compensation in persons with Tourette syndrome. J Psychosom Res. 2009; 67:559–573. [PubMed: 19913660]

Pomarol-Clotet E, Canales-Rodriguez EJ, Salvador R, Sarro S, Gomar JJ, Vila F, Ortiz-Gil J, Iturria-Medina Y, Capdevila A, McKenna PJ. Medial prefrontal cortex pathology in schizophrenia as revealed by convergent findings from multimodal imaging. Mol Psychiatry. 2010; 15:823–830. [PubMed: 20065955]

Posner J, Hellerstein DJ, Gat I, Mechling A, Klahr K, Wang Z, McGrath PJ, Stewart JW, Peterson BS. Antidepressants Normalize the Default Mode Network in Patients With Dysthymia. JAMA Psychiatry. 2013:1–10.

Rauch SL, Jenike MA, Alpert NM, Baer L, Breiter HC, Savage CR, Fischman AJ. Regional cerebral blood flow measured during symptom provocation in obsessive-compulsive disorder using oxygen 15-labeled carbon dioxide and positron emission tomography. Arch Gen Psychiatry. 1994; 51:62–70. [PubMed: 8279930]

Rauh V, Arunajadai S, Horton M, Perera F, Hoepner L, Barr DB, Whyatt R. Seven-year neurodevelopmental scores and prenatal exposure to chlorpyrifos, a common agricultural pesticide. Environ Health Perspect. 2011; 119:1196–1201. [PubMed: 21507777]

Rauh VA, Perera FP, Horton MK, Whyatt RM, Bansal R, Hao X, Liu J, Barr DB, Slotkin TA, Peterson BS. Brain anomalies in children exposed prenatally to a common organophosphate pesticide. Proc Natl Acad Sci U S A. 2012; 109:7871–7876. [PubMed: 22547821]

Rubia K, Halari R, Cubillo A, Smith AB, Mohammad AM, Brammer M, Taylor E. Methylphenidate normalizes fronto-striatal underactivation during interference inhibition in medication-naive boys with attention-deficit hyperactivity disorder. Neuropsychopharmacology. 2011a; 36:1575–1586. [PubMed: 21451498]

Rubia K, Halari R, Mohammad AM, Taylor E, Brammer M. Methylphenidate normalizes frontocingulate underactivation during error processing in attention-deficit/hyperactivity disorder. Biol Psychiatry. 2011b; 70:255–262. [PubMed: 21664605]

Rubin RT, Ananth J, Villanueva-Meyer J, Trajmar PG, Mena I. Regional 133xenon cerebral blood flow and cerebral 99mTc-HMPAO uptake in patients with obsessive-compulsive disorder before and during treatment. Biol Psychiatry. 1995; 38:429–437. [PubMed: 8672603]

Schobel SA, Chaudhury NH, Khan UA, Paniagua B, Styner MA, Asllani I, Inbar BP, Corcoran CM, Lieberman JA, Moore H, Small SA. Imaging patients with psychosis and a mouse model establishes a spreading pattern of hippocampal dysfunction and implicates glutamate as a driver. Neuron. 2013; 78:81–93. [PubMed: 23583108]

Schobel SA, Lewandowski NM, Corcoran CM, Moore H, Brown T, Malaspina D, Small SA. Differential targeting of the CA1 subfield of the hippocampal formation by schizophrenia and related psychotic disorders. Arch Gen Psychiatry. 2009; 66:938–946. [PubMed: 19736350]

Schumann CM, Barnes CC, Lord C, Courchesne E. Amygdala enlargement in toddlers with autism related to severity of social and communication impairments. Biol Psychiatry. 2009; 66:942–949. [PubMed: 19726029]

Schumann CM, Bloss CS, Barnes CC, Wideman GM, Carper RA, Akshoomoff N, Pierce K, Hagler D, Schork N, Lord C, Courchesne E. Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism. J Neurosci. 2010; 30:4419–4427. [PubMed: 20335478]

Schwartz JM, Stoessel PW, Baxter LR Jr. Martin KM, Phelps ME. Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of obsessive-compulsive disorder. Arch Gen Psychiatry. 1996; 53:109–113. [PubMed: 8629886]

Schwartz S, Susser E. The use of well controls: an unhealthy practice in psychiatric research. Psychol Med. 2011; 41:1127–1131. [PubMed: 20810003]

Sexton CE, Allan CL, Le Masurier M, McDermott LM, Kalu UG, Herrmann LL, Maurer M, Bradley KM, Mackay CE, Ebmeier KP. Magnetic resonance imaging in late-life depression: multimodal examination of network disruption. Arch Gen Psychiatry. 2012; 69:680–689. [PubMed: 22752234]

Shaw P, Eckstrand K, Sharp W, Blumenthal J, Lerch JP, Greenstein D, Clasen L, Evans A, Giedd J, Rapoport JL. Attention-deficit/hyperactivity disorder is characterized by a delay in cortical maturation. Proc Natl Acad Sci U S A. 2007; 104:19649–19654. [PubMed: 18024590]

Sowell ER, Thompson PM, Leonard CM, Welcome SE, Kan E, Toga AW. Longitudinal mapping of cortical thickness and brain growth in normal children. J Neurosci. 2004; 24:8223–8231. [PubMed: 15385605]

Sowell ER, Thompson PM, Welcome SE, Henkenius AL, Toga AW, Peterson BS. Cortical abnormalities in children and adolescents with attention-deficit hyperactivity disorder. Lancet. 2003; 362:1699–1707. [PubMed: 14643117]

Stanfield AC, McIntosh AM, Spencer MD, Philip R, Gaur S, Lawrie SM. Towards a neuroanatomy of autism: a systematic review and meta-analysis of structural magnetic resonance imaging studies. Eur Psychiatry. 2008; 23:289–299. [PubMed: 17765485]

Sui J, Pearlson G, Caprihan A, Adali T, Kiehl KA, Liu J, Yamamoto J, Calhoun VD. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. Neuroimage. 2011; 57:839–855. [PubMed: 21640835]

Swedo SE, Pietrini P, Leonard HL, Schapiro MB, Rettew DC, Goldberger EL, Rapoport SI, Rapoport JL, Grady CL. Cerebral glucose metabolism in childhood-onset obsessive-compulsive disorder. Revisualization during pharmacotherapy. Arch Gen Psychiatry. 1992; 49:690–694. [PubMed: 1514873]

Swedo SE, Schapiro MB, Grady CL, Cheslow DL, Leonard HL, Kumar A, Friedland R, Rapoport SI, Rapoport JL. Cerebral glucose metabolism in childhood-onset obsessive-compulsive disorder. Arch Gen Psychiatry. 1989; 46:518–523. [PubMed: 2786402]

Talbot JD, Marrett S, Evans AC, Meyer E, Bushnell MC, Duncan GH. Multiple representations of pain in human cerebral cortex. Science. 1991; 251:1355–1358. [PubMed: 2003220]

Vaidya CJ, Austin G, Kirkorian G, Ridlehuber HW, Desmond JE, Glover GH, Gabrieli JD. Selective effects of methylphenidate in attention deficit hyperactivity disorder: a functional magnetic resonance study. Proc Natl Acad Sci U S A. 1998; 95:14494–14499. [PubMed: 9826728]

Valera EM, Faraone SV, Murray KE, Seidman LJ. Meta-analysis of structural imaging findings in attention-deficit/hyperactivity disorder. Biol Psychiatry. 2007; 61:1361–1369. [PubMed: 16950217]

van den Bulk BG, Koolschijn PC, Meens PH, van Lang ND, van der Wee NJ, Rombouts SA, Vermeiren RR, Crone EA. How stable is activation in the amygdala and prefrontal cortex in adolescence? A study of emotional face processing across three measurements. Dev Cogn Neurosci. 2013; 4:65–76. [PubMed: 23043904]

Van Erp TG, Saleh PA, Rosso IM, Huttunen M, Lonnqvist J, Pirkola T, Salonen O, Valanne L, Poutanen VP, Standertskjold-Nordenstam CG, Cannon TD. Contributions of genetic risk and fetal hypoxia to hippocampal volume in patients with schizophrenia or schizoaffective disorder, their unaffected siblings, and healthy unrelated volunteers. Am J Psychiatry. 2002; 159:1514–1520. [PubMed: 12202271]

Velakoulis D, Wood SJ, Wong MT, McGorry PD, Yung A, Phillips L, Smith D, Brewer W, Proffitt T, Desmond P, Pantelis C. Hippocampal and amygdala volumes according to psychosis stage and diagnosis: a magnetic resonance imaging study of chronic schizophrenia, first-episode psychosis, and ultra-high-risk individuals. Arch Gen Psychiatry. 2006; 63:139–149. [PubMed: 16461856]

Waber DP, De Moor C, Forbes PW, Almli CR, Botteron KN, Leonard G, Milovan D, Paus T, Rumsey J, Brain Development Cooperative, G. The NIH MRI study of normal brain development: performance of a population based sample of healthy children aged 6 to 18 years on a neuropsychological battery. J Int Neuropsychol Soc. 2007; 13:729–746. [PubMed: 17511896]

Weissman MM, Wickramaratne P, Nomura Y, Warner V, Pilowsky D, Verdeli H. Offspring of depressed parents: 20 years later. Am J Psychiatry. 2006; 163:1001–1008. [PubMed: 16741200]

Zanto TP, Rubens MT, Thangavel A, Gazzaley A. Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. Nat Neurosci. 2011; 14:656–661. [PubMed: 21441920]

Zhu CZ, Zang YF, Cao QJ, Yan CG, He Y, Jiang TZ, Sui MQ, Wang YF. Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder. Neuroimage. 2008; 40:110–120. [PubMed: 18191584]

Key points

- Ascertainment bias can preclude the ability of imaging studies to draw conclusions about the general population, a limitation that can be circumvented through epidemiological ascertainment of study samples.
- Moving from cross-sectional studies to repeated-measures, longitudinal designs may enhance our ability to identify developmental trajectories in health and illness.
- Yoking imaging technologies to high-risk and RCT study designs can help draw more valid conclusions about mechanisms of pathogenesis and treatment.
- Multi-modal imaging approaches may help integrate findings across imaging modalities and contribute to a more holistic view of the pathophysiology of mental illness.
- Imaging tools for clinical diagnosis and prognosis based on multivariate pattern classification are already at our disposal, but further validation of the techniques and their clinical applicability is needed.