Statistical Approaches to Complex Phenotypes: Evaluating Neuropsychological Endophenotypes for Attention-Deficit/Hyperactivity Disorder

Irwin D. Waldman

There is renewed interest among psychiatric geneticists in endophenotypes, constructs posited to be more directly and strongly influenced by candidate genes than manifest disorders. Researchers have proposed various criteria for the selection of endophenotypes useful in finding genes that underlie psychiatric disorders. These criteria include the endophenotype's psychometric properties, its relationship to the disorder in the population and within families, its expression in probands' unaffected relatives, its heritability and common genetic influences with the disorder, its association with candidate genes that underlie the disorder, and its mediation and moderation of association between the candidate gene and the disorder. In this article, analytic methods for evaluating the validity and utility of putative endophenotypes consistent with these proposed criteria are reviewed. The use of such analyses is illustrated with data on childhood attention-deficit/hyperactivity disorder and time to complete Trails A and B from both a candidate gene study of clinically referred children and a study of non-referred twins. It is demonstrated that both putative endophenotypes show association with the dopamine D4 receptor gene and meet most (but not all) of the criteria proposed for their validity and utility.

Key Words: Endophenotypes, validity, analytic methods, neuropsychological measures, candidate genes, association, linkage

There is a large gap between candidate genes and the manifest symptoms of disorders, such as attention-deficit/hyperactivity disorder (ADHD), as typically assessed by interviews or rating scales. It is desirable from both conceptual and empirical perspectives to find valid and meaningful mediational or intervening constructs that might help to bridge this gap. The term "endophenotype" is often used to describe such constructs and the variables that are used to measure them. Endophenotypes were first described with respect to psychiatric disorders by Gottesman and Shields more than 35 years ago in their application to the genetics of schizophrenia (Gottesman and Shields 1967) as "internal phenotypes discoverable by a biochemical test or microscopic examination" (Gottesman and Gould 2003; Gottesman and Shields 1967). More generally, endophenotypes refer to constructs that are thought to underlie psychiatric disorders and to be more directly influenced by the genes relevant to the disorder than are the manifest symptoms. As such, they are closer to the immediate products of such genes (i.e., the proteins they code for) and are thought to be more strongly influenced by the genes that underlie them than the manifest symptoms that they in turn undergird. Endophenotypes also are thought to be "genetically simpler" in their etiology than are complex traits, such as manifest disorders or their symptom dimensions (Gottesman and Gould 2003). This means that the underlying structure of genetic influences on endophenotypes is simpler than that of complex disorders and traits, in that there are fewer individual genes (or sets thereof) that contribute to their etiology.

A number of researchers have outlined criteria for evaluating the validity and utility of putative endophenotypes (e.g., Almasy and Blangero 2001; Castellanos and Tannock 2002; Cornblatt and Malhotra 2001; Gottesman and Gould 2003; Doyle et al 2005). These criteria include 1) the endophenotype has good psychometric properties; 2) the endophenotype is related to the disorder and its symptoms in the general population; 3) the endophenotype is stable over time (i.e., is expressed regardless of whether or not the disorder is currently manifest); 4) the endophenotype is expressed at a higher rate in the unaffected relatives of probands than in randomly selected individuals from the general population; 5) the endophenotype and disorder are associated within families (i.e., they "co-segregate"); 6) the endophenotype is heritable; 7) there are common genetic influences underlying the endophenotype and the disorder; 8) the endophenotype must show association and/or linkage with one (or more) of the candidate genes or genetic loci that underlie the disorder, and should show association with the gene over and above the gene's association with the diagnosis or symptoms; 9) the endophenotype should *mediate* association and/or linkage between the candidate gene and the disorder, meaning that the effects of a particular gene or locus on a disorder are expressed—either in full or in part—through the endophenotype; and 10) the endophenotype should *moderate* association and/or linkage between the candidate gene and the disorder, meaning that the effects of a particular gene or locus on a disorder are stronger in disordered individuals who also show the endophenotype. It should be noted that criteria 1–7 are important for evaluating the validity of putative endophenotypes and for suggesting their promise for inclusion in molecular genetic studies, whereas criteria 8–10 indicate the utility of putative endophenotypes within candidate gene studies.

In this article, I describe various analyses that might be useful for evaluating the validity and utility of putative endophenotypes according to the above criteria. Although several researchers have proposed research designs and analyses to address selected criteria above (e.g., Almasy and Blangero 2001; Cornblatt and Malhotra 2001; Gottesman and Gould 2003; Seidman et al 2000; Doyle et al 2005), to my knowledge there has not previously been a comprehensive treatment of this topic in which analyses were proposed to address all of the aforementioned criteria for evaluating endophenotypes. I illustrate the analyses for addressing the above criteria using data on executive functions, ADHD symptoms, and the dopamine D4 receptor gene (*DRD4*) from both a population-based twin study and a candidate gene study in a clinically referred sample. The results suggest that although

From the Department of Psychology, Emory University, Atlanta, Georgia.
Address reprint requests to Irwin D. Waldman, Ph.D., Emory University, Department of Psychology, 532 North Kilgo Circle, Atlanta, GA 30322; E-mail: psyiw@emory.edu.

Received June 14, 2004; revised March 2, 2005; accepted March 2, 2005.

0006-3223/05/$30.00
doi:10.1016/j.biopsych.2005.03.002

most of the criteria were met for the putative endophenotype variables tested, some were not, which could lead to the mistaken rejection of putative endophenotypes that are valid and useful for finding genes for disorders.

Methods and Materials

Participants

The clinically referred sample consists of 379 children from 220 families recruited through the Center for Learning and Attention Deficit Disorders at the Emory University School of Medicine and the Emory University Psychological Center in Atlanta, Georgia, as well as from psychiatrists in private practice in Tucson, Arizona. The Center for Learning and Attention Deficit Disorders is a clinic that specializes in the assessment and treatment of childhood externalizing disorders such as ADHD, oppositional defiant disorder, and conduct disorder. The Psychological Center is part of the Department of Psychology's clinical training program and receives referrals for the assessment and treatment of children with ADHD and learning disorders. Any diagnosis assigned to a child remained confidential and did not influence inclusion in the study. The participants represent an expanded sample from that in previous publications (e.g., Rowe et al 1998; Waldman et al 1998). Families were assessed in their homes over the course of a 3-hour period. Participating children completed a comprehensive assessment of executive function measures, while their parents completed questionnaires assessing the family's demographic characteristics as well as symptoms of childhood psychiatric disorders. The sample included 195 probands who met criteria for a diagnosis of ADHD, 101 unaffected siblings who did not meet criteria for ADHD, and 38 affected siblings who met ADHD criteria. Data on Trails A and B were available for 137 probands, 71 unaffected siblings, and 25 affected siblings. Children ranged in age from 6 to 18 years (mean [SD]: 10.4 [3.2] years), and 63% were male. The ethnic background of the sample was 76% Caucasian, 10% African American, 1% Hispanic, 2% Asian, and 10% of mixed ethnicity. This study received approval from the institutional review boards of both Emory University and the University of Arizona.

The twin sample was drawn from the Georgia Twin Registry, a sample of twins from the general population of Georgia born between 1980 and 1991 and recruited through birth records. The twin families in the current laboratory study had previously participated in a questionnaire study of child psychopathology. Parents of the twins were contacted by telephone to participate in the current study, and the twin families were assessed in our laboratory at Emory University for a 3-hour period. Data from 86 twin pairs (51 monozygotic [MZ] and 35 dizygotic [DZ] pairs) were available for this study. Zygosity was determined with a 9-item questionnaire on which parents rated the physical similarity of their twins. Similar questionnaires have shown high accuracy (≥95%) in determining zygosity when compared with the use of DNA polymorphisms (Spitz et al 1996). Participating twins completed a comprehensive assessment of executive function measures that was virtually identical to that used in the clinic-referred sample, and their parents completed the same symptom rating scales on their twins. Twin pairs ranged in age from 6 to 18 years (13.2 [2.5] years), and 42% of the twins were male. The ethnic background of the sample was 94% Caucasian, 3% African American, and 1% Asian. This study received approval from the Emory University institutional review board.

Measures

Emory Diagnostic Rating Scale. Symptom ratings were obtained for each child with the Emory Diagnostic Rating Scale from mothers and fathers (whenever possible). The Emory Diagnostic Rating Scale was developed in our laboratory to assess symptoms of the major DSM-IV childhood psychiatric disorders, including disruptive disorders such as ADHD, oppositional defiant disorder, and conduct disorder, and internalizing disorders such as depression and anxiety disorders. Parents rated symptoms on a 0–4 scale, and the symptom scores were summed to create hyperactive–impulsive and inattentive ADHD symptom dimensions. Questionnaire-based diagnoses of ADHD were made using appropriate DSM-IV diagnostic cutoffs (i.e., six or more of the nine symptoms on either of the ADHD symptom dimensions).

Trails A and B. In Trails A, children were presented with a page of circles with numbers in them and were required to connect the circles in order (i.e., 1, 2, 3, and so on). In Trails B, children were presented with a similar page of circles containing letters and numbers presented in a random pattern and were required to connect the two sets of ordered stimuli (letters and numbers) in an alternating fashion (e.g., 1, A, 2, B, and so on). The speed with which children complete this task is hypothesized to assess the ability to shift attention between sets and served as the primary endophenotype measure in this study, with speed to complete Trails A included as a comparison measure.

Trails A has been interpreted as a measure of processing speed that relies on visual perceptual ability and motor speed (Crowe 1998). Factor analytic studies relating performance on this test to other processing speed measures have provided some empirical support for this claim. For example, Shute and Huertas (1990) administered a battery of eight tests, including the Digit-Symbol subtest of the Wechsler Intelligence Scale for Children and Trails A and B, to a sample of 58 normal college students. An exploratory factor analysis revealed that performance on the Digit-Symbol test and the time to complete Trails A loaded on a single factor, which was interpreted to assess processing speed. Such evidence suggests that Trails A is not a measure of executive function per se, which has been further supported by studies of patients with prefrontal lesions, suggesting that Trails A cannot reliably differentiate between individuals with frontal lobe damage and normal control subjects (Lezak 1995).

Trails B has been widely interpreted as a measure of set-shifting ability in the neuropsychological literature. Factor analytic studies relating performance on this test to other executive function measures have provided some empirical support for this claim. For example, Shute and Huertas' (1990) exploratory factor analysis of a battery of eight tests, including the Wisconsin Card Sorting Task (WCST) and Trails A and B, revealed that the perseverative errors score from the WCST and the time to complete Trails B loaded on a single factor, which was interpreted to assess set-shifting. Similarly, Kortte et al (2002) administered a battery of executive function measures to a group of adults and found that after controlling for age, gender, and performance on Trails A, only perseverative errors on the WCST predicted performance on Trails B. Additional evidence supporting the validity of Trails B as a measure of set-shifting comes from studies of brain-damaged patients. As a component of executive functioning, set-shifting has been hypothesized to be under the control of the prefrontal cortex. Several studies have demonstrated that individuals with prefrontal lesions display deficits on Trails B, but these deficits are not specific to frontal lobe damage, because other studies have demonstrated that lesions in a variety