Journal of Consulting and Clinical Psychology
1991, Vol. 59, No. 5, 732-738

Copyright 1991 by the American Psychological Association, Inc.
0022-006X/91/$3.00

Diagnostic Efficiency of Symptoms for Oppositional Defiant Disorder and Attention-Deficit Hyperactivity Disorder

Irwin D. Waldman and Scott O. Lilienfeld
University of Minnesota

We examined the diagnostic efficiency of symptoms for experimentally diagnosed oppositional defiant disorder (OD) and attention-deficit hyperactivity disorder (ADHD), and for their differential diagnosis, in a nonreferred sample. More specifically, we used four conditional probability indices—positive predictive power, negative predictive power, sensitivity, and specificity—in an attempt to evaluate the diagnostic efficiency of these symptoms. Teacher ratings of OD and ADHD symptoms were gathered on 102 boys, ages 8 to 12. On average, ADHD symptoms were as useful as OD symptoms as exclusion criteria for OD, while OD symptoms were nearly as effective as ADHD symptoms as inclusion criteria for ADHD. Nonetheless, a number of OD and ADHD symptoms appeared to be useful as both inclusion and exclusion criteria for their respective disorders. These findings illustrate the utility of conditional probability indices in both the diagnosis and differentiation of childhood disorders.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The distinction between externalizing (i.e., disruptive, as defined in the *Diagnostic and Statistical Manual of Mental Disorders*, 3rd ed., rev., *DSM-III-R*; American Psychiatric Association, 1987) and internalizing childhood disorders has received considerable empirical support (Achenbach & Edelbrock, 1978). In many cases, however, the validity of the distinction among narrow-band syndromes within these two broad-band domains remains unclear (Hinshaw, 1987). In the domain of childhood externalizing disorders, for example, only recently has a consensus been reached that attention-deficit hyperactivity disorder (ADHD) and conduct disorder (CD) are moderately to highly correlated, but separable, problem areas (Hinshaw, 1987; Lilienfeld & Waldman, 1990). Moreover, the construct validity of the third major externalizing disorder in *DSM-III-R*, oppositional defiant disorder (OD; American Psychiatric Association, 1987), has received little research attention.

Introduced as oppositional disorder in *DSM*, 3rd ed., (*DSM-III*; American Psychiatric Association, 1980), OD is characterized by a "pattern of negativistic, hostile, and defiant behavior" (American Psychiatric Association, 1987, p. 56) that is developmentally inappropriate and not accompanied by the more severe symptoms[1] of CD. OD has been reported to overlap moderately with ADHD and moderately to highly with CD (Rey et al., 1988; Werry, Methven, Fitzpatrick, & Dixon, 1983). Along these lines, Ferguson and Rapoport (1984) argued that ADHD and OD may be difficult to distinguish and that some individuals with ADHD may be mistakenly diagnosed as having OD. Nevertheless, the distinguishability of ADHD and OD has not been extensively explored. In particular, the efficiency of the symptoms of ADHD and OD for the differential diagnosis of these two disorders has, to our knowledge, never been examined.

Four indices are particularly useful for examining the efficiency of symptoms in the diagnosis and differential diagnosis of a disorder: sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP). Sensitivity and specificity, by far the most commonly used of these four indices, are the conditional probability of the presence of a symptom given the presence of a disorder and the conditional probability of the absence of a symptom given the absence of a disorder, respectively. Although useful for a number of purposes, sensitivity and specificity do not provide information directly relevant to the diagnostic decision-making process (Milich, Widiger, & Landau, 1987; Widiger, Hurt, Frances, Clarkin, & Gilmore, 1984). Instead, the statistics most useful for this purpose are PPP and NPP, which are the conditional probability of the presence of a diagnosis given the presence of a symptom and the conditional probability of the absence of a diagnosis given the absence of a symptom, respectively. PPP and NPP indicate the utility of symptoms as inclusion and exclusion criteria, respectively. These indices, unlike sensitivity and specificity, provide the clinician and researcher with the hit rate of a diagnosis based on a symptom or set of symptoms, and they are affected by the base rates of the disorder(s) of interest. This sensitivity to base rates frequently causes PPP and NPP to fluctuate greatly across samples (Baldessarini, Finkelstein, & Arana, 1983; Meehl & Rosen, 1955), but this is a potential advantage because PPP and NPP reflect the changes in diagnostic accuracy that result from changes in these base rates. For example, a symptom may have extremely high sensitivity and specific-

We would like to thank William M. Grove, who contributed considerably to the conceptual and statistical health of this article, and Paul J. Frick and Benjamin B. Lahey, who offered helpful comments on the issue of chance correction of conditional probability indices.

Correspondence concerning this article should be addressed to Irwin D. Waldman, who is now at Emory University, Department of Psychology, Atlanta, Georgia 30322.

[1] Like Widiger, Hurt, Frances, Clarkin, and Gilmore (1984), we use the word *symptom* to mean any expression of a patient's disorder, whether observable or not.

ity for a disorder, yet have poor PPP and NPP when applied to samples with low base rates of this disorder (e.g., Baldessarini et al., 1983).

Milich et al. (1987), using parental structured psychiatric interviews in an outpatient clinic sample, examined the diagnostic efficiency of symptoms of *DSM–III* attention deficit disorder (ADD) and CD. They found that some symptoms of these disorders had high PPP, some had high NPP, only a few had both high PPP and NPP, and some had both low PPP and NPP. In addition, some symptoms of ADD had equally high or higher NPP for CD as for ADD. Although Milich et al.'s findings suggest that some *DSM–III* symptoms of ADD and CD are relatively poor for diagnostic purposes, it should be borne in mind that PPP and NPP are sensitive to base rate fluctuations and therefore may change when applied to a different sample. Thus, it is necessary to examine PPP and NPP across a variety of settings (Widiger et al., 1984).

In this article, we attempt to extend Milich et al.'s findings by (a) examining, for the first time, the efficiency of the symptoms of ADHD and OD in the differential diagnosis of each disorder; (b) using *DSM–III–R*, rather than *DSM–III*, criteria; (c) using a nonclinic, rather than a clinic, sample; and (d) using teacher ratings, rather than parental interviews.

With regard to the first point, we were not able to assign strict diagnoses of ADHD and OD, as age of onset and duration of symptoms were not assessed. Nonetheless, our classification of children as ADHD or OD on the basis of teacher-rated symptoms is similar to that used in many other studies of these disorders (see below).[2] With regard to the last point, there is some suggestion that diagnoses of ADHD that are based on teacher ratings are more valid than are diagnoses based on parental interviews (Ross & Pelham, 1978), probably because teachers possess greater knowledge of developmentally appropriate norms. Consequently, our study may provide a stronger test of the diagnostic efficiency of symptoms than that of Milich et al. More important, our study should clarify the extent to which the *DSM–III–R* criteria can distinguish between ADHD and OD (Ferguson & Rapoport, 1984).[3]

Method

Subjects

One hundred and five boys, aged 8–12 years (Grades 3 through 6), were sampled from nine suburban parochial and public schools in southwestern Ontario serving a primarily White, middle-class population. None of the children had pervasive developmental disorder, which is the sole exclusion criterion for ADHD in *DSM–III–R*. The presence of CD, which is the sole exclusion criterion for OD, was not assessed.

Participating boys were those who met the selection criteria for membership in one of four peer status groups. Classification into the four groups was based on peer sociometric ratings of popularity (Singleton & Asher, 1977) and peer nominations of aggression and isolation (Masten, Morrison, & Pelligrini, 1985). The groups derived from these measures were (a) average peer status (average popularity, scores below the mean on both aggression and isolation); (b) isolate status (extremely low scores on popularity, extremely high scores on isolation); (c) aggressive status (extremely low scores on popularity, extremely high scores on aggression); and (d) isolate–aggressive status (extremely low scores on popularity, extremely high scores on both isolation and aggression). More information on the peer status classification is presented in Waldman (1988, 1990).

Although this is neither a clinical nor a random community sample, there are at least two reasons for its relevance to evaluating the current diagnostic criteria for externalizing disorders. First, children identified as having peer relations difficulties might be expected to have an elevated rate of current or future clinical referrals, given that such difficulties are commonly found in children with externalizing behavior problems (Pelham & Bender, 1982). Second, as discussed earlier, PPP and NPP may vary markedly across samples with differing characteristics. To our knowledge, our study represents the first investigation of these statistics in a sample defined on the basis of peer relations difficulties, even though such samples often have been used by researchers in the developmental psychopathology literature (e.g., Dodge, 1980; Ledingham, 1981).

Procedure

Teacher rating scales were available for 102 of the subjects. The *DSM–III–R* criteria for ADHD and OD were converted into a rating scale format similar to that used by other authors (Halperin et al., 1988; Lahey et al., 1988; Milich, et al., 1987; Pelham, Atkins, Murphy, & White, 1981). The 14 ADHD symptoms and the 9 OD symptoms were rated by teachers on a 0–3 scale (*Not at all, Just a little, Pretty much, Very much*) on the basis of how characteristic they were of a child's behavior. In this sample, Cronbach's alphas for the ADHD and OD scales were .95 and .95, respectively.

A *DSM–III–R* symptom was coded present if teachers rated it as *Pretty much* or *Very much* (similar to Milich et al., 1987). Boys were assigned an experimental diagnosis by *DSM–III–R* criteria: Children with an ADHD score ≥ 8 and an OD score < 5 were categorized as ADHD; children with an ADHD score < 8 and an OD score ≥ 5 were categorized as OD; children with an ADHD score ≥ 8 and an OD score ≥ 5 were catagorized as both ADHD and OD; and children with an ADHD score < 8 and an OD score < 5 were categorized as having neither disorder. The number of symptoms required for each diagnosis were as specified in *DSM–III–R* (pp. 52–53, 57–58).

It should be noted that teacher rating scales do not permit the researcher to make strict *DSM–III–R* diagnoses of ADHD and OD, as these scales do not typically assess age of onset, duration, or exclusion criteria for these disorders (although as noted earlier, there were no children with pervasive developmental disorder in this sample). Nevertheless, such scales, which have been extensively used in past studies to make diagnoses of externalizing disorders (e.g., Atkins, Pelham, &

[2] As an expository convenience, we will use the term *experimental diagnosis* throughout this article to refer to the classification of children into ADHD and OD groups on the basis of our use of the *DSM–III–R* symptom cutoffs. We urge the reader to keep in mind the limitations of ADHD and OD "diagnoses" made without data on age of onset and symptom duration.

[3] Spitzer, Davies, and Barkley (1990) have recently presented analyses of the conditional probabilities used to determine the criteria for ADHD and OD in *DSM–III–R*. Nevertheless, they did not report the PPPs and NPPs for individual symptoms of these disorders or for their differential diagnosis. As these represent the two major foci of the current article, their results will not be further discussed here.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Licht, 1985; August & Garfinkel, 1989; Offord et al., 1987), appear to possess good construct validity for ADHD and related diagnoses. For example, children identified as ADD by the SNAP rating scale (Pelham et al., 1981), which is extremely similar in format to the rating scale used here, differ from non-ADD children in observed classroom behaviors relevant to inattention and impulsivity (Atkins et al., 1985), peer-rated aggression and likability (Johnston, Pelham, & Murphy, 1985), and teacher-rated conduct problems (Pelham et al., 1981).

Results

Results of this study will be presented in a format similar to that used by Milich et al. (1987), given the similarity of analyses used. We first examined the number and percentage of the sample that met the criteria for ADHD, OD, or both diagnoses. Sixty-six boys (65%) had neither disorder, 15 boys (15%) had ADHD only, 7 boys (7%) had OD only, and 14 boys (14%) had both ADHD and OD. (By way of comparison, in Milich et al. [1987] 39% of the sample had neither disorder, 37% had ADD only, 8% had CD only, and 16% had both ADD and CD.) Although the base rate for ADHD in our study was approximately three times higher than Ontario population norms for ADD (Offord et al., 1987), this is not surprising given that the majority of our sample was selected on the basis of peer relations difficulties, which are common in children with ADHD (Pelham & Bender, 1982). There was significant overlap between ADHD and OD, indicated both by an analysis of the overall contingency table, $\chi^2(1, N = 102) = 19.00$, $p < .001$; phi = .43, and by the discrepancy between observed and expected numbers of subjects in various cells. The number of boys with both ADHD and OD was significantly greater than expected (standardized residual = 3.29, $p < .001$), whereas the number of boys with OD who did not have ADHD was less than expected (standardized residual = −2.07, $p < .02$), and the number of boys with ADHD who did not have OD was less than expected (standardized residual = −1.67, $p < .05$). The probability of an experimental diagnosis of ADHD given an experimental diagnosis of OD (i.e., the PPP of ADHD given OD) was .67, whereas the probability of an experimental diagnosis of OD given an experimental diagnosis of ADHD (i.e., the PPP of OD given ADHD) was .50.

The base rates, sensitivity, specificity, PPP, and NPP of each ADHD and OD symptom for the experimental diagnoses of ADHD and OD are presented in Table 1. The mean PPP of the ADHD symptoms for the experimental diagnosis of ADHD was .70, only slightly higher than the mean PPP of the OD symptoms for the experimental diagnosis of ADHD (.64).[4] In contrast, the mean PPP of the OD symptoms for the experimental diagnosis of OD was higher (.80), whereas the mean PPP of the ADHD symptoms for the experimental diagnosis of OD was substantially lower (.44). The mean NPP of the ADHD symptoms for the experimental diagnosis of ADHD was .91 and was virtually equivalent to the mean NPP of these symptoms for the experimental diagnosis of OD (.90). In contrast, there were substantial differences in the mean NPP of the OD symptoms for the two experimental diagnoses. The mean NPP of the OD symptoms for the experimental diagnosis of OD was .95, whereas the mean NPP of the OD symptoms for the experimental diagnosis of ADHD was lower (.81). These findings indicate that, in this sample, the *DSM–III–R* ADHD symptoms were excellent exclusion criteria for both ADHD and OD, whereas the OD symptoms were moderately good inclusion criteria for both disorders.

The mean specificity was greater than the mean sensitivity when comparing each set of symptoms with its own experimental diagnosis, although this difference was greater for the OD symptoms (.94 and .78, respectively) than for the ADHD symptoms (.85 and .78, respectively). Interestingly, the mean specificity of the OD symptoms for the experimental diagnosis of ADHD (.89) was higher than that of the ADHD symptoms for their own experimental diagnosis, whereas the mean sensitivity of the OD symptoms for the experimental diagnosis of ADHD (.46) was lower than the corresponding value for the ADHD symptoms. Although the mean specificity and the mean sensitivity of the ADHD symptoms for the experimental diagnosis of OD were moderately high (.76 and .67, respectively), they were not as high as the corresponding values for the OD symptoms.[5]

The point-biserial correlation of each symptom with the sum of the remaining symptoms for its experimental diagnosis is also presented in Table 1. The median correlations of symptoms with their corresponding experimental diagnosis were high; the median point-biserial correlation of the ADHD symptoms with the ADHD experimental diagnosis was .68 ($SD = .08$), and the median point-biserial correlation of the OD symptoms with the OD experimental diagnosis was .72 ($SD = .06$).

Like Milich et al. (1987), we examined the differential utility of symptoms as inclusion and exclusion criteria. Unlike those authors' findings, a number of the ADHD and OD symptoms in the present sample appeared to be particularly useful as both inclusion and exclusion criteria for their respective disorders. As shown in Table 1, the ADHD symptoms "physically dangerous," "blurts answers," "difficulty waiting turn," and "often interrupts" were most indicative of the presence of ADHD (PPPs were .94, .87, .82, and .81, respectively) but also were useful in ruling out the disorder by their absence (NPPs were .84, .89, .86, and .89, respectively). These symptoms occurred relatively

[4] Although we use the terms *higher* and *lower* when comparing conditional probability indices, there are no adequate procedures for drawing statistical inferences regarding their relative magnitudes (T. A. Widiger, personal communication, May 30, 1989), nor for constructing confidence intervals around these indices.

[5] We also administered the IOWA-revised Conners rating scale (Loney & Milich, 1982), which is composed of two factor-analytically derived subscales, inattention/overactivity (IO) and aggression (A). Nevertheless, as the correlations between the IO and the ADHD scales, and between the A and OD scales were very high ($r = .91$ and .94, respectively) we do not report these data in the text. Similar patterns and magnitudes of PPP, NPP, specificity, and sensitivity were observed for the IOWA–IO and IOWA–A subscales as were observed for the ADHD and OD symptoms. One interesting finding was that given a diagnosis of OD, a child was more likely to have items on the IO than on the A scale, whereas given the absence of a diagnosis of ADHD, a child was more likely not to have items on the A than on the IO scale.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Conditional Probabilities and Base Rates of ADHD and OD Symptoms

Symptom	ADHD BR	ADHD SEN	ADHD SPE	ADHD PPP	ADHD NPP	ADHD PB	OD SEN	OD SPE	OD PPP	OD NPP	OD PB
ADHD											
Fidgets	.36	.90	.85	.70	.95	.73	.71	.73	.41	.91	.48
Difficulty remaining seated	.26	.72	.92	.78	.89	.68	.52	.80	.41	.87	.42
Easily distracted	.49	.93	.68	.54	.96	.70	.81	.59	.34	.92	.37
Difficulty waiting turn	.22	.62	.95	.82	.86	.66	.57	.88	.55	.89	.54
Blurts answers	.23	.69	.96	.87	.89	.64	.48	.84	.43	.86	.40
Fails to finish chores	.43	.86	.74	.57	.93	.63	.86	.68	.41	.95	.45
Difficulty sustaining attention	.43	.90	.75	.59	.95	.70	.76	.65	.36	.91	.38
Shifts activities	.38	.97	.85	.72	.98	.75	.76	.72	.41	.92	.46
Difficulty playing quietly	.32	.79	.86	.70	.91	.69	.67	.77	.42	.90	.49
Often talks	.34	.86	.86	.71	.94	.74	.62	.73	.37	.88	.44
Interrupts	.25	.72	.93	.81	.89	.69	.76	.88	.62	.93	.64
Doesn't listen	.39	.79	.77	.58	.90	.59	.76	.70	.40	.92	.45
Loses things	.35	.62	.75	.50	.83	.44	.62	.72	.36	.88	.29
Physically dangerous	.16	.52	.99	.94	.84	.62	.48	.93	.62	.87	.52
Messy/sloppy[a]	.37	.76	.78	.58	.89	.58	.67	.70	.37	.89	.36
M[b]	.33	.78	.85	.70	.91	.68	.67	.76	.44	.90	.45
SD	.09	.13	.10	.13	.05	.08	.12	.10	.09	.03	.08
OD											
Loses temper	.21	.48	.90	.67	.81	.53	.81	.95	.81	.95	.79
Often argues	.17	.48	.96	.82	.82	.56	.67	.96	.82	.92	.70
Actively defies	.19	.41	.90	.63	.80	.43	.81	.98	.89	.95	.77
Annoys	.28	.55	.82	.55	.82	.56	.86	.86	.62	.96	.72
Blames others	.28	.59	.84	.59	.84	.55	.86	.86	.62	.96	.71
Touchy	.25	.59	.89	.68	.84	.59	.86	.91	.72	.96	.78
Angry/resents	.22	.48	.89	.64	.81	.49	.95	.98	.91	.99	.85
Spiteful	.15	.31	.92	.60	.77	.38	.67	.99	.93	.92	.69
Swears	.13	.28	.93	.62	.76	.39	.57	.99	.92	.90	.68
Bullies[a]	.19	.31	.86	.47	.76	.36	.52	.90	.58	.88	.57
M[b]	.21	.46	.89	.64	.81	.53	.78	.94	.80	.95	.72
SD	.05	.11	.04	.08	.03	.08	.12	.05	.12	.03	.06

Note. For purposes of comparison, the structure of this table is identical to Table 2 in Milich et al., 1987. ADHD = attention-deficit hyperactivity disorder. OD = oppositional defiant disorder. BR = base rate (subjects with symptom divided by total subjects); SEN = sensitivity (subjects with disorder who have symptom divided by subjects with disorder); SPE = specificity (subjects without disorder who do not have symptom divided by subjects without disorder); PPP = positive predictive power (subjects with symptom who have disorder divided by subjects with symptom); NPP = negative predictive power (subjects without symptom who do not have disorder divided by subjects without symptom); PB = correlation of symptom and total number of symptoms of the particular diagnosis (symptom is excluded from total symptoms for its own diagnosis). All correlations significant, $p < .001$, except for "loses things," $p = .002$. [a] Statistics for the symptoms "messy/sloppy" and "bullies" are included in the table for purposes of comparison, as these symptoms appeared in an earlier draft of *DSM–III–R* used in field trials. These symptoms were not included in the sum of symptoms used to calculate point-biserial correlations, nor were values for their conditional probabilities used in computing *M* and *SD*. [b] Median point-biserial correlations are reported, whereas means are reported for all other indices.

infrequently (base rates were .16, .23, .22, and .25, respectively); identified a moderate proportion of ADHD children (sensitivities were .52, .69, .62, and .72, respectively); and were highly specific to the disorder (specificities were .99, .96, .95, and .93, respectively). Although the NPPs for these four symptoms were high, other ADHD symptoms were more optimal as exclusion criteria for this disorder. The absence of the symptoms "shifts activities," "easily distracted," "fidgets," and "difficulty sustaining attention" were strongly predictive of the absence of ADHD (NPPs were .98, .96, .95, and .95, respectively). Interestingly, Milich et al. also found the similar "easily distracted" *DSM–III* symptom to have very high NPP, although the symptom "shifts activities" did not.

Four of the OD symptoms appeared to be particularly useful as both inclusion and exclusion criteria for OD in this sample. The symptoms "spiteful," "swears," "angry/resents," and "actively defies" were highly useful as both inclusion and exclusion criteria (PPPs were .93, .92, .91, and .89, respectively; NPPs

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

were .92, .90, .99, and .95, respectively). These symptoms occurred relatively infrequently (base rates were .15, .13, .22, and .19, respectively); identified a moderate to high proportion of boys with OD (sensitivities were .67, .57, .95, and .81, respectively); and were highly specific to OD (specificities were .99, .99, .98, and .98, respectively). A number of other OD symptoms, although not as useful for identifying the presence of OD, were also very effective exclusion criteria. Absence of the symptoms "annoys," "touchy," "blames others," and "loses temper" reliably indicated the absence of the disorder (NPPs were .96, .96, .96, and .95, respectively), although these symptoms were only moderately useful as inclusion criteria (PPPs were .62, .72, .62, and .81, respectively).

The statistics in Table 1 also demonstrate that several ADHD and OD symptoms possess good validity for the experimental diagnosis of the other disorder. One OD symptom, "often argues," was both a useful inclusion and exclusion criterion for ADHD, as its presence was a strong indicator of the presence of ADHD (PPP = .82), and its absence was a strong indicator of the absence of ADHD (NPP = .82). Although none of the other OD symptoms were especially useful as exclusion criteria for ADHD, several (e.g., "touchy" and "loses temper") might serve as effective inclusion criteria, as their PPPs for ADHD were almost as high as the average of the ADHD symptoms for this disorder. The opposite appeared to be true in examining the utility of the ADHD symptoms for the experimental diagnosis of OD. Although none of the ADHD symptoms were especially effective as inclusion criteria for OD, several (e.g., "fails to finish chores" and "often interrupts") were effective exclusion criteria, as their NPPs for OD were almost as high as the average of the OD symptoms for this disorder.

Discussion

We examined the utility of symptoms for the experimental diagnoses of OD and ADHD using the conditional probability indices of PPP, NPP, sensitivity, and specificity. We found considerable overlap between ADHD and OD, consistent with previous research examining the association among childhood externalizing problems (e.g., Hinshaw, 1987). Nevertheless, the results also indicate the capacity of some ADHD and OD symptoms to differentiate between these disorders. This ability to discriminate among psychopathological syndromes is a prerequisite for their validation (Robins & Guze, 1970).

It is of interest that our PPPs and NPPs were substantially higher than those reported by Milich et al. This does not appear to be a function of the base rates of the disorders, as these were either equal to or lower than those of Milich and colleagues. A more plausible explanation is that these differences are due to our use of teacher ratings, as opposed to parental interview data, as teachers appear to be a better source of information regarding some childhood behavioral problems than are parents (Ross & Pelham, 1978).

In contrast to the findings of Milich et al., many of the ADHD and OD symptoms were particularly useful as both inclusion and exclusion criteria for their respective experimental diagnoses. Interestingly, these symptoms were not always among those most highly correlated with their experimental diagnosis. Another interesting observation is that some of these ADHD and OD symptoms are not among those listed in *DSM-III-R* as most discriminating for their own disorder; in fact, a number are listed as among the least discriminating (e.g., "physically dangerous" for ADHD and "spiteful" for OD). It is worth noting that PPP and NPP were not used in the selection of individual symptoms for these diagnoses in *DSM-III-R* (Spitzer, Davies, & Barkley, 1990). Like Milich et al., we found that some ADHD and OD symptoms were quite useful in the experimental diagnosis of the other disorder. For example, the OD symptom "often argues" was useful as both an inclusion and an exclusion criterion for ADHD. A number of OD symptoms appeared to be useful inclusion criteria for ADHD (e.g., "touchy"), whereas a number of ADHD symptoms appeared to be useful exclusion criteria for OD (e.g., "fails to finish chores").

A methodological difficulty that frequently arises in studies of the relation between symptoms and diagnoses concerns the possible inflation of conditional probability indices due to chance agreement between the symptoms and the diagnosis. For example, although it is known that PPP and NPP for a particular symptom frequently differ *across* samples as the base rate of the symptom and diagnosis change, the problem of comparing conditional probability indices for two symptoms whose base rates considerably differ *within* a sample appears less well known. This problem is highly similar to that of correcting for chance agreement between observers (Brennan & Prediger, 1981), which led to the use of the kappa coefficient (Cohen, 1960).

Although a number of authors have attempted solutions to the problem of chance agreement (e.g., Brennan & Prediger, 1981; House, House, & Campbell, 1981; Millon, Bockian, Tringone, Antoni, & Green, 1989), defining chance agreement is actually a complex issue. Despite the common use of subtracting the product of marginal frequencies from a conditional probability as a correction for chance agreement, two observers may agree in their marginal frequencies because they both possess knowledge regarding the true base rate of the event observed (Brennan & Prediger, 1981). Hence, subtracting the product of their marginals as a means of correcting for chance agreement may actually be penalizing the raters for their shared expertise. This problem is even more difficult in studies of agreement between symptoms and diagnoses because of the lack of their observational independence, as the presence of diagnostically useful symptoms often covaries strongly with the presence of their disorder.[6]

The present results indicate the utility of PPP and NPP for discerning the best symptoms for the diagnosis of ADHD and OD, as well as for their differential diagnosis. Moreover, PPP

[6] As suggested by a reviewer, we examined the possible effects of chance agreement by correcting each of the PPPs presented in Table 1 by subtracting from them the product of the relevant marginal proportions—in this case the base rates of occurrence for both the particular symptom and the experimental diagnosis being compared. We then correlated the original and corrected PPPs and found that the rank orderings of these were identical (r = .99 or 1.0) in all four combinations of ADHD and OD symptoms and experimental diagnoses.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and NPP could be used to facilitate the diagnostician's decision-making process. Symptoms with high NPP across multiple settings, for example, may be particularly effective screening criteria and could thus allow interviewers to "skip out" of unnecessary assessment questions.

The present results are also consistent with previous literature (Ross & Pelham, 1978) in suggesting the diagnostic utility of teacher assessments of symptoms as an adjunct to the much more commonly used parental assessments of child functioning. Data on nonreferred school samples are necessary for the proper interpretation of teacher-rated externalizing symptoms. Thus, our study represents a further step toward the inclusion of teacher ratings in the clinician's assessment armamentarium.

Nonetheless, it is important to bear several caveats in mind. First, differences between the present results and those of Milich et al. may be due to a number of factors, including the difference in samples (clinic-referred vs. nonreferred) and in the type of information gathered (parent interview data vs. teacher ratings). This highlights the importance of replication and the necessity of using conditional probability indices to study the relation of symptoms to diagnoses in a variety of populations. Second, there are limitations inherent in examining the predictive efficiency of symptoms for the diagnosis they determine. The ultimate worth of diagnoses, and of the symptoms that embody them, rests on their validation against criteria such as treatment response, biological markers, course, and outcome. Thus, if researchers construct diagnostic criteria consisting of symptoms with high NPPs and high PPPs, these criteria must be subject to such validation. Third, computing such statistics implicitly assumes that the diagnostic entities are taxa, rather than dimensions. Such assumptions should be directly investigated, ideally using taxometric procedures (e.g., Meehl & Golden, 1982), especially given the assumptions of dimensionality implicit in the numerous factor analyses of scales similar to the symptom lists used herein (e.g., Hinshaw, 1987). Finally, statistical methods are required for comparing the levels of conditional probabilities for different symptoms. Such methods have yet to be applied to this problem (T. A. Widiger, personal communication, May 30, 1989), but they are necessary for drawing accurate inferences regarding the relative diagnostic efficiency of symptoms.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

References

Achenbach, T. M., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85,* 1275–1301.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev.). Washington, DC: Author.

Atkins, M. S., Pelham, W. E., & Licht, M. H. (1985). A comparison of objective classroom measures and teacher ratings of attention deficit disorder. *Journal of Abnormal Child Psychology, 13,* 155–167.

August, G. J., & Garfinkel, B. D. (1989). Behavioral and cognitive subtypes of ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 739–748.

Baldessarini, R. J., Finkelstein, S., & Arana, G. W. (1983). The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry, 40,* 569–573.

Brennan, R. L., & Prediger, D. J. (1981). Coefficient kappa: Some uses, misuses, and alternatives. *Educational and Psychological Measurement, 41,* 687–699.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37–46.

Dodge, K. A. (1980). Social cognition and children's aggressive behavior. *Child Development, 51,* 162–170.

Ferguson, H. B., & Rapoport, J. L. (1984). Nosological issues and biological validation. In M. Rutter (Ed.), *Developmental neuropsychiatry* (pp. 369–384). London: Churchill Livingstone.

Halperin, J. M., Wolf, L. E., Pascualvaca, D. M., Newcorn, J. H., Healey, J. M., O'Brien, J. D., Morganstein, A., & Young, J. G. (1988). Differential assessment of attention and impulsivity in children. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 285–290.

Hinshaw, S. P. (1987). On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. *Psychological Bulletin, 101,* 443–463.

House, A. E., House, B. J., & Campbell, M. B. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment, 3,* 37–57.

Johnston, C., Pelham, W. E., & Murphy, H. A. (1985). Peer relationships in ADDH and normal children: A developmental analysis of peer and teacher ratings. *Journal of Abnormal Child Psychology, 13,* 89–100.

Lahey, B. B., Pelham, W. E., Schaughency, E. A., Atkins, A. S., Murphy, H. A., Hynd, G., Russo, M., Hartdagen, S., & Lorys-Vernon, A. (1988). Dimensions and types of attention deficit disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 330–335.

Ledingham, J. E. (1981). Developmental patterns of aggressive and withdrawn behavior in childhood: A possible method for identifying preschizophrenics. *Journal of Abnormal Child Psychology, 9,* 1–22.

Lilienfeld, S. O., & Waldman, I. D. (1990). The relation between childhood attention-deficit hyperactivity disorder and adult antisocial behavior reexamined: The problem of heterogeneity. *Clinical Psychology Review, 10,* 699–725.

Loney, J., & Milich, R. (1982). Hyperactivity, inattention, and aggression in clinical practice. In M. Wolraich & D. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 3, pp. 113–147). Greenwich, CT: JAI Press.

Masten, A. S., Morrison, P., & Pelligrini, D. S. (1985). A revised class play method of peer assessment. *Developmental Psychology, 21,* 523–533.

Meehl, P. E., & Golden, R. R. (1982). Taxometric methods. In P. C. Kendall & J. N. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–181). New York: Wiley.

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin, 52,* 194–216.

Milich, R., Widiger, T. A., & Landau, S. (1987). Differential diagnosis of attention deficit and conduct disorders using conditional probabilities. *Journal of Consulting and Clinical Psychology, 55,* 762–767.

Millon, T., Bockian, N., Tringone, R., Antoni, M., & Green, C. J. (1989). New diagnostic efficiency statistics: Comparative sensitivity and predictive/prevalence ratio. *Journal of Personality Disorders, 3,* 163–173.

Offord, D. R., Boyle, M. H., Szatmari, P., Rae-Grant, N. I., Links, P. S., Cadman, D. T., Byles, J. A., Crawford, J. W., Blum, H. M., Byrne, C., Thomas, H., & Woodward, C. A. (1987). Ontario Child Health Study: 2. Six-month prevalence of disorder and rates of service utilization. *Archives of General Psychiatry, 44,* 832–836.

Pelham, W. E., Atkins, M. S., Murphy, H. A., & White, K. S. (1981). Operationalization and validation of attention deficit disorders. Paper presented at the annual meeting of the Association for the Advancement of Behavior Therapy, Toronto, Ontario, Canada.

Pelham, W. E., & Bender, M. E. (1982). Peer relationships in hyperactive children: Description and treatment. In K. D. Gadow & I. Bialer (Eds.), *Advances in learning and behavioral disabilities* (Vol. 1, pp. 365–436). Greenwich, CT: JAI Press.

Rey, J. M., Bashir, M. R., Schwarz, M., Richards, I. M., Plapp, J. M., & Stewart, G. W. (1988). Oppositional disorder: Fact or fiction? *Journal of the American Academy of Child and Adolescent Psychiatry, 27,* 157–162.

Robins, E., & Guze, S. B. (1970). Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. *American Journal of Psychiatry, 126,* 983–987.

Ross, A. D., & Pelham, W. E. (1978). Child psychopathology. *Annual Review of Psychology, 32,* 278–342.

Singleton, L. C., & Asher, S. R. (1977). Peer preferences and social interaction among third-grade children in an integrated school district. *Journal of Educational Psychology, 69,* 330–336.

Spitzer, R. L., Davies, M., & Barkley, R. A. (1990). The *DSM–III–R* field trial of disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 690–697.

Waldman, I. D. (1988). *Relationships between non-social information processing, social perception, and social status in 7- to 12-year-old boys.* Unpublished doctoral dissertation, University of Waterloo, Waterloo, Ontario.

Waldman, I. D. (1990). *Aggressive children's hostile biases: The role of attention and impulsivity.* Manuscript submitted for publication.

Werry, J. S., Methven, R. J., Fitzpatrick, J., & Dixon, H. (1983). The *DSM–III* diagnoses of New Zealand Children. In R. L. Spitzer, J. B. Williams, & A. E. Skodol (Eds.), *International perspectives on DSM–III* (pp. 291–307). Washington, DC: American Psychiatric Press.

Widiger, T. A., Hurt, S. W., Frances, A., Clarkin, J. F., & Gilmore, M. (1984). Diagnostic efficiency and *DSM–III. Archives of General Psychiatry, 41,* 1005–1012.

Received May 7, 1990
Revision received April 19, 1991
Accepted April 29, 1991 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Call for Nominations for *Psychology, Public Policy, and Law*

The Publications and Communications (P&C) Board has opened nominations for the editorship of *Psychology, Public Policy, and Law,* a new journal in development by APA. The journal will include articles that integrate and critically evaluate existing areas of research and original large-scale empirical research with significant public policy and legal implications.

Candidates must be members of APA and should be available to start receiving manuscripts in the late spring of 1992. Please note that the P&C Board encourages more participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. To nominate candidates, prepare a statement of one page or less in support of each candidate. Submit nominations to

Howard E. Egeth
Department of Psychology
Johns Hopkins University
Charles & 34th Streets
Baltimore, Maryland 21218

Other members of the search committee are Shari S. Diamond, J. Thomas Grisso, and Felice J. Levine. First review of nominations will begin December 15, 1991.