

# Prospective Follow-Up of Girls With Attention-Deficit/Hyperactivity Disorder Into Adolescence: Evidence for Continuing Cross-Domain Impairment

Stephen P. Hinshaw, Elizabeth B. Owens, Nilofar Sami, and Samantha Fargeon  
University of California, Berkeley

The authors performed 5-year prospective follow-up (retention rate = 92%) with an ethnically diverse sample of girls, aged 11–18 years, who had been diagnosed in childhood with attention-deficit/hyperactivity disorder (ADHD;  $N = 140$ ) and a matched comparison group ( $N = 88$ ). Hyperactive-impulsive symptoms were more likely to abate than inattentive symptoms. Across multiple domains of symptoms and functional impairment, girls with ADHD continued to display deficits of moderate to large effect size in relation to the comparison girls, but few differences emerged between the inattentive versus combined types. Follow-up effects withstood statistical control of crucial covariates for most outcomes, meaning that there were specific effects of childhood ADHD on follow-up status; in other instances, baseline disruptive disorders accounted for adolescent effects. For outcomes identical at baseline and follow-up, girls with ADHD showed more improvement across time than comparison girls (except for math achievement). Overall, ADHD in girls portends continuing impairment 5 years after childhood ascertainment.

*Keywords:* attention-deficit/hyperactivity disorder (ADHD), girls, prospective follow-up, longitudinal, impairment

It is now well known that ADHD exists in girls and that it yields considerable impairment in female samples (Gaub & Carlson, 1997; Gershon, 2002). Yet information regarding the long-term course of ADHD in girls is lacking, meaning that little is known about whether girls with ADHD display continued symptomatology and impairment across the life span. Several key prospective studies have studied males exclusively (Biederman et al., 1996; Loney, Kramer, & Milich, 1981; Mannuzza, Klein, Bessler, Mallow, & LaPadula, 1998; Satterfield, Swanson, Schell, & Lee, 1994); others have extremely large ratios of male to female participants (e.g., Barkley, Fischer, Smallish, & Fletcher, 2002; see review by Waschbusch, 2002). Mannuzza and Klein (1999) cited one controlled, prospective study (Mannuzza & Gittelman, 1984), with a sample of 12 (up to 19 in Mannuzza & Klein, 2000).

A recent search for published prospective studies of girls with normal-range intelligence, including systematic control groups and at least a 5-year prospective interval, yielded six reports, with a combined sample of 102 girls with ADHD and 79 comparison girls (Barkley et al., 2002; Latimer et al., 2003; Mannuzza & Gittelman,

1984; Rasmussen & Gillberg, 2000; Weiss & Hechtman, 1993; Young, Hepinstall, Sonuga-Barke, Chadwick, & Taylor, 2005). Female subsamples were small enough that either tests of sex differences yielded extremely low statistical power or separate analyses for girls could not be performed. Furthermore, almost all participants were White, and follow-up of the inattentive type of ADHD (ADHD-I) was not performed.

Additionally, Dalsgaard, Mortensen, Frydenberg, and Thomsen (2002) found that 32% of 25 Scandinavian female participants with ADHD-related symptoms, as opposed to 21% of 183 similar male participants, had been admitted to psychiatric units 10–30 years after childhood, a statistically significant difference. However, the follow-back nature of the design and the subaverage IQ scores of the sample precluded definitive conclusions. Babinski, Hartsough, and Lambert (1999) followed up 230 boys and 75 girls, ascertained in the 1970s, into their mid 20s but with nonstandard ascertainment procedures and without systematic analysis of control participants. A sample of key importance for developmental psychopathology—the Dunedin, New Zealand, birth cohort—did not yield enough girls with ADHD for viable follow-up investigations (T. Moffitt, personal communication, August 2004). Lahey et al. (2004) have followed preschool-aged children with ADHD, plus matched comparison participants, via yearly prospective assessments. Although the 16 girls with ADHD have shown impairment through childhood, statistical power is low for sex comparisons and persistence of ADHD into adolescence is not yet known. The Multimodal Treatment Study of Children with ADHD featured 116 girls among its 579 participants (20%), initially diagnosed with ADHD-combined type (ADHD-C; MTA Cooperative Group, 1999a). Published findings on the girls have reported mainly an absence of Sex  $\times$  Treatment interactions for core

---

Stephen P. Hinshaw, Elizabeth B. Owens, Nilofar Sami, and Samantha Fargeon, Department of Psychology and Institute of Human Development, University of California, Berkeley.

Work on this project was supported by National Institute of Mental Health Grant 45064. We gratefully acknowledge the girls who have participated in our continuing study, their families, and the many graduate students and research assistants who have contributed their time and effort to making the project a reality.

Correspondence concerning this article should be addressed to Stephen P. Hinshaw, Department of Psychology, Tolman Hall #1650, University of California, Berkeley, CA 94720-1650. E-mail: hinshaw@berkeley.edu

outcomes (MTA Cooperative Group, 1999b; Owens et al., 2003). The large and well-characterized female sample of Biederman et al. (1999), with a sample of 140 plus 122 matched comparison girls, will yield important follow-up data.

In addition, considerable debate exists about the distinctiveness of youth with the ADHD-I from those with ADHD-C or ADHD-hyperactive-impulsive (HI) type (McBurnett, Pfiffner, & Frick, 2001; Milich, Balentine, & Lynam, 2001). Given the potential importance of ADHD-I for girls (American Psychiatric Association, 2000) and the almost complete absence of longitudinal research on this subcategory of ADHD in either sex, prospective follow-up is a key area for investigation and constitutes a major goal of this report.

Hinshaw (2002) presented systematic data on a large sample of preadolescent girls with ADHD ( $N = 140$ ), including ADHD-C ( $n = 93$ ) and ADHD-I ( $n = 47$ ), plus an age- and ethnicity-matched comparison group ( $n = 88$ ). All participated in extensive clinical assessments and naturalistic summer research programs. Across domains of clinical symptomatology (Externalizing, Internalizing) and impairment (academic, cognitive, peer-related, familial)—featuring multiinformant and in some cases objective indicators—the ADHD sample displayed marked problems relative to the comparison group, even with statistical control of IQ, demographic indicators, and comorbid disorders. Additional investigations with this sample have documented neuropsychological and executive deficits (Hinshaw, Carte, Sami, Treuting, & Zupan, 2002), problems with friendships (Blachman & Hinshaw, 2002), deficits in independent play and adult likability (Mikami & Hinshaw, 2003), relational aggression (Zalecki & Hinshaw, 2004), and hostile expressed emotion from caregivers (Peris & Hinshaw, 2003). Our present goal is to appraise continuing symptoms and impairment.

The dearth of prospective data on female samples means that hypotheses must be based on follow-up studies of boys with ADHD, theoretical accounts of ADHD in female samples, and longitudinal research with related populations. First, most boys with ADHD continue to show clinically significant symptoms and impairments through adolescence (e.g., Barkley, 2003), with hyperactive-impulsive (HI) symptomatology revealing a steeper decline than the inattentive symptom cluster (Hart, Lahey, Loeber, Applegate, & Frick, 1995). Second, the theoretical perspective known as the *gender paradox* (see Eme, 1992) posits that the sex in which a given disorder is less prevalent should show greater levels of symptoms and impairment than the sex in which the disorder is more prevalent. Although supportive data are sparse (Hinshaw & Blachman, 2005), we did predict that girls with ADHD would show a notable pattern of continuing problems over time. Third, Pajer (1998) showed that girls with conduct problems are, like boys, at risk for antisocial outcomes, but such girls display additional risk for Internalizing disorders (depression, suicide, somatization disorders), early mortality (often from violent causes), and the development of notable personal and social problems (Robins & Price, 1991; Zoccolillo, Pickles, Quinton, & Rutter, 1992). ADHD may be likely to yield such multifinal outcomes in girls (Cicchetti & Rogosch, 1996).

We hypothesized that, like boys, girls with ADHD would show a greater decline in HI than in inattentive symptoms from childhood and that they would be at high risk for both Internalizing (anxious, depressed) and Externalizing or disruptive comorbidity

across time, as well as being at high risk for substance abuse. For eating symptomatology, we made no predictions because of a dearth of relevant literature. We also predicted that impairments in important life domains (school achievement, peer relations, social skills, academic performance, service utilization) would persist, even with control of age, demographic characteristics, IQ, and relevant additional disorders at baseline. Indeed, because it is important to understand whether continuing impairments are specific to initial ADHD status or whether features co-occurring with ADHD in childhood help to explain problematic outcomes, we rigorously controlled potential confounding variables. The preadolescent age of the sample at baseline, its diversity, the childhood measures available as covariates, the inclusion of ADHD-I and ADHD-C, and the multiinformant, multimethod assessments are key features.

## Method

### *Overview of Procedures*

The participants from Hinshaw's (2002) study were invited to participate in a prospective follow-up investigation during the academic year between 4 and 5 years after their baseline participation. Three separate research summer camp programs had been conducted (1997, 1998, and 1999), so that girls and families from those summer camps were invited in the school years 2001–2002, 2002–2003, and 2003–2004, respectively. Each evaluation was designed to span two half-day, clinic-based assessment sessions. In several cases for which clinic participation was not possible, telephone interviews or home visits were performed. Priority was placed on obtaining multisource, multiinformant data across both symptoms and domains of functional impairment.

### *Participants*

During the baseline assessments in childhood, a multigated screening and diagnostic procedure was implemented to gather a viable sample of preadolescent girls with ADHD and an age- and ethnicity-matched comparison sample (see Hinshaw, 2002, for full details). Girls with ADHD were recruited through pediatricians, mental health centers, schools, and direct advertisement, and comparison girls were recruited through pediatricians, community centers, and direct advertisement. Preliminary rating scale criteria (parent and teacher) were intentionally set with liberal, sex-specific thresholds, in order to prevent premature exclusion of potentially eligible girls, but final study entry depended on the participant's having met full criteria for ADHD, through the parent-administered Diagnostic Interview Schedule for Children (4th ed., DISC-IV; Shaffer et al., 2000). Common comorbidities (oppositional defiant disorder [ODD], conduct disorder [CD], anxiety disorders, depression, learning disorders) were allowed. Comparison girls, screened to match the ADHD sample with respect to age and ethnicity, could not meet diagnostic criteria for ADHD via either adult ratings or structured interview criteria. Exclusion criteria were mental retardation, evidence of psychosis or overt neurological disorder, lack of English spoken in the home, and medical problems prohibiting summer camp participation. The summer programs were not therapeutic in nature but instead constituted enrichment programs, free of charge, allowing the participation of a diverse, largely referred sample.

At baseline, the girls spanned the ages of 6–12 years. The sample was ethnically diverse (53% White, 27% African American, 11% Latina, 9% Asian American). The clinic and summer camp procedures yielded multiinformant, multimethod data on both symptoms and a wide range of domains of functional impairment (see Hinshaw, 2002, for details). The follow-up evaluations were performed on 209 of the 228 participants (92%), who ranged in age from 11.3 to 18.2 years ( $M = 14.2$  years).

Reasons for nonparticipation included (a) loss of the family to all tracking efforts ( $n = 4$ ), (b) refusal of the family to participate ( $n = 5$ ), and (c) difficulty in scheduling assessments although the family had been contacted ( $n = 10$ ). Comparison of the retained sample with those lost to attrition revealed that, for 29 of 31 demographic, diagnostic, and symptom variables gathered at baseline, differences were not statistically significant. The two significant differences were found with respect to (a) single-parent versus dual-parent family status—53% of the 19 girls lost to attrition were from single-parent homes versus 28% of the 209 follow-up participants and (b) teacher-reported Internalizing scores—those lost to attrition had higher baseline scores, with a small effect size. Although the null hypothesis cannot, of course, be proven, the follow-up sample appears representative of the total sample.

## Measures

Follow-up assessment staff were highly trained, bachelor's-level research assistants or graduate students in clinical psychology. With two exceptions, they were entirely separate from those involved in baseline assessments; staff members were not informed of the baseline diagnostic status of participants. Whereas responses to interview questions and issues regarding medication status could clearly suggest ADHD, objective variables were included (e.g., academic testing; computerized structured interviews) and diagnostic status did change for some girls at follow-up (see Results). Thus, data should not be biased by any breaking of blinds.

Follow-up measures were selected to reflect functioning in terms of ADHD-related and other symptomatology as well as core domains of impairment and service utilization. Note that 57% of the ADHD-C sample and 44% of the ADHD-I sample had been receiving psychotropic medications within the year before the follow-up assessments ( $p > .05$ ). Regarding stimulants, 45% of the ADHD-C sample and 27% of the ADHD-I sample had been receiving stimulant medications at the time of the follow-up visit, a contrast just missing statistical significance ( $p = .05$ ). On rating scales and interviews, parents and teachers were asked to respond regarding ADHD-related symptomatology for periods during which the girls were not medicated with stimulants, and on one of the two assessment days, girls participated stimulant free.

Space permits only brief descriptions of core measures. Those repeated from baseline afford examination of change across time; additional measures were required for adolescent symptomatology and impairment. Table 1 provides a listing of domains and measures.

## Symptom Measures

*Diagnostic Interview Schedule for Children—4th edition (DISC-IV; Shaffer et al., 2000).* This is a well-validated, highly structured diagnostic interview yielding both categorical diagnoses and symptom counts for the major disorders in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., *DSM-IV*; American Psychiatric Association, 2000). It was administered to parents and separately to adolescents; we used parent reports for this article and feature disorders present within the preceding year (rather than lifetime diagnoses, which would have been problematic for a longitudinal investigation).

*Swanson, Nolan, and Pelham Rating Scale—4th edition (SNAP-IV; Swanson, 1992).* This parent and teacher rating scale includes a dimensionalized checklist of the nine *DSM* items for ADHD-I, the nine items for ADHD-HI, and the eight items for ODD, with each scored on a metric that ranges from 0 (*not at all*) to 3 (*very much*) metric. It has been used extensively in ADHD assessment and treatment research (e.g., MTA Cooperative Group, 1999a).

*Child Behavior Checklist and Teacher Report Form (CBCL, TRF; Achenbach, 1991a, 1991b).* These extensively used scales comprise eight narrowband factors as well as the broadband factors of Externalizing (Aggressive Behavior and Delinquent Behavior scales) and Internalizing

(Withdrawn, Somatic Problems, and Anxious/Depressed Behavior scales) symptoms. All CBCL scales have excellent internal consistency and test-retest reliability as well as validity. Each of the 113 constituent CBCL items is rated on a 0–2 metric; we used *T* scores in all analyses.

*Self-Reported Delinquency (SRD; Elliott, Huizinga, & Ageton, 1985).* This is a widely used, self-report measure of the frequency and characteristics of a variety of overt and covert antisocial activities. Similar to Elliott et al. (1985), we created a score reflecting the number of different types of antisocial acts committed of the 36 possible types from the SRD. This score indexes the variety of antisocial behavior within the previous 6 months rather than a frequency count of the number of acts committed. It correlates moderately ( $r =$  from .34 to .41) with key criterion measures (e.g., parent-reported CBCL Delinquency and TRF Delinquency).

*Children's Depression Inventory (CDI; Kovacs, 1992).* This is a widely used self-report instrument tapping symptoms of depression in youth. Its psychometric properties compare favorably with those from other instruments in the field, with internal consistency ranging from .71 to .87 and with test-retest reliability figures averaging .70 (Kovacs, 1992). Each of the 27 items is scored on a 0–2 metric. Although the CDI has yielded a single factor solution in many analyses, we addressed the possibility that some CDI items index behavioral or learning problems rather than depression per se by eliminating the six items most reflective of such tendencies. No differences in any finding emerged, so we report the total score herein.

*Substance Use Questionnaire (SUQ; Molina & Pelham, 2003).* The SUQ is a structured questionnaire and interview adapted and expanded from existing measures, including the Health Interview Questionnaire (Jessor, Donovan, & Costa, 1989) and the National Institute on Drug Abuse's National Household Survey of Drug Abuse (1992). The SUQ includes both lifetime exposure questions and quantity/frequency questions. Kappas for 2-week test-retest reliability for "ever trying" one of five substances averaged .84, ranging from .70 (cigarettes) to .91 (marijuana). We created a severity score, reflecting the variety of substances used within the past year and the frequency with which these were used. It shows moderately large correlations ( $r =$  from .45 to .53) with substance abuse or dependence symptom levels from the DISC-IV.

*Eating Disorders Inventory and Eating Attitudes Test (EDI-2; Garner, 1991; EAT-26, Garner, Olmstead, Bohr, & Garfinkel, 1982).* These are two well-validated, self-report measures of (a) symptomatology of eating disorders and (b) features related to eating pathology. We analyzed total scores from the EAT-26 and the subscales of Bulimia, Drive for Thinness, and Body Dissatisfaction from the EDI-2. The EAT score yields an alpha between .8 and .9, discriminating adolescents with anorexia nervosa from comparison youths (Garner & Garfinkel, 1979). Internal consistencies of these EDI-2 subscales range from .69 to .93, with a mean of .87; short-term test-retest reliabilities range from .77 to .97; validity is extensively documented (Garner, 1991).

## Measures of Functional Impairment

*Columbia Impairment Scale (CIS; Bird, 1999).* Parents rated, on a 0–4 metric, the extent of problems their adolescent was having across 13 items in the home, peer, and school domains. We used the total score. Bird (1999) reported that this score is internally consistent ( $\alpha = .89$ ) and reliable across time ( $r = .68$ ), that it shows convergent validity with other measures of psychological dysfunction, and that it discriminates clinical from community participants. It was a primary outcome in the MTA Study (Hinshaw et al., 1997).

*Social Skills Rating System (SSRS; Gresham & Elliott, 1990).* With parent and teacher forms, the SSRS scales tap cooperation, self-control, and assertiveness (skill domains) plus Internalizing and Externalizing psychopathology. We scored the 30 items tapping skill domains to form the Total Social Skills subscale, which is internally consistent ( $\alpha = .91$  parent; .95 teacher). This subscale differentiates clinical from control samples; criterion validity in the current sample is shown through correlations

Table 1  
*Functioning at Follow-Up of Girls with Two Types of Attention-Deficit/Hyperactivity Disorder and Girls in a Comparison Group Across Domains by Baseline Diagnostic Status*

Diagnosis	No Covariates															
	F	df	Comparison			Inattentive			Combined			p <sup>a</sup>	Effect Sizes <sup>b</sup>			Covariates p <sup>c</sup>
			n	M	SD	n	M	SD	n	M	SD		0-1	0-2	1-2	
ADHD symptoms	19.49***	8, 278														
P SNAP-I (R)			78	0.5 <sup>a</sup>	0.5	40	2.1 <sup>b</sup>	0.7	81	1.9 <sup>b</sup>	0.8	.000	1.62	1.46	0.15	.000
P SNAP-HI (R)			78	0.1 <sup>a</sup>	0.2	40	0.8 <sup>b</sup>	0.5	81	1.2 <sup>c</sup>	0.8	.000	0.83	1.41	0.58	.000
T SNAP-I (R)			55	0.4 <sup>a</sup>	0.5	36	1.2 <sup>b</sup>	0.8	63	1.2 <sup>b</sup>	0.8	.000	1.01	0.95	0.06	.004
T SNAP-HI (R)			55	0.1 <sup>a</sup>	0.2	36	0.4 <sup>a</sup>	0.4	63	0.7 <sup>b</sup>	0.8	.000	0.41	0.91	0.51	.055
Externalizing symptoms	8.23***	10, 280														
P DISC-IV ODD (%)			6	7.4 <sup>a</sup>		21	51.2 <sup>b</sup>		43	50.6		.000	13.13 <sup>b</sup>	12.80 <sup>b</sup>	0.98 <sup>b</sup>	.010
P DISC-IV CD (%)			1	1.2 <sup>a</sup>		2	4.9 <sup>a</sup>		15	17.6 <sup>b</sup>		.001	4.10 <sup>b</sup>	17.14 <sup>b</sup>	4.18 <sup>b</sup>	.008
P CBCL Externalizing (R)			79	45.8 <sup>a</sup>	9.9	40	58.8 <sup>b</sup>	8.4	84	63.0 <sup>b</sup>	12.4	.000	0.98	1.29	0.32	.000
T TRF Externalizing (R)			56	49.0 <sup>a</sup>	8.3	36	55.3 <sup>b</sup>	9.0	62	58.3 <sup>b</sup>	11.1	.000	0.60	0.89	0.29	.033
S SRD Total			81	0.8	1.3	40	1.1	1.8	84	0.9	1.9	.712	0.16	0.07	0.09	—
Internalizing symptoms	4.02***	10, 282														
P DISC-IV Anxiety (%)			2	2.5		4	9.8		9	10.6		.103	4.27 <sup>b</sup>	4.68 <sup>b</sup>	1.09 <sup>b</sup>	—
P DISC-IV Depression (%)			3	3.7		4	9.8		10	11.8		.154	14.86 <sup>b</sup>	3.47 <sup>b</sup>	1.23 <sup>b</sup>	—
P CBCL Internalizing (R)			79	45.7 <sup>a</sup>	10.9	40	56.7 <sup>b</sup>	9.8	84	56.6 <sup>b</sup>	12.0	.000	0.89	0.87	0.02	.005
T TRF Internalizing (R)			56	49.8 <sup>a</sup>	8.7	36	54.5 <sup>b</sup>	10.4	62	54.8 <sup>b</sup>	8.3	.006	0.51	0.54	0.03	.519
S CDI Total (R)			82	5.3 <sup>a</sup>	5.7	40	6.5 <sup>a,b</sup>	4.0	84	7.8 <sup>b</sup>	6.5	.016	0.21	0.43	0.22	.412
Substance abuse/dependence	2.62*	4, 400														
P DISC-IV Sub A/D (%)			1	1.2 <sup>a</sup>		0	0.0 <sup>a</sup>		6	7.1 <sup>a</sup>		.047	—	6.08 <sup>b</sup>	—	.266
S SUQ Severity			82	-0.1	0.6	40	0.2	1.1	83	0.0	1.0	.160	0.33	0.11	0.22	—
Eating disorder symptoms	2.31*	8, 398														
S EAT Total			82	45.5 <sup>a</sup>	11.2	40	45.3 <sup>a,b</sup>	11.2	83	53.1 <sup>b</sup>	24.6	.012	0.01	0.42	0.43	.005
S EDI Bulimia			82	9.8 <sup>a</sup>	2.9	40	10.2 <sup>a,b</sup>	3.3	84	11.7 <sup>b</sup>	5.0	.005	0.10	0.47	0.37	.081
S EDI Drive for Thinness			82	12.6 <sup>a</sup>	5.6	40	13.4 <sup>a,b</sup>	4.8	84	15.8 <sup>b</sup>	8.0	.007	0.12	0.48	0.36	.060
S EDI Body Dissatisfaction			82	21.1 <sup>a</sup>	8.9	39	24.4 <sup>a</sup>	8.7	83	24.7 <sup>a</sup>	11.2	.046	0.33	0.36	0.03	.370
General impairment	42.11***	2, 208														
P CIS (R)			82	0.6 <sup>a</sup>	0.5	41	1.4 <sup>b</sup>	0.5	85	1.5 <sup>b</sup>	0.8	.000	1.07	1.20	0.13	.020
Social skills	12.93***	4, 294														
P SSRS Total			79	1.5 <sup>a</sup>	0.2	40	1.2 <sup>b</sup>	0.3	85	1.1 <sup>b</sup>	0.3	.000	0.88	1.18	0.29	.014
T SSRS Total			56	1.6 <sup>a</sup>	0.3	36	1.3 <sup>b</sup>	0.3	63	1.3 <sup>b</sup>	0.4	.000	0.73	0.73	0.00	.002
Peer relationships	7.70***	8, 268														
T Dishion Social Pref. (R)			56	3.1 <sup>a</sup>	1.3	34	2.3 <sup>a</sup>	1.6	62	1.3 <sup>b</sup>	2.6	.000	0.37	0.83	0.46	.041
S SRI Delinquent Peers			83	3.2	5.5	40	4.6	7.7	79	3.1	4.1	.329	0.25	0.02	0.27	—
P SRQ Friendship			78	1.0 <sup>a</sup>	0.4	39	0.5 <sup>b</sup>	0.6	85	0.3 <sup>b</sup>	0.7	.000	0.76	1.08	0.32	.169
P SRQ Peer Conflict			78	0.1 <sup>a</sup>	0.2	39	0.4 <sup>b</sup>	0.4	85	0.7 <sup>c</sup>	0.7	.000	0.44	1.07	0.62	.002
Achievement	11.34***	6, 294														
O WIAT Math (R)			81	112.9 <sup>a</sup>	14.2	39	95.6 <sup>b</sup>	15.5	84	93.5 <sup>b</sup>	16.8	.000	0.96	1.08	0.12	.001
O WIAT Reading (R)			81	107.7 <sup>a</sup>	8.2	39	97.8 <sup>b</sup>	13.1	84	98.1 <sup>b</sup>	11.5	.000	0.85	0.82	0.03	.073
T TRF Acad. Perf. (R)			56	54.0 <sup>a</sup>	9.5	36	44.1 <sup>b</sup>	8.4	60	43.8 <sup>b</sup>	8.2	.000	0.99	1.02	0.03	.070
Self-perceptions	4.92***	6, 402														
S Harter Self-Worth			82	3.4	0.5	40	3.3	0.6	83	3.2	0.7	.054	0.16	0.32	0.16	—
S Harter Social (R)			82	3.4 <sup>a</sup>	0.5	40	3.2 <sup>a,b</sup>	0.5	83	3.1 <sup>b</sup>	0.7	.011	0.33	0.50	0.17	.299
S Harter Scholastic (R)			82	3.2 <sup>a</sup>	0.6	40	2.8 <sup>b</sup>	0.5	83	2.8 <sup>b</sup>	0.6	.000	0.62	0.62	0.00	.107
Service utilization	28.25***	4, 404														
P School services (%)			11	13.4 <sup>a</sup>		32	82.1 <sup>b</sup>		66	78.6 <sup>b</sup>		.000	29.51 <sup>b</sup>	23.67 <sup>b</sup>	0.80 <sup>b</sup>	.000
P Nonschool service (%)			26	31.7 <sup>a</sup>		28	71.8 <sup>b</sup>		59	70.2 <sup>b</sup>		.000	5.48 <sup>b</sup>	5.08 <sup>b</sup>	0.93 <sup>b</sup>	.703

Note. 0 = Comparison, 1 = Inattentive, 2 = Combined; P = parent report, T = teacher report, S = self report, O = objective test; R = identical measure used at baseline and follow-up for repeated measurement. Higher means indicate worse functioning, except for Dishion Social Preference (Dishion Social Pref.) and Social Relationships Questionnaire (SRQ) Friendship measures, measures in Social Skills and Self-Perceptions domains, and Wechsler Individual Achievement Test (WIAT) scores; ADHD = attention-deficit/hyperactivity disorder; SNAP = Swanson, Nolan, and Pelham rating scale; I = inattention; HI = hyperactivity/impulsivity; DISC-IV = Diagnostic Interview Schedule for Children 4.0-Parent Report; CBCL = Child Behavior Checklist; TRF = Teacher Report Form; ODD = oppositional defiant disorder; CD = conduct disorder; SRD = Self-Report of Delinquency; CDI = Child Depression Inventory; Sub. A/D = substance abuse or dependence; SUQ = Substance Use Questionnaire; EAT = Eating Attitudes Test; EDI = Eating Disorders Inventory; CIS = Columbia Impairment Scale; SSRS = Social Skills Rating System; SRI = Social Relationships Interview; Acad. Perf. = academic performance.

<sup>a</sup> Significance: One-way analysis of variance (ANOVA) for continuous variables; Pearson chi-square statistic for categorical variables. Means with different superscripts in a given row differ significantly, according to the results of Tukey's post hoc comparisons or 2 × 2 chi-square tests.

<sup>b</sup> Effect size type: Cohen's *d* for continuous variables; odds ratios for categorical variables. For subgroups with zero percentages, no odds ratio exists.

<sup>c</sup> Significance: One-way ANCOVA for continuous variables; Wald statistic from logistic regression for categorical variables. Baseline covariates included age, family income, maternal education, child IQ, ODD or CD diagnosis (except when predicting Externalizing problems), any anxiety or depressive diagnosis (except when predicting Internalizing problems), reading disorder (except when predicting reading achievement), and medication status.

\*  $p < .05$ . \*\*\*  $p < .001$ .



between the parent scale and CBCL Social Competence subscale ( $r = .65$ ) and between the teacher scale and TRF Behaving Appropriately subscale ( $r = .66$ ).

*Dishion Social Preference Scale (Dishion, 1990).* This is a three-item, teacher-completed measure of the proportion of peers who accept, reject, and ignore the adolescent in question, with each item rated on a 5-point scale. Dishion has reported moderately strong correlations of these items with peer-derived sociometric indicators. From these ratings, we derived a widely used and well-validated social preference score (Coie, Dodge, & Coppotelli, 1982; Lahey et al., 2004) by subtracting the *reject* rating from the *accept* rating.

*Social Relationships Interview.* This new interview includes items related to deviant peers, friendships, and romantic relationships. Using items based on the Peer Delinquency Scale (Loeber, Stouthamer-Loeber, van Kammen, & Farrington, 1994), we obtained the participant's count of the number of their peers engaging in 15 types of antisocial behavior. Additional questions were based on conceptual models of friendship attainment and social/dating relationships.

*Social Relationships Questionnaire.* This is a parent-reported measure of an adolescent's relationships with peers and friends containing 12 items, each of which is appraised on a 4-point metric. A principal components analysis of these items yielded two factors with eigenvalues greater than 1, accounting for 44% and 11% of the variance, respectively. An oblique rotation yielded two factors, each comprising 6 items, which we termed Peer Conflict ( $\alpha = .83$ ) and Friendship ( $\alpha = .77$ ). These scales correlate moderately with parent and teacher reports of problem behavior and social competence in the current sample.

*Wechsler Individual Achievement Test (WIAT; Wechsler, 1992).* We administered the Basic Reading and Math Reasoning subtests. Normed on the same sample as the Wechsler Intelligence Scale for Children, (3rd ed., WISC-III; Wechsler, 1991), the WIAT is a psychometrically sound, widely used test of achievement. Test-retest reliabilities for the Reading and Math subtest scores range from .85 to .92 (Wechsler, 1992).

*Teacher Report Form (Achenbach, 1991b).* The TRF contains a scale of teacher reports of performance below, at, or above grade level in various academic subjects.

*Self-Perception Profile for Adolescents (Harter, 1988).* This is an upward extension of Harter's extensively used scale for children; adolescents make self-reports of the extent to which they agree or disagree with statements reflecting perceived competence across several domains (scholastic, behavioral conduct, social, athletic, close friends, job, romantic relationships, physical appearance, and global self-worth). We analyzed Social Acceptance, Scholastic Competence, and Global Self-Worth subscales. As reported by Harter (1982), internal consistencies of these scales range from .75 to .84, with test-retest reliabilities ranging from .69 to .80.

### Service Measures

All families completed a year-by-year grid of the services their families and daughters had received from baseline to follow-up. For school services, we counted (1 vs. 0) the use of any special education, tutoring, or other school services across the 5-year interval. For nonschool services, we included (1 vs. 0) individual, group, or family therapy in the community.

### Covariates

To ascertain whether adolescent symptoms and impairments are related specifically to the girls' original ADHD status, we controlled for key measures of baseline status. These included, first, demographic information (family income, maternal education) and participant age (given the 6–7 year age span across the sample), which had distinguished subgroups at baseline. We also included full-scale IQ (from the WISC-III; Wechsler, 1991) and additional disorders (i.e., comorbidities for the girls with ADHD) from the parent DISC-IV, dummy coded as 1 versus 0 for the

presence versus absence of ODD or CD, for the presence versus absence of anxiety or mood disorders (note that anxiety disorders had to include the presence of one or more conditions beyond specific phobias), and for the presence versus absence of a reading disorder. IQ and comorbid diagnoses had not moderated ADHD versus comparison group differences with respect to our baseline criterion measures (see Hinshaw, 2002). We also included the presence versus absence of psychotropic medication use during the year preceding the follow-up interview (see Lahey et al., 2004). Because of the age span of the sample, we also created a three-level age variable related to the girls' age groups in the summer programs; it significantly moderated the effect of diagnostic status on only one outcome. Thus, age effects were not salient.

### Data Analytic Plan

All statistical analyses were performed with SPSS for Windows (Version 12; SPSS, 2003). Because cohort effects could complicate interpretation of findings, we performed an initial set of analyses of variance (ANOVAs) with the independent variable of the initial year of participation: 1997, 1998, or 1999. Only 4 of 35 analyses were significant, with small effects; we do not consider these further. The first primary analysis involved examination of the proportion of participants who were classified as comparison girls versus girls with one of the ADHD types at baseline and again at follow-up. For this categorical cross-classification, the diagnostic procedures used in Hinshaw (2002) to designate ADHD types constituted the baseline diagnoses; at follow-up, we performed a parallel procedure. That is, each of the 18 *DSM-IV* ADHD symptoms was considered present if endorsed on the DISC-IV or if the mother or teacher rated it as a 2 (*pretty much*) or 3 (*very much*) on the SNAP. Girls with at least 6 ADHD-I and 6 ADHD-HI symptoms (with at least 4 in each domain based on the DISC-IV; see Hinshaw et al., 1997) were designated as having ADHD-C; girls with at least 6 ADHD-I symptoms (with at least 4 based on the DISC-IV) but fewer than 6 ADHD-HI as having ADHD-I; girls with at least 6 ADHD-HI symptoms (with at least 4 based on the DISC-IV) but fewer than 6 ADHD-I as having ADHD-HI; and girls with fewer than 6 ADHD-I and 6 ADHD-HI symptoms as not having ADHD.

For the second set, we grouped outcomes into 11 domains (see Table 1). After obtaining a highly significant multivariate analyses of variance (MANOVA) across all 35 outcomes, we then performed 11 MANOVAs, one per domain. (For global impairment, the sole measure was the CIS.) The independent variable was baseline diagnostic subgroup, comprising the three levels of comparison, ADHD-I, and ADHD-C. For significant MANOVAs ( $\alpha < .05$ ), we then examined separate outcomes via ANOVAs plus Tukey's post hoc comparisons of each subgroup contrast. Power was high, even for the ADHD-C versus ADHD-I contrasts, for which our sample sizes yielded power between .65 and .80 to detect (2-tailed) a difference of medium effect size (Faul & Erdfelder, 1992). We displayed effect sizes (Cohen's  $d$ ), with the difference between means as the numerator and the pooled standard deviation as the denominator (Cohen, 1988). For categorical variables, we performed univariate analyses via 3 (subgroup)  $\times$  2 (present vs. absent) chi-square tests and decomposing significant findings into a series of 2  $\times$  2 chi-square tests for subgroup comparisons. Here, effect sizes were calculated as odds ratios.

We supplemented these with MANCOVAs for domains with significant MANOVA results, controlling for age, family income, maternal education, full-scale IQ, ODD/CD diagnoses (except when testing Externalizing outcomes), Internalizing diagnoses (except when testing Internalizing outcomes), reading disorder status (except when testing reading achievement), and the designation of having received psychotropic medication during the year preceding follow-up. We performed (a) ANCOVAs for each continuous dependent measure with a significant ANOVA or (b) hierarchical logistic regressions for each categorical dependent measure with a significant chi-square test. Note that exploratory analyses in which race or ethnicity was examined as a moderator revealed only four significant

effects across all outcomes; these had low cell sizes and idiosyncratic, difficult-to-interpret findings.

Finally, to address change across time, we examined separately the 16 continuous variables that were identical or nearly identical from baseline to follow-up (see in Table 1). We performed an initial multivariate, repeated measures MANOVA, with the independent variables of baseline diagnostic status and time, followed by univariate repeated measures ANOVAs. Of interest are the Diagnostic Status  $\times$  Time interactions, which would provide evidence for differential patterns of change across time for the groups.

## Results

### ADHD Status

Table 2 reveals the baseline versus follow-up diagnostic status of the girls in the follow-up assessments (among the 209, we lost 2 additional girls because of DISC-IV data errors). The vast majority of the comparison group at baseline maintained non-ADHD status at follow-up (77/81). Most girls classified as having ADHD-I at baseline maintained this classification at follow-up (26/41; 63%), with the remainder classified as comparison (10/41) or as having ADHD-C/HI (5/41), but less than half initially classified as having ADHD-C (33/85; 39%) maintained this status, with the remainder classified as comparison (29/85), as having ADHD-I (20/85), or as having ADHD-HI (3/85). There was greater persistence of ADHD-I than of ADHD-C,  $\chi^2(1, N = 126) = 4.91, p = .027$ .

### Symptom and Impairment Domains

The omnibus MANOVA was highly significant,  $F(70, 179) = 3.60, p = .000$ , (Pillai's trace = 1.18), and each of the 11 MANOVAs was significant as well (see Table 1). Regarding the follow-up univariate tests, we first note that the separate dependent measures were generally moderately correlated: Within domains, the average correlations between variables ranged from .24 to .52 (average  $r = .38$ ).

Regarding the ADHD symptom domain, all four outcome measures revealed significant ANOVAs. For maternal and teacher

SNAP-IV ADHD-I, girls with both ADHD types had significantly higher scores than did the comparison girls, with large to extremely large effect sizes, but did not differ from each other. For maternal and teacher SNAP-IV ADHD-HI, a stepwise pattern emerged, whereby comparison girls scored lowest, followed by girls with ADHD-I and then those with ADHD-C (the latter two differences were medium); but the teacher-rated comparison versus ADHD-I contrast was nonsignificant (incomplete data collection yielded lower sample sizes for teacher measures).

For the Externalizing domain, all measures except the SRD yielded significant omnibus findings. For ODD, both ADHD types had rates of approximately 50%, far higher than the comparison rate of 7%; for CD, the comparison and ADHD-I rates (1% and 5%, respectively) were lower than the ADHD-C rate (18%). For CBCL and TRF Externalizing, both ADHD types had higher scores than the comparison group (with medium to strong effects) but did not differ significantly from each other (these latter effect sizes were small:  $d = .32$  and  $.29$ , respectively).

For the Internalizing domain, DISC-IV diagnoses did not yield significant effects, but the CBCL and TRF Internalizing scales showed similar patterns: Girls with both ADHD types displayed higher scores than girls in the comparison group (with large effects for the CBCL and medium for the TRF) but were nearly indistinguishable from each other. The CDI revealed that the girls in the ADHD-C group reported higher scores than the girls in the comparison group (small to medium effect). For substance abuse or dependence, specific contrasts were nonsignificant. For eating symptomatology, the clearest finding was that the girls with ADHD-C scored significantly higher than the comparison girls (EAT Total, EDI Bulimia, EDI Drive for Thinness), with effect sizes ranging from .42 to .48.

Regarding impairment, on the CIS, girls with both ADHD types were far more impaired than the comparison girls ( $ds > 1$ ) but did not differ themselves. A parallel pattern emerged for social skills: Both girls with ADHD-C and girls with ADHD-I were rated as less skilled than the girls in the comparison group (effect sizes medium to large) but were themselves indistinguishable. For peer relationships, self-reported contact with delinquent peers did not significantly distinguish the groups, but for teacher-reported social preference, the ADHD-C group was more rejected than either the girls with ADHD-I or the comparison girls (latter effect large). Parents reported girls with each ADHD type to have fewer friends than comparison girls (large effects); for the peer conflict factor, comparison girls scored lowest, girls with ADHD-I next, and girls with ADHD-C highest, with all three contrasts significant (medium-large effects).

For academic performance, the girls with both ADHD types showed lower scores than the comparison girls (effect sizes large) but did not differ from each other (effect sizes miniscule). For self-perceptions, Global Self-Worth scores did not distinguish the groups; for Perceived Social Competence, girls with ADHD-C scored lower than the comparison girls (medium effect); and for Perceived Scholastic Competence, girls with both ADHD types scored lower than the comparison girls (medium effect) but did not differ themselves. For service utilization, fewer than one seventh of the comparison girls received special services at school between baseline and follow-up, compared with 79% of girls with ADHD-C and 82% of girls with ADHD-I. As for individual, group, or family therapy (outside school), the respective percent-

Table 2  
Attention-Deficit/Hyperactivity Disorder Status at Baseline and Follow-Up

Follow-up diagnosis	Baseline Diagnosis					
	No ADHD diagnosis ( <i>n</i> = 81)		ADHD-inattentive ( <i>n</i> = 41)		ADHD-combined ( <i>n</i> = 85)	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
No ADHD diagnosis	77	95	10	24	29	34
ADHD-inattentive	4	5	26	63	20	24
ADHD-hyperactive/impulsive	0	0	1	2	3	4
ADHD-combined	0	0	4	10	33	39

Note. ADHD = attention-deficit/hyperactivity disorder; *n* = no. at follow-up. Percentages are for diagnostic group at baseline; 2 of the 209 participants were excluded because of problems with the Diagnostic Interview Schedule for Children-IV data.

ages were 32%, 70%, and 72%, with both ADHD types revealing higher rates than the comparison girls but not differing between themselves.

*Covariates*

MANCOVAs revealed that significant diagnostic status findings remained for 7 of the 11 domains: ADHD symptoms, externalizing, global impairment, social skills, peer relations, academic achievement, and service utilization. Thus, in these areas, childhood ADHD yielded problems at follow-up despite stringent statistical control of age, demographic factors (family income, maternal education), child IQ, the presence of comorbid disorders at baseline, and current medication treatment status, suggesting specific effects of childhood ADHD related to key adolescent outcomes. Still, ANCOVAs revealed a few instances in which specific measures that had yielded significant findings from the ANOVAs were no longer significant (see Table 1).

For the four domains that did not survive the inclusion of covariates (Internalizing symptoms, substance abuse, eating symptoms, and self-perceptions), we tried to discern which covariates accounted for the loss of significance by conducting a series of MANCOVAs in which each covariate was considered individually. For Internalizing and eating symptoms and self-perceptions, baseline ODD or CD status was the main covariate responsible. For substance abuse or dependence, several covariates (ODD or

CD, Internalizing symptoms, IQ) led to nonsignificance. Overall, baseline ODD or CD was a major contributor to nonspecificity of prediction in selected domains.

*Change Across Time*

For the 16 continuous outcomes that were identical or highly parallel between baseline and follow-up, the repeated measures MANOVA yielded significant effects for diagnostic status,  $F(32, 212) = 11.91, p = .000$ ; time,  $F(16, 105) = 9.75, p = .000$ ; and the Status  $\times$  Time interaction,  $F(32, 212) = 3.49, p = .000$ . Follow-up ANOVAs revealed time effects for 11 outcomes and Status  $\times$  Time interactions for 10. Plots plus tests for simple effects (see Table 3) reveal that for WIAT Math, the girls with ADHD-C declined in performance, whereas the comparison girls improved (the apparent decline of the girls with ADHD-I was not significantly different from 0). For all but one of the remaining interactions, however, the effects were essentially opposite, such that girls with ADHD (particularly ADHD-C) showed more improvement than did the comparison girls. Notably, unlike the girls with ADHD, the comparison girls showed a significant increase in TRF Internalizing problems. Furthermore, with demographics, age, and comorbid disorders controlled, the Status  $\times$  Time interaction survived for 8 of the 10 outcomes (TRF Internalizing and Dishion Social Preference interactions were reduced to marginal significance). Even so, girls with ADHD were performing signif-

Table 3  
*Baseline to Follow-Up Changes by Baseline Attention-Deficit/Hyperactivity Disorder Diagnostic Status*

Dependent Measure	Comparison	Baseline to follow-up change scores		<i>p</i> <sup>a</sup>
		Inattentive	Combined	
1. P SNAP-I (Rev)	-0.08 <sup>a</sup>	<b>0.35<sup>b</sup></b>	<b>0.39<sup>b</sup></b>	.000
2. P SNAP-HI (Rev)	<b>0.14<sup>a</sup></b>	<b>0.35<sup>a</sup></b>	<b>0.88<sup>b</sup></b>	.000
3. T SNAP-I (Rev)	-0.14 <sup>a</sup>	<b>0.73<sup>b</sup></b>	<b>0.70<sup>b</sup></b>	.000
4. T SNAP-HI (Rev)	0.04 <sup>a</sup>	<b>0.25<sup>a</sup></b>	<b>0.83<sup>b</sup></b>	.000
5. P CBCL Externalizing (Rev)	0.0 <sup>a</sup>	0.18 <sup>a</sup>	<b>5.6<sup>b</sup></b>	.001
6. T TRF Externalizing (Rev)	-0.2 <sup>a</sup>	3.3 <sup>a</sup>	<b>9.7<sup>b</sup></b>	.000
7. P CBCL Internalizing (Rev)	1.0	<b>3.4</b>	<b>4.4</b>	.146
8. T TRF Internalizing (Rev)	<b>-3.6<sup>a</sup></b>	4.3 <sup>b</sup>	<b>3.1<sup>b</sup></b>	.002
9. S CDI Total (Rev)	-0.7	2.3	1.3	.077
10. P CIS (Rev)	-0.12 <sup>a</sup>	0.18 <sup>b</sup>	<b>0.49<sup>c</sup></b>	.000
11. T Dishion Soc. Pref.	-0.32 <sup>a</sup>	0.48 <sup>a,b</sup>	<b>0.98<sup>b</sup></b>	.032
12. O WIAT Math	<b>2.8<sup>a</sup></b>	-2.4 <sup>a,b</sup>	<b>-4.1<sup>b</sup></b>	.001
13. O WIAT Reading	<b>-4.7</b>	<b>-4.0</b>	<b>-3.2</b>	.582
14. T TRF Acad. Perf.	1.0	<b>5.2</b>	2.2	.147
15. S Harter Social	<b>0.20</b>	<b>0.32</b>	<b>0.27</b>	.691
16. S Harter Scholastic	0.01	0.11	-0.01	.720

*Note:* P = parent report, T = teacher report, S = self-report, O = objective test, SNAP = Swanson, Nolan, and Pelham rating scale; I = inattention; HI = hyperactivity/impulsivity; CBCL = Child Behavior Checklist; TRF = Teacher Report Form; CDI = Child Depression Inventory; CIS = Columbia Impairment Scale; Dishion Social Pref. = Dishion Social Preference; WIAT = Wechsler Individual Achievement Test. Measures 1–10 tap symptoms or impairment; Measures 11–16 tap skills or competence. We reverse scored Items 1–10 (indicated by “Rev”) for consistency of interpretation, so that positive values signify improvement. Change scores are follow-up minus baseline scores, meaning that positive values denote improvements in functioning across time and negative values denote declines. Change scores that are significantly different from zero are placed in bold; scores with different superscripts in a given row differ significantly according to Tukey’s post hoc comparisons. <sup>a</sup>Significance of the *F* value associated with the Time  $\times$  Diagnostic Status interaction in repeated measures analysis of variance.

icantly worse than comparison girls on all of the outcomes showing Status  $\times$  Time interactions, with effect sizes ranging from moderate to very large.

### Discussion

We prospectively followed a large, diverse, and carefully diagnosed sample of preadolescent girls with ADHD (retention rate = 92%), 5 years after their diagnosis, which had occurred at 6–12 years of age. Our findings were as follows: (a) Although the comparison group almost invariably maintained their non-ADHD status and although a majority of girls with ADHD–I kept this classification, over half the girls with ADHD–C had “lost” enough HI symptoms that their adolescent diagnoses reverted to ADHD–I or even to nonclinical status. Thus, symptoms of HI abated more sharply than did those related to inattention. (b) Girls with baseline ADHD had significantly more problems at follow-up than did girls in the comparison group across all 11 domains of symptoms and impairments that we examined, with effect sizes ranging from medium to large (slightly smaller, on average, than those present at baseline; see Hinshaw, 2002). (c) Repeated measures analyses for 16 variables that were identical or highly parallel at baseline and follow-up indicated that comparison girls stayed level or slightly worsened, whereas the girls with ADHD (particularly the ADHD–C type) actually improved in some instances; the exception was WIAT Math, for which the ADHD–C group showed declining performance across time. Such ADHD-related improvement did not, however, erase the large ADHD-related deficits that had been present since baseline. (d) Significant ADHD–C versus ADHD–I differences at follow-up were rarely found; the exceptions (diagnoses of conduct disorder and indicators of peer rejection and conflict) revealed that girls with ADHD–C showed greater problems than those with ADHD–I. (e) For most domains, ADHD-related deficits survived stringent statistical control of age, baseline demographics and comorbidity, IQ, and medication status during the preceding year. Thus, many effects of childhood ADHD on adolescent impairment appear specific. For other domains, effects dissipated with such statistical control, particularly of baseline ODD or CD. Overall, the key conclusion is that ADHD in girls portends noteworthy problems 5 years later.

Space permits only brief elaboration of the major findings. First, as predicted, we found that ADHD–C was less stable over time than ADHD–I because of informant-reported abatement of HI symptoms (see Hart et al., 1995). Note, however, that when categorical classifications of ADHD are monitored yearly (e.g., Lahey et al., 2004), measurement error and regression to the mean occur. In other words, classifications may fluctuate as a function of changes in just one or two symptoms. It will take additional follow-up into late adolescence and adulthood to ascertain the ultimate stability of our diagnostic classifications from childhood (see Barkley et al., 2002).

Second, our key finding was that the girls with childhood-diagnosed ADHD continued to show greater psychiatric symptomatology across multiple symptom areas (ADHD, Externalizing, Internalizing, eating, substance abuse or dependence) and larger functional impairments (global, social skills, peer relations, academic performance, self-perceptions, and service utilization rates) than did comparison girls. A major question pertains to the clinical significance of the ADHD–comparison differences during our

follow-up. Despite general improvement in functioning of the girls with ADHD over time, some indicators for the ADHD sample (e.g., Dishion scores, which indicated substantial levels of peer rejection; rates of service utilization spanning 70%–80% between baseline and follow-up; large deficits in academic achievement) reveal clear evidence of major problems during adolescence. In terms of clinical implications, these findings indicate that ADHD in girls is a problem of real importance; services may well be required in most instances. Indeed, this investigation affirms the public health significance of ADHD in girls, given the likelihood of persisting impairments in crucial domains through 5-year follow-up.

Third, just as at baseline (Hinshaw, 2002; Hinshaw et al., 2002), ADHD–C versus ADHD–I differences were rarely significant and almost always of small effect size. One explanation may be that standard definitions of the inattentive form of ADHD are not sufficiently stringent, in that a child with near-threshold levels of HI symptoms could still be classified as ADHD–I. To rule out this possibility, we performed additional analyses using a more restricted definition of childhood ADHD–I (see Hinshaw, 2002), featuring a low threshold of HI symptoms plus the presence of sluggish cognitive tempo symptoms from the SNAP–IV. Yet this group with restricted ADHD–I was virtually identical to the remainder of the girls with *DSM-IV*-defined ADHD–I on all outcomes; like the “main” ADHD–I subgroup, the group did not yield a clear divergence from the ADHD–C sample. Thus, at first glance, our data do not appear to support the contention of Milich et al. (2001) that the inattentive form of ADHD is a qualitatively distinct variant.

Still, this issue deserves further scrutiny. Note that the domains that did reveal subtype differences were, in addition to HI symptomatology (an artifact of sample definition), comorbid conduct disorder and peer rejection or conflict, for which the ADHD–C group was more impaired at follow-up. These are clinically meaningful outcomes, and the relatively better status of girls with ADHD–I in these areas is noteworthy. From another perspective, it could well be the case that two (or more) conditions yield evidence for similar patterns of impairment, even though fundamentally different etiologic processes led to these conditions. Our data do not, obviously, address this possibility; it may be that at least some individuals with ADHD–I are distinct from those with ADHD–C with respect to key risk factors or causal variables despite overall similarity at follow-up. The distinctiveness of diagnostic subcategories of ADHD is an unresolved issue.

Fourth, given the presence of 35 separate outcomes, which we retained in the interest of clinical interpretability, were there informant effects? The majority of both parent-reported outcomes (10 of 15) and teacher-reported outcomes (4 of 7) yielded significance (even though medication issues may have reduced the sensitivity of teacher reports), compared with only 2 of 11 self-reported outcomes (and 1 of 2 objective tests). Parents, who provided the crucial data for sample ascertainment at baseline, continued to detect adolescent problems for most variables. Teachers, who were clearly independent of any baseline measurements, also revealed effects (as did the WIAT Math subtest). As might be expected given the relatively poor self-monitoring of individuals with ADHD, the girls themselves were far less likely to disclose problematic functioning (see Barkley et al., 2002). An exception pertained to the CDI, for which the ADHD–C group self-reported



greater levels of depressive symptomatology than did comparison girls at follow-up.

Fifth, the specificity of impairment from early ADHD was shown by the analyses of covariance, which revealed that for ADHD and Externalizing symptoms, as well as for global impairment, social skills, peer relations, academic performance, and service utilization, diagnostic group differences survived stringent statistical control. The covariates analyzed (age, family income, maternal education, IQ scores, presence of other diagnoses, and medication status) are all important correlates of adolescent status; controlling them provided a stringent test. Thus, for these outcomes, it appears that early ADHD itself is the major factor that is associated with continuing impairment. For the other outcomes, our statistical controls rendered ADHD-related group differences nonsignificant, with baseline ODD or CD most likely to reduce levels of significance. Hence, comorbidity is still an important aspect of the psychopathology of ADHD.

Overall, along with a host of cross-sectional findings (see Gaub & Carlson, 1997; Gershon, 2002, for reviews), the present longitudinal results provide clear evidence for the impairments related to ADHD in female samples. Indeed, the current findings show that the risk for ADHD-related psychiatric problems and functional impairments extends at least until early to mid adolescence. Continued follow-up of the present sample (as well as other carefully ascertained female samples) into adulthood is a priority on both clinical and conceptual grounds.

Study limitations include, first, the nonrepresentative nature of the present ADHD sample. As explained in Hinshaw (2002), however, we chose not to obtain our sample via epidemiologic methods at study entry, given our objective of involving a large female sample in summer research programs that could yield an objective and multisource database. The use of multiple recruitment sources and the diversity of the sample suggest that the girls reflect reasonably well the nature of ADHD in the San Francisco Bay Area. In addition, the sample did not include the ADHD-HI type at baseline, given that (a) this category appears to be concentrated in samples of young children (Lahey et al., 1994) and (b) we wished to preserve statistical power for the crucial contrast of the inattentive versus combined types of ADHD. Next, although our follow-up data collection efforts yielded an acceptable retention rate of participants (92%), we lacked teacher data for 26% of the follow-up sample. Next, some parents of girls on nonstimulant medications and a number of teachers who did provide reports on participants could not accurately appraise behavior patterns during periods in which the girls were not medicated. If anything, this problem would tend to provide a conservative bias on our results, potentially leading to underestimating the problem behavior and impairments in some girls with ADHD. Finally, although the diversity of the sample can be considered a strength, racial and ethnic subgroups within ADHD types were small in size, and we did not find meaningful moderation of key outcomes by race or ethnicity.

In all, the chief conclusion is that ADHD in girls portends continuing problems that are of substantial magnitude and that exist across multiple domains of symptoms and functional impairment. Such findings require replication but provide strong evidence of (a) the clinical impact of ADHD in female samples, (b) the public health importance of this condition in girls, and (c) the need for the continued search for underlying mechanisms (Hin-

shaw & Blachman, 2005). Greater understanding of ADHD in female samples, especially its longitudinal course, is a continuing research priority; the current evidence suggests that such understanding is crucial, given the nontransitory nature of key comorbidities and impairments.

## References

- Achenbach, T. M. (1991a). *Manual for the Child Behavior Checklist/4-18 and 1991 profile*. Burlington, VT: University Associates in Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Teacher Report Form and 1991 Profile*. Burlington, VT: University Associates in Psychiatry.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Babinski, L. M., Hartsough, C. S., & Lambert, N. M. (1999). Childhood conduct problems, hyperactivity/impulsivity, and inattention as predictors of adult criminal activity. *Journal of Child Psychology and Psychiatry*, *40*, 347-355.
- Barkley, R. A. (2003). Attention-deficit hyperactivity disorder. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (2nd ed., pp. 75-143). New York: Guilford Press.
- Barkley, R. A., Fischer, M., Smallish, L., & Fletcher, K. (2002). The persistence of attention-deficit/hyperactivity disorder into young adulthood as a function of reporting source and definition of disorder. *Journal of Abnormal Psychology*, *111*, 279-289.
- Biederman, J., Faraone, S. V., Mick, E., Williamson, S., Wilens, T., Spencer, T. J., et al. (1999). Clinical correlates of ADHD in females: Findings from a large group of girls ascertained from pediatric and psychiatric referral sources. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*, 966-975.
- Biederman, J., Faraone, S. V., Milberger, S., Curtis, S., Chen, L., Marris, A., et al. (1996). Predictors of persistence and remission of ADHD into adolescence: Results from a four-year prospective follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*, 343-351.
- Bird, H. R. (1999). The assessment of functional impairment. In D. Shaffer, C. P. Lucas, & J. E. Richters (Eds.), *Diagnostic assessment in child and adolescent psychopathology* (pp. 209-229). New York: Guilford Press.
- Blachman, D. R., & Hinshaw, S. P. (2002). Patterns of friendship in girls with and without attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology*, *30*, 625-640.
- Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology*, *8*, 597-600.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross-age perspective. *Developmental Psychology*, *18*, 557-570.
- Dalsgaard, S., Mortensen, P. B., Frydenberg, M., & Thomsen, P. H. (2002). Conduct problems, gender, and adult psychiatric outcome with attention-deficit hyperactivity disorder. *British Journal of Psychiatry*, *181*, 416-421.
- Dishion, T. (1990). The peer context of troublesome child and adolescent behavior. In P. E. Leone (Ed.), *Understanding troubled and troubling youth* (pp. 128-153). Thousand Oaks, CA: Sage.
- Elliott, D. S., Huizinga, D., & Ageton, S. (1985). *Explaining delinquency and drug use*. Thousand Oaks, CA: Sage.
- Eme, R. F. (1992). Selective female affliction in the developmental disorders of childhood: A review. *Journal of Clinical Child Psychology*, *21*, 354-364.
- Faul, F., & Erdfelder, E. (1992). *G-power: A priori, post-hoc, and compromise power analyses for MS-DOS* [Computer software]. Bonn, Germany: Bonn University Department of Psychology.

- Garner, D. M. (1991). *Eating Disorder Inventory* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Garner, D. M., & Garfinkel, P. E. (1979). The Eating Attitudes Test: An index of the symptoms of anorexia nervosa. *Psychological Medicine*, *9*, 273–279.
- Garner, D. M., Olmstead, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological Medicine*, *12*, 871–878.
- Gaub, M., & Carlson, C. L. (1997). Gender differences in ADHD: A meta-analysis and critical review. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 1036–1046.
- Gershon, J. (2002). A meta-analytic review of gender differences in ADHD. *Journal of Attention Disorders*, *5*, 143–154.
- Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System: Parent, teacher, and child forms*. Circle Pines, MN: American Guidance Systems.
- Hart, E. L., Lahey, B. B., Loeber, R., Applegate, B., & Frick, P. J. (1995). Developmental change in attention deficit hyperactivity disorder in boys: A four-year longitudinal study. *Journal of Abnormal Child Psychology*, *23*, 729–749.
- Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development*, *53*, 87–97.
- Harter, S. (1988). *Manual for the Self-Perception Profile for Adolescents*. Unpublished manuscript, University of Denver.
- Hinshaw, S. P. (2002). Preadolescent girls with attention-deficit/hyperactivity disorder: I. Background characteristics, comorbidity, cognitive and social functioning, and parenting practices. *Journal of Consulting and Clinical Psychology*, *70*, 1086–1098.
- Hinshaw, S. P., & Blachman, D. R. (2005). Attention-deficit/hyperactivity disorder. In D. Bell-Dolan, S. Foster, & E. J. Mash (Eds.), *Handbook of behavioral and emotional problems in girls* (pp. 117–147). New York: Kluwer Academic/Plenum Press.
- Hinshaw, S. P., Carte, E. T., Sami, N., Treuting, J. J., & Zupan, B. A. (2002). Preadolescent girls with attention-deficit/hyperactivity disorder: II. Neuropsychological performance in relation to subtypes and individual classification. *Journal of Consulting and Clinical Psychology*, *70*, 1099–1111.
- Hinshaw, S. P., March, J. S., Abikoff, H., Arnold, L. E., Cantwell, D. P., Conners, C. K., et al. (1997). Comprehensive assessment of childhood attention-deficit hyperactivity disorder in the context of a multisite, multimodal clinical trial. *Journal of Attention Disorders*, *1*, 217–234.
- Jessor, R., Donovan, J. E., & Costa, F. M. (1989). *Health Behavior Questionnaire*. Boulder: University of Colorado Institute of Behavioral Science.
- Kovacs, M. (1992). *Manual: Children's Depression Inventory*. Toronto, Canada: Multihealth Systems.
- Lahey, B. B., Applegate, B., McBurnett, K., Biederman, J., Greenhill, L., Hynd, G. W., et al. (1994). *DSM-IV* field trials for attention deficit hyperactivity disorder in children and adolescents. *American Journal of Psychiatry*, *151*, 1673–1685.
- Lahey, B. B., Pelham, W. E., Loney, J., Kipp, H., Ehrhardt, A., Lee, S. S., et al. (2004). Three-year predictive validity of *DSM-IV* attention-deficit/hyperactivity disorder in children diagnosed at 4–6 years of age. *American Journal of Psychiatry*, *161*, 2014–2020.
- Latimer, W. W., August, G. J., Newcomb, M. D., Realmuto, G. M., Hektner, J. M., & Mahty, R. M. (2003). Child and familial pathways to academic achievement and behavioral adjustment: A prospective six-year study of children with and without ADHD. *Journal of Attention Disorders*, *7*, 101–116.
- Loeber, R., Stouthamer-Loeber, M., van Kammen, V., & Farrington, D. (1994). *Delinquency, substance use, and mental health problems during childhood and adolescence*. Unpublished manuscript, University of Pittsburgh, Pittsburgh, PA.
- Loney, J., Kramer, J., & Milich, R. (1981). The hyperkinetic child grows up: Predictors of symptoms, delinquency, and achievement at follow-up. In K. Gadow & J. Loney (Eds.), *Psychosocial aspects of drug treatment for hyperactivity* (pp. 381–415). Boulder, CO: Westview Press.
- Mannuzza, S., & Gittelman, R. (1984). The adolescent outcome of hyperactive girls. *Psychiatry Research*, *13*, 19–29.
- Mannuzza, S., & Klein, R. G. (1999). Adolescent and adult outcomes of attention-deficit/hyperactivity disorder. In H. C. Quay & A. E. Hogan (Eds.), *Handbook of disruptive behavior disorders* (pp. 279–294). New York: Kluwer Academic.
- Mannuzza, S., & Klein, R. G. (2000). Long-term prognosis in attention-deficit/hyperactivity disorder. *Child and Adolescent Psychiatric Clinics of North America*, *9*, 711–726.
- Mannuzza, S., Klein, R. G., Bessler, A., Malloy, P., & LaPadula, M. (1998). Adult psychiatric status of hyperactive boys grown up. *American Journal of Psychiatry*, *155*, 493–498.
- McBurnett, K., Pfiffner, L. J., & Frick, P. J. (2001). Symptom properties as a function of ADHD type: An argument for continued study of sluggish cognitive tempo. *Journal of Abnormal Child Psychology*, *29*, 207–213.
- Mikami, A. Y., & Hinshaw, S. P. (2003). Buffers of peer rejection among girls with and without ADHD: The role of popularity with adults and goal-directed solitary play. *Journal of Abnormal Child Psychology*, *31*, 381–397.
- Milich, R., Balentine, A. C., & Lynam, D. R. (2001). ADHD combined type and ADHD predominantly inattentive type are distinct and unrelated disorders. *Clinical Psychology: Science and Practice*, *8*, 463–488.
- Molina, B. S. G., & Pelham, W. E. (2003). Childhood predictors of adolescent substance use in a longitudinal study of children with ADHD. *Journal of Abnormal Psychology*, *112*, 497–507.
- MTA Cooperative Group. (1999a). A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, *56*, 1073–1086.
- MTA Cooperative Group. (1999b). Moderators and mediators of treatment response for children with ADHD: The MTA Study. *Archives of General Psychiatry*, *56*, 1088–1096.
- National Institute on Drug Abuse. (1992). National Household Survey of Drug Abuse. Bethesda, MD: U.S. Government Printing Office.
- Owens, E. B., Hinshaw, S. P., Kraemer, H. C., Arnold, L. E., Abikoff, H. B., Cantwell, D. P., et al. (2003). Which treatment for whom for ADHD? Moderators of treatment response in the MTA. *Journal of Consulting and Clinical Psychology*, *71*, 540–552.
- Pajer, K. A. (1998). What happens to “bad” girls? A review of the adult outcomes of antisocial adolescent girls. *American Journal of Psychiatry*, *155*, 862–870.
- Peris, T., & Hinshaw, S. P. (2003). Family dynamics and preadolescent girls with ADHD: The relationship between expressed emotion, ADHD symptomatology, and comorbid disruptive behavior. *Journal of Child Psychology and Psychiatry*, *44*, 1177–1190.
- Rasmussen, P., & Gillberg, C. (2000). Natural outcome of ADHD with developmental coordination disorder at age 22 years: A controlled, longitudinal, community-based study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 1424–1431.
- Robins, L., & Price, R. K. (1991). Adult disorders predicted by childhood conduct problems: Results from the NIMH Epidemiologic Catchment Area project. *Psychiatry*, *54*, 116–132.
- Satterfield, J., Swanson, J. M., Schell, A., & Lee, F. (1994). Prediction of antisocial behavior in attention deficit hyperactivity disorder boys from aggressive/defiance scores. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 185–190.
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH Diagnostic Interview Schedule for Children, Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 28–38.
- SPSS. (2003). *SPSS for Windows* (Version 12.0). Chicago: Author.

- Swanson, J. M. (1992). *School-based assessments and interventions for ADD students*. Irvine, CA: K. C. Press.
- Waschbusch, D. (2002). A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological Bulletin*, *128*, 118–150.
- Wechsler, D. (1991). *Wechsler Intelligence Scales for Children* (3rd ed.). New York: Psychological Corporation.
- Wechsler, D. (1992). *Wechsler Individual Achievement Test*. New York: Psychological Corporation.
- Weiss, G., & Hechtman, L. T. (1993). *Hyperactive children grown up* (2nd ed.). New York: Guilford Press.
- Young, S., Hepinstall, E., Sonuga-Barke, E., Chadwick, O., & Taylor, E. (2005). The adolescent outcome of hyperactive girls: Self-report of psychosocial status. *Journal of Child Psychology and Psychiatry*, *46*, 255–262.
- Zalecki, C., & Hinshaw, S. P. (2004). Overt and relational aggression in girls with attention-deficit hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology*, *33*, 131–143.
- Zoccolillo, M., Pickles, A., Quinton, D., & Rutter, M. (1992). The outcome of childhood conduct disorder: Implications for defining adult personality disorder and conduct disorder. *Psychological Medicine*, *22*, 971–986.

Received April 14, 2005

Revision received September 12, 2005

Accepted November 2, 2005 ■