Journal of Consulting and Clinical Psychology
2012, Vol. 80, No. 6, 1041–1051

© 2012 American Psychological Association
0022-006X/12/$12.00 DOI: 10.1037/a0029451

Prospective Follow-Up of Girls With Attention-Deficit/Hyperactivity Disorder Into Early Adulthood: Continuing Impairment Includes Elevated Risk for Suicide Attempts and Self-Injury

Stephen P. Hinshaw, Elizabeth B. Owens, and Christine Zalecki
University of California, Berkeley

Suzanne Perrigue Huggins
University of Maryland

Adriana J. Montenegro-Nevado
Palo Alto University

Emily Schrodek and Erika N. Swanson
University of California, Berkeley

Objective: We performed a 10-year prospective follow-up of a childhood-ascertained (6–12 years), ethnically and socioeconomically diverse sample of girls with attention-deficit/hyperactivity disorder (ADHD; $N = 140$: combined type [ADHD-C] $n = 93$; inattentive type [ADHD-I] $n = 47$) plus a matched comparison group ($N = 88$). Girls were recruited from schools, mental health centers, pediatric practices, and via advertisements; extensive evaluations confirmed ADHD versus comparison status. ***Method:*** Ten-year outcomes (age range 17–24 years; retention rate = 95%) included symptoms (ADHD, externalizing, internalizing), substance use, eating pathology, self-perceptions, functional impairment (global, academic, service utilization), self-harm (suicide attempts, self-injury), and driving behavior. ***Results:*** Participants with childhood-diagnosed ADHD continued to display higher rates of ADHD and comorbid symptoms, showed more serious impairment (both global and specific), and had higher rates of suicide attempts and self-injury than the comparison sample, with effect sizes from medium to very large; yet the groups did not differ significantly in terms of eating pathology, substance use, or driving behavior. ADHD-C and ADHD-I types rarely differed significantly, except for suicide attempts and self-injury, which were highly concentrated in ADHD-C. Domains of externalizing behavior, global impairment, service utilization, and self-harm (self-injury and suicide attempts) survived stringent control of crucial childhood covariates (age, demographics, comorbidities, IQ). ***Conclusions:*** Girls with childhood ADHD maintain marked impairment by early adulthood, spreading from symptoms to risk for serious self-harm. Our future research addresses the viability of different diagnostic conceptions of adult ADHD and their linkages with core life impairments.

Keywords: attention-deficit/hyperactivity disorder (ADHD), females, longitudinal, suicide, self-injury

When attention-deficit/hyperactivity disorder (ADHD) occurs in girls, it is associated with major impairments, both in childhood and through midadolescence (Biederman et al., 1999, 2006; Hinshaw, 2002; Hinshaw, Owens, Sami, & Fargeon, 2006; Young, Hepinstall, Sonuga-Barke, Chadwick, & Taylor, 2005; see reviews in Hinshaw & Blachman, 2005, and Rucklidge, 2010). Yet little is known about symptoms and impairments in adulthood. Prior research with small female subsamples has revealed continuing problems (e.g., Barkley, Murphy, & Fischer, 2008; Rasmussen & Gillberg, 2000; Weiss & Hechtman, 1993). Babinski et al. (2011) reported on a subsample of the Pittsburgh Adolescent Longitudinal Study (34 young women with ADHD and 24 matched comparisons), followed to a mean age of 20 years. The ADHD group had worse academic and vocational outcomes, more conflicts with family and friends, and lower self-esteem than the comparison sample, but rates of self-reported substance use did not differ significantly. In a far larger investigation, Biederman et al. (2010) described an 11-year follow-up of a sample of girls with ADHD ($n = 140$) and comparison girls ($n = 122$), ascertained between the ages of 6 and 17. By early adulthood (mean age of 22 years at follow-up) those originally diagnosed with ADHD displayed significantly higher lifetime prevalence rates than the comparison women across six psychopathology domains: mood disor-

This article was published Online First August 13, 2012.

Stephen P. Hinshaw, Department of Psychology, University of California, Berkeley; Elizabeth B. Owens and Christine Zalecki, Institute of Human Development, University of California, Berkeley; Suzanne Perrigue Huggins, Department of Psychology, University of Maryland; Adriana J. Montenegro-Nevado, Pacific Graduate School of Psychology-Stanford PsyD Consortium, Palo Alto University; Emily Schrodek, Institute of Human Development, University of California, Berkeley; Erika N. Swanson, Department of Psychology, University of California, Berkeley.

This project was supported by National Institute of Mental Health Grant 45064. We gratefully acknowledge the girls—now young women—who have participated in our ongoing longitudinal investigation, along with their caregivers and our large numbers of graduate students and research assistants. Their remarkable dedication has made this research possible.

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

ders, anxiety disorders, antisocial disorders, developmental disorders, substance dependence, and eating disorders, with differences surviving statistical control of baseline rates of these disorders.

The largest additional female sample comes from the Multimodal Treatment Study of Children with ADHD, featuring 116 girls among its 579 participants. All had been diagnosed in childhood with ADHD-combined type (MTA Cooperative Group, 1999a). Published findings on the girls include an absence of sex-by-treatment interactions for key outcomes (MTA Cooperative Group, 1999b; Owens et al., 2003) or a lack of explicit consideration of sex effects (Molina et al., 2009).

A limitation of much extant prospective research is the limited socioeconomic or ethnic diversity of most participants. Furthermore, retention rates across childhood through adulthood have not always been superior (e.g., Biederman et al., 2010; Weiss & Hechtman, 1993). Given the clear need to elucidate the young-adult outcomes of girls and women with major forms of psychopathology—especially childhood-onset conditions that typically show large male:female sex disparities, such as autism spectrum disorders, early onset conduct problems, and ADHD—our key aim is to understand patterns of symptoms and impairment over time in a large, diverse, well-characterized, and highly retained sample of girls with ADHD, ascertained in childhood.

What would constitute outcomes of major importance? Beyond ADHD symptoms, externalizing behavior patterns, academic underachievement, and substance abuse, all of which have been elevated in follow-up studies of predominantly male samples (e.g., Barkley et al., 2008; Mannuzza & Klein, 2000), additional domains—for example, internalizing symptoms and self-harmful behavior—may be particularly salient for girls. In a prospective investigation of ADHD ascertained during the preschool years, comprising mostly boys (86%) but with a small female subgroup, Chronis-Tuscano et al. (2010) found that probands showed higher rates of major depression, suicidal ideation, and suicide attempts than those in the matched comparison group. Both boys and girls with ADHD showed increased rates of suicidal behavior, but additional analyses revealed that the small female subsample (n = 18) was at significantly higher risk than the male subsample (n = 107). Comorbid internalizing and externalizing symptomatology at baseline, as well as maternal depression, contributed to such risk.

A key unanswered question pertains to the predictive validity of ADHD subtypes—the predominantly inattentive type (ADHD-I) versus the combined type (ADHD-C)—over time. Debate has continued about the distinctiveness of the inattentive type from forms marked by hyperactive-impulsive behavior (e.g., Milich, Balentine, & Lynam, 2001). Because of the potential importance of the inattentive type for girls (American Psychiatric Association, 2000; Hinshaw & Blachman, 2005), our analyses include differentiation of these ADHD types in childhood. Chronis-Tuscano et al. (2010) found that depression was predicted longitudinally by both early inattention and hyperactivity/impulsivity (HI) but suicide attempts were related largely to HI symptoms (i.e., in ADHD-C and ADHD-HI types but not the ADHD-I). Our sampling strategy did not include a baseline group of girls with ADHD-HI), which is rarely found above the preschool years; our contrasts herein pertain to ADHD-I versus ADHD-C. Of note, ADHD subtypes tend to fluctuate markedly over time (e.g., Lahey et al., 2004), making longitudinal predictions to impairment a challenging endeavor. We predict long-term outcomes from dimensionalized symptom counts of HI versus inattentive behaviors elsewhere (Loya, Guelzow, & Hinshaw, 2012).

Source of information is an important methodological and conceptual issue with regard to tracking longitudinal outcomes in ADHD samples. Through late adolescence and early adulthood, at least in largely male samples (Barkley, Fischer, Smallish, & Fletcher, 2002; Barkley et al., 2008), parent/caregiver report is correlated with youth/young adult self-report at low levels. Parent report better predicts long-term negative outcomes, presumably because of the difficulties young adults with ADHD have with reporting accurately on symptoms and impairment. We therefore feature both self-reported and parent-reported outcomes, as well as objective indictors of performance when possible (e.g., reading and mathematics scores).

Our overall objective is to understand young-adult outcomes of girls with ADHD, ascertained during childhood, through a range of clinically interpretable outcomes, with an explicit attempt to have outcome domains parallel or identical to those from our 5-year follow-up (Hinshaw et al., 2006), and to target developmentally salient extensions of key outcomes. Note that, during adolescence, each of the 11 domains of functioning examined yielded significant ADHD versus comparison differences (Hinshaw et al., 2006). For our 10-year, young-adult follow-up, we emphasized the following domains: ADHD-related, externalizing, and internalizing symptoms; eating pathology (Mikami, Hinshaw, Patterson, & Lee, 2008); substance use; academic achievement; self-perceptions of competence; service utilization; self-harm (including the crucial outcomes of suicide attempts and self-injurious behavior); and driving behavior. Because of space limitations, we discuss neuropsychological predictors and outcomes elsewhere (M. Miller, Ho, & Hinshaw, 2012; M. Miller, Nevado-Montenegro, & Hinshaw, 2012). Related to persistence of ADHD symptoms, Faraone, Biederman, and Mick (2006) discussed the thorny issues involved in appraising adult ADHD status, given the natural diminution of symptoms found in clinical and normative samples. Given the extensive work needed to provide data on alternative conceptions of adult diagnostic criteria (e.g., Barkley et al., 2008), we adhere to *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994) criteria herein and present data on alternate thresholds for adult ADHD in forthcoming research.

We base our hypotheses on the above-cited literature and on our own previous findings of (a) major comorbidity and impairment in the present sample during childhood (Hinshaw, 2002; Hinshaw, Carte, Sami, Treuting, & Zupan, 2002) and (b) continuing cross-domain impairment at our follow-up in early to midadolescence (Hinshaw, Carte, Fan, Jassy, & Owens, 2007; Hinshaw et al., 2006; Mikami et al., 2008; Owens, Hinshaw, Lee, & Lahey, 2009). First, we predict that despite a continuing decline in reported symptoms of ADHD over time, particularly in the HI domain (see Hart, Lahey, Loeber, Applegate, & Frick, 1995; Loya et al., 2012), girls with childhood diagnosed ADHD will continue to show elevations in core symptomatology (ADHD-related, internalizing, and externalizing), young-adult relevant symptom areas (e.g., eating pathology; see, e.g., Biederman et al., 2010; Mikami et al., 2008), substance use severity (Hinshaw et al., 2006; but see also Babinski et al., 2011, for opposing findings), several aspects of impairment (service utilization, academic and global), and crucial

outcomes such as self-injurious behavior and suicide attempts (see Chronis-Tuscano et al., 2010).

Second, as emphasized in reports of our adolescent follow-up (Hinshaw et al., 2006), the "gender paradox" (see Eme, 1992) posits that the less-prevalent sex with a given disorder should show greater levels of symptoms, comorbidity, and impairment than the sex with higher prevalence (see also Hinshaw & Blachman, 2005). We predict that girls with ADHD will show a notable pattern of continuing problems over time, even though we lack specific comparisons to male samples.

Third, we predict that some, but not all, domains of continued ADHD-related impairment in young adulthood will withstand statistical control of important covariates, which include childhood-appraised demographic, comorbidity-related, and cognitive (e.g., IQ) variables (Hinshaw et al., 2006, provided parallel analyses at our 5-year follow-up). We predict that academic impairment, service utilization, and substance use will be specifically linked to early ADHD (e.g., Hinshaw, 1992; Molina & Pelham, 2003).

Method

Overview of Procedures

From the San Francisco Bay area, we recruited girls, aged 6–12 years, from schools, mental health centers, pediatric practices, and through direct advertisements to participate in research summer programs in 1997, 1998, and 1999. These programs were designed as enrichment rather than therapeutic endeavors, with emphasis on ecologically valid measures. After extensive diagnostic assessments, 140 girls with ADHD and 88 age- and ethnicity-matched comparison girls were selected (Hinshaw, 2002). Five years later, we invited all participants for prospective follow-up (Wave 2; Hinshaw et al., 2006). Subsequently, we invited all participants and parents for a 10-year follow-up (Wave 3), involving two half-day, clinic-based assessment sessions. When necessary, we performed telephone interviews or home visits; some of the latter involved travel to other states or nations to which participants had moved. To be thorough in documenting impairments and potential resilience (e.g., Owens et al., 2009), we prioritized multidomain, multisource, and multi-informant data collection.

Participants

At Wave 1, preliminary parent and teacher rating-scale criteria were intentionally set with low, sex-specific thresholds, in order to prevent premature exclusion of potentially eligible girls. Ultimately, study participation was contingent on meeting full criteria for ADHD via the parent-administered Diagnostic Interview Schedule for Children (4th ed.; DISC–IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 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 on age and ethnicity, could not meet diagnostic criteria for ADHD via either adult ratings or structured interview criteria. Some (3.4%) met criteria for internalizing disorders (anxiety/depression) or for disruptive behavior disorders (6.8%), but our goal was not to match comparison participants to those with ADHD on comorbid conditions (which would have yielded a nonrepresentative comparison group). Exclusion criteria for both groups were mental retardation, pervasive developmental disorders, psychosis or overt neurological disorder, lack of English spoken in the home, and medical problems prohibiting summer camp participation.

For the Wave 1 summer programs, the girls spanned the ages of 6–12 years. The sample was ethnically diverse (53% White, 27% African American, 11% Latina, 9% Asian American); family income levels ranged from professional parents to receipt of public assistance. The clinic and summer camp procedures yielded multi-informant, multimethod data on both symptoms and a wide range of domains of functional impairment (Hinshaw, 2002). For the current, 10-year follow-up (Wave 3), we made extensive efforts to track all participants, including those whom we were unable to contact at the adolescent follow-up. Aided by use of social media in some cases, we located, consented, and obtained at least some data from 216 of the 228 original participants (95%). The age range was 17–24 ($M = 19.6$; $Mdn = 20$). Although some participants were 16 years of age at the precise 10-year follow-up point, we waited until age 17 to assess them, in order to use a standard assessment battery for all participants, reflecting young-adult status. For specific measures, sample sizes were lower, as highlighted in the Results section.

To evaluate the representativeness of the retained sample, we contrasted Wave 1 measures for the 12 participants lost to the Wave 3 follow-up versus those retained. Of 23 analyses, on measures ranging from demographics, core ADHD symptoms, comorbid symptoms, and functional impairments, five were significant: The nonretained subsample had lower family incomes and full-scale IQ scores and higher Wave 1 teacher-rated ADHD, externalizing, and internalizing symptoms. Although the Wave 3 sample appears generally representative of the total sample, the nonretained subgroup was more impaired cognitively and behaviorally.

Measures

Assessment staff were bachelor-degree-level research assistants or graduate students in clinical psychology, all highly trained and not informed of participants' Wave 1 diagnostic status. Whereas responses to interview questions and issues regarding medication status might suggest ADHD, (a) objective measures were included (e.g., academic testing, computerized structured interviews), and (b) diagnostic status changed for some girls by Wave 3 (see Results). Thus, we do not believe that data were biased by any breaking of blinds.

As noted above, we selected Wave 3 measures to reflect both symptomatic and impairment-related functioning. Note that 58% of the ADHD-C sample and 44% of the ADHD-I sample had received ADHD-related medications for at least some of the time period since the Wave 2 follow-up (vs. 1% of the comparison sample). On rating scales and interviews, caregivers (referred to as parents because most were mothers or fathers) and young adults were asked to respond regarding ADHD-related symptoms for periods during which the participant had not received ADHD medications. On one of the two assessment days (the one on which neuropsychological testing was performed), participation was stimulant free.

Space permits only brief descriptions of core measures (see Table 1). Except for the Diagnostic Interview Schedule for Children, which yielded diagnoses at Waves 1, 2, and 3, all measures

Table 1
Functioning at W3 Across Domains by W1 Diagnostic Status

Dependent variable	Comparison		Inattentive		Combined			ES[b] and post hoc[c]			Covariates
	N	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*p*[a]	0–1	0–2	1–2	*p*[d]
ADHD symptoms, $F(8, 310) = 15.89$, $p = .000$											
P SNAP Inattention	71	0.4 (0.5)	35	1.6 (0.9)	70	1.8 (0.8)	.000	1.27*	1.46*	0.19	.000
P SNAP HI	71	0.1 (0.2)	35	0.6 (0.6)	70	1.0 (0.8)	.000	0.69*	1.33*	0.64*	.000
Y SNAP Inattention	83	0.6 (0.5)	37	1.0 (0.7)	77	1.1 (0.7)	.000	0.60*	0.73*	0.14	.029
Y SNAP HI	83	0.4 (0.4)	37	0.7 (0.5)	77	0.9 (0.6)	.000	0.47*	0.89*	0.43	.014
Externalizing symptoms, $F(8, 324) = 10.25$, $p = .000$											
DISC–IV CD/ODD (%)	86	4.7%	41	48.8%	86	40.7%	.000	19.5*	14.1*	0.7	.009
ACBL Externalizing	69	46.7 (9.0)	36	56.8 (9.0)	69	62.3 (10.9)	.000	0.84*	1.30*	0.46*	.000
ASR Externalizing	85	50.2 (10.2)	40	56.8 (12.6)	83	57.6 (12.5)	.000	0.54*	0.61*	0.07	.042
SRD Total	86	1.8 (2.1)	40	1.7 (1.8)	85	1.8 (2.1)	.980	0.03	0.00	0.04	N/A
Internalizing symptoms, $F(10, 322) = 4.85$, $p = .000$											
DISC–IV Dep/Dys (%)	86	7.0%	41	19.5%	85	21.2%	.024	3.2*	3.6*	1.1	.858
DISC–IV Anxiety (%)	86	10.5%	41	34.1%	87	32.2%	.001	4.4*	4.1*	0.9	.066
ACBL Internalizing	69	44.9 (10.8)	36	55.2 (10.1)	69	59.3 (13.1)	.000	0.77*	1.08*	0.31	.113
ASR Internalizing	85	51.8 (11.7)	40	54.7 (13.2)	83	55.2 (13.0)	.189	0.23	0.27	0.04	N/A
BDI Total	85	8.2 (10.5)	40	10.1 (10.6)	83	11.6 (10.1)	.094	0.19	0.33	0.15	N/A
Substance use											
SUQ Severity	86	−0.05 (0.8)	40	0.09 (1.0)	84	0.01 (0.9)	.703	0.16	0.06	0.10	N/A
Eating disorder symptoms, $F(8, 384) = 0.90$, $p = .519$											
EAT Total	84	50.8 (17.4)	39	49.8 (14.0)	81	54.9 (18.4)	.199	0.06	0.23	0.30	N/A
EDI Bulimia	82	12.8 (5.4)	40	12.7 (6.0)	79	13.3 (5.4)	.828	0.02	0.08	0.10	N/A
EDI Drive for Thinness	82	17.3 (8.1)	40	15.7 (7.4)	79	18.5 (8.3)	.195	0.19	0.16	0.35	N/A
EDI Body Dissatisfaction	82	26.6 (9.9)	40	26.5 (12.0)	79	29.2 (10.3)	.223	0.01	0.25	0.26	N/A
Global impairment											
CIS	71	0.5 (0.5)	36	1.3 (0.7)	71	1.5 (0.9)	.000	0.98*	1.14*	0.16	.033
Academic achievement, $F(6, 404) = 9.38$, $p = .000$											
WIAT Math	84	105.9 (12.5)	41	91.0 (15.8)	87	91.3 (16.3)	.000	0.91*	0.89*	0.02	.193
WIAT Reading	84	108.8 (8.5)	41	97.2 (15.9)	87	97.3 (14.7)	.000	0.83*	0.83*	0.00	.012
Years of education	81	13.1 (1.6)	42	12.7 (1.0)	85	12.4 (1.5)	.004	0.30	0.53*	0.23	.135
Well-being, $F(6, 390) = 3.01$, $p = .007$											
Harter Self-worth	84	3.2 (0.7)	38	3.0 (0.7)	77	3.0 (0.7)	.076	0.27	0.35	0.07	N/A
Harter Social	84	3.3 (0.7)	38	3.1 (0.6)	77	3.1 (0.7)	.085	0.34	0.30	0.04	N/A
Harter Scholastic	84	3.1 (0.7)	38	2.7 (0.7)	77	2.7 (0.7)	.000	0.59*	0.58*	0.01	.581
Service utilization, $F(8, 398) = 12.65$, $p = .000$											
Any school services (%)	80	21.3%	42	61.9%	84	66.7%	.000	6.0*	7.4*	1.2	.000
Any mental health tx (%)	80	51.3%	42	71.4%	87	73.6%	.006	2.4*	2.6*	1.1	.072
Any stimulant (%)	81	1.2%	41	43.9%	86	58.1%	.000	62.6*	111.1*	1.8	.000
Any other med (%)	81	18.5%	41	17.1%	86	32.6%	.053	0.9	2.1*	2.3	.168
Self-harm, $F(4, 384) = 5.51$, $p = .000$											
Suicide attempts (%)	84	6.0%	39	7.7%	85	22.4%	.004	1.3	4.5*	3.5*	.020
Self-injury (%)	79	19.0%	38	28.9%	81	50.6%	.000	1.7	4.4*	2.5*	.028
Driving											
Y DBQ	73	1.4 (1.1)	29	1.6 (1.1)	68	1.3 (1.0)	.593	0.12	0.10	0.22	N/A

Note. W = Wave. For the effect size (ES) and post hoc comparisons, 0 = comparison, 1 = inattentive, 2 = combined. Y = young adult self-report; P = parent's report on young adult. SNAP = Swanson, Nolan, and Pelham; HI = hyperactive/impulsive; DISC–IV = Diagnostic Interview Schedule for Children–IV; CD/ODD = conduct disorder/oppositional defiant disorder; ACBL = Adult Behavior Checklist; ASR = Adult Self-Report; SRD = Self-Report of Delinquency; Dep/Dys = depression/dysthymia; BDI = Beck Depression Inventory–II; SUQ = Substance Use Questionnaire; EAT = Eating Attitudes Test; EDI = Eating Disorder Inventory–2; CIS = Columbia Impairment Scale; WIAT = Wechsler Individual Achievement Test–II; DBQ = Driving Behavior Questionnaire; N/A = not applicable (i.e., no analysis of covariance because of lack of significance of analysis of variance).

[a] Significance: One-way ANOVA for continuous variables; Pearson chi-square statistic for categorical variables. [b] Effect sizes: Cohen's *d* for continuous variables and odds ratio for dichotomous variables. [c] Tukey's test for each pairwise comparison. [d] Covariates: W1 age, W1 maternal education, W1 family income, W1 child full-scale IQ, W1 comorbid CD/ODD; from P DISC), W1 comorbid anxiety or depression (from P DISC), W1 reading disorder, medication status (any stimulants or other psychotropic medication taken between W2 and W3). With medication status and nonschool treatments as outcomes, we did not covary medication status. With W3 externalizing variables as outcomes, we did not covary comorbid CD/ODD. With W3 internalizing variables as outcomes, we did not covary comorbid depression/dysthymia or anxiety. With W3 achievement variables as outcomes, we did not covary W1 reading disorder.

* $p < .05$.

in the next section reflect Wave 3 (outcome) status. As explained in the Covariates section, all covariates were measured at Wave 1, with the exception of the previous year's medication use.

Symptomatology.

Diagnostic Interview Schedule for Children (4th ed.; Young Adult version; DISC–IV–YA; 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 *DSM–IV* (American Psychiatric Association, 1994). It was administered to young adult participants and separately to parents by trained research staff, who recorded responses instantaneously on laptop computers. We utilized algorithms to capture disorders present within the past year, as lifetime diagnoses would be problematic for a longitudinal investigation. Note that the DISC–IV–YA is designed for individuals between 18 and 25 years of age; it features a continuation of the ADHD module typically used for children and adolescents and is therefore the logical developmental extension for our age range. DISC-derived diagnoses contributed outcome measures in the domains of ADHD, externalizing (oppositional defiant disorder, conduct disorder) and internalizing (depression, dysthymia, anxiety disorders).

Swanson, Nolan, and Pelham Rating Scale (4th ed.; SNAP–IV; Swanson, 1992). This parent and young-adult rating scale includes a dimensionalized checklist of the nine *DSM–IV* items for inattention, the nine items for HI, and the eight items for ODD, with each scored on a 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).

Adult Behavior Checklist and Adult Self-Report (ABCL and ASR; Achenbach & Rescorla, 2003). These extensively used scales are parallel versions of the Child Behavior Checklist, used at our previous waves, providing dimensional symptom measures. ABCL (completed by parents) and ASR (completed by participants themselves) scales have good to excellent reliability and validity (Achenbach & Rescorla, 2003). Each constituent item is rated on a 0 to 2 metric; we utilized externalizing and internalizing *T* scores in our analyses.

Self-Reported Delinquency (SRD; Elliott, Huizinga, & Ageton, 1985). This is a widely used, self-report measure of the frequency and characteristics of overt and covert antisocial actions. 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. At Wave 2 it correlated moderately (r = .34 to .41) with key criterion measures (parent- and teacher-reported delinquency).

Beck Depression Inventory–II (BDI–II; Beck, Steer, Ball, & Ranieri, 1996). This is a widely used and extensively validated self-report instrument tapping symptoms of depression in adults, replacing the Children's Depression Inventory, which we used at prior waves. Its psychometric properties are excellent.

Substance Use Questionnaire (SUQ; Molina & Pelham, 2003). The SUQ is a structured questionnaire/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. 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 = .45 to .53) with substance abuse/dependence symptom levels from the DISC–IV.

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

Functional impairment.

Columbia Impairment Scale (CIS; Bird, 1999). Parents rated, on a 0 to 3 metric, the extent of problems their young adults have across 13 items in the home, peer, and school domains. We utilized 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).

Wechsler Individual Achievement Test, Second Version (WIAT–II; Wechsler, 2001). We administered the Word Reading and Math Reasoning subtests. The WIAT–II is a psychometrically sound, widely used test of achievement. Note that some participants were administered the Basic Reading Test from the original WIAT before we changed forms to the WIAT–II; this subtest is equivalent to Word Reading from the WIAT–II. Established test–retest reliabilities for the Reading and Math scores on the original WIAT range from .85 to .92 (Wechsler, 1992).

Self-perceptions/well-being.

Self-Perception Profile for Adolescents (Harter, 1988). In this upward extension of Harter's extensively used scale for children, adolescents/young adults make self-reports of the extent to which they agree or disagree with statements reflecting perceived competence in several domains. We analyzed the Social Acceptance, Scholastic Competence, and Global Self-Worth subscales. As reported by Harter (1982), internal consistencies of these scales ranged from .75 to .84, with test–retest reliabilities ranging from .69 to .80.

Service utilization and educational attainment.

Family Information Profile (FIP). Each family completed a comprehensive, year-by-year grid requesting information on key life events and demographic information between Wave 2 and Wave 3. We quantified each participant's years of education received. Also, regarding service utilization, we counted (1 vs. 0) the use of special education services at school (e.g., restricted placement, classroom aides, occupational or speech therapy, onsite mental health services). For nonschool services we counted (1 vs. 0) individual, group, or family therapy in the community, plus hospitalizations for mental illness. Although changes in family

structure or status might mediate Wave 3 outcomes, we defer their consideration to other papers.

Self-harm.

Barkley Suicide Questionnaire (Barkley, 2006). This is a three-item self-report scale: "Have you ever considered suicide?"; "Have you ever attempted suicide?"; and "Have you ever been hospitalized for an attempt?" A positive endorsement to any question is followed up with a frequency question. We analyzed the dichotomous suicide attempts item. In addition, the family-completed FIP inquired about suicide attempts; there was one case where the FIP reported an attempt but the Barkley scale did not. We added this individual to the count of attempted suicide.

Self-Injury Questionnaire (SIQ). We assessed variety and frequency of nonsuicidal self-injury (NSSI) using a modification of Claes, Vandereycken, and Vertommen's (2001) SIQ. Vanderlinden and Vandereycken (1997) provided data supporting the validity and reliability of that measure within eating-disordered samples. Participants were asked whether, in the past year, they had deliberately injured themselves (e.g., scratched or cut their skin with objects, burned themselves, hit themselves hard, pulled hair out) and how often (1 = *only once*, 6 = *a couple of times a day*). We created a dichotomous self-injury variable indicating whether or not a participant had ever injured herself on purpose.

Driving behavior.

Driving Behavior Questionnaire (DBQ). We employed a self-report version based on Barkley, Murphy, and Kwasnik (1996), which assesses illegal driving behavior and traffic violations (as opposed to driving errors and attentional lapses). Our total score summed the following dichotomous items: ever driven illegally, had a permit suspended or revoked, had a license suspended or revoked, involved in an accident, and received a traffic violation.

Covariates. To ascertain whether young adult symptoms and impairments were related specifically to the girls' original ADHD status rather than to confounding factors, we controlled for key Wave 1 measures. These include, first, participant age, given the 6- to 7-year age span across the sample, and demographic information (family income, maternal education). We also included full-scale IQ from the Wechsler Intelligence Scale for Children (3rd ed.; Wechsler, 1991). We controlled, as well, for additional mental disorders (i.e., comorbidities for the girls with ADHD), from the Wave 1 DISC–IV, coded as 1 versus 0 for the presence versus absence of (a) ODD or CD, (b) depression/dysthymic disorder or anxiety disorder (the latter had to include the presence of one or more conditions beyond specific phobias), and (c) reading disorder, defined as basic reading score < 85. We also included the presence versus absence of psychotropic medication use during the year before the Wave 3 visit (Lahey et al., 2004, followed a parallel procedure).

Data analytic plan. All statistical analyses were performed with SPSS for Windows, Version 19. We categorized our Wave 3 outcomes into the following 11 domains: ADHD symptoms, externalizing symptoms, internalizing symptoms, substance use, eating disorder symptoms, global impairment, academic achievement, well-being, service and medication utilization, self-harm (self-injury, suicide attempts), and driving behavior. See Table 1 for a complete listing of domains and respective measures in each domain.

We first examined ADHD classification (below threshold, ADHD-I, ADHD-C) across Waves 1 and 3; second, we repeated this analysis across Waves 2 and 3 (see Hinshaw et al., 2006, for Wave 1 to Wave 2 classifications). For these categorical cross-classifications, the diagnostic procedures of Hinshaw (2002) to designate ADHD types constituted the Wave 1 diagnoses. At Waves 2 and 3, we performed parallel procedures: Each of the 18 *DSM–IV* ADHD symptoms was considered present if endorsed on the DISC–IV *or* if the parent or young adult rated it as a 2 (*pretty much*) or 3 (*very much*) on the SNAP (at Waves 1 and 2 teacher report was used instead of young adult). Girls with at least six inattentive and six HI symptoms (with at least four in each domain based on young adult or parent DISC–IV; Hinshaw et al., 1997) were designated ADHD-C; girls with at least six inattentive (with at least four based on young adult or parent DISC–IV) but fewer than six HI as ADHD-I; girls with at least six HI (with at least four based on young adult or parent DISC–IV) but fewer than six inattentive as ADHD-HI; and girls with fewer than six inattentive and six HI as falling below symptom thresholds for ADHD. We prioritized the DISC because of the greater specificity of questions with respect to time frame (and associated impairment) than is available from rating scales.

The second set of analyses involved parametric subgroup comparisons. Following a highly significant multivariate analyses of variance (MANOVA) across all 32 specific outcome measures, we performed eight MANOVAs, one per each of the eight domains with multiple measures. (For global impairment, the sole measure was the CIS; for substance use, the SUQ; for driving behavior, the DBQ.) The independent variable was Wave 1 diagnostic subgroup, comprising three levels: comparison, ADHD-I, and ADHD-C. We then examined separate outcomes via analyses of variance (ANOVAs) plus Tukey 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 (two-tailed) a difference of medium effect size (Faul & Erdfelder, 1992). In Table 1 we display effect sizes (Cohen's *d*), with the difference between means as the numerator and the pooled standard deviation as the denominator (Cohen, 1988), even in cases of nonsignificant ANOVAs or per-subgroup contrasts. 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 2×2 chi-square tests for subgroup comparisons. Here, effect sizes were calculated as odds ratios (ORs). Even if a MANOVA for a particular domain did not attain significance, we present univariate tests for heuristic purposes (especially because of the highly significant omnibus MANOVA).

Finally, we performed multivariate analyses of covariance (MANCOVA) for domains with significant MANOVA results, controlling for Wave 1 family income, maternal education, child age, full-scale IQ, ODD/CD diagnoses (except when testing externalizing outcomes), internalizing diagnoses (except when testing internalizing outcomes), reading disorder status (except when testing achievement outcomes), and the designation of having received psychotropic medication during the year preceding follow-up (except when testing medication or service utilization outcomes). We also 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. Specific covariates employed in each ANCOVA and logistic regression are listed in the note for Table 1. The impact of continuing comorbidities at Waves 2 or 3 might serve as an important additional explanatory factor, but performing such complex analyses would place us far beyond current page limits.

Please note that our data analytic procedure is parallel to that from the core articles from Wave 1 (Hinshaw, 2002) and Wave 2 (Hinshaw et al., 2006), emphasizing measure-by-measure examination to enhance clinical interpretability of outcomes and to preserve dimensional measures when possible. However, to ensure the validity of analyses, we examined distributions of the 23 continuous outcome measures. In a majority of instances (17), underlying distributions were nonnormal, via visual inspection and skewness statistics that were 2 or more times the standard error. In such cases, we applied an inverse transformation. For the positively skewed variables, we added 1 to each value (if the original distribution contained values between 0 and 1), and then took the inverse ($1/x$). For the negatively skewed variables, we did the same, except we first multiplied each value by -1. When we reconducted our analyses with the transformed data, in every case the significance level of the new analysis was identical to the nontransformed data. As an additional approach to assure the viability of our analyses, we reconducted analyses for these 17 nonnormally distributed outcome variables via generalized linear models, as suggested by an anonymous reviewer of the article. In these conservative analyses, every single significance level was unchanged, both with and without inclusion of covariates. Thus, we present our analyses below (see also Table 1) via raw scores and MANOVA/ANOVA models. Finally, as a stringent measure, we applied a Bonferroni correction for our univariate ANOVAs (.05 divided by 5, the largest number of dependent measures per any domain = .01).

Results

ADHD Status

Table 2 presents the *DSM–IV* diagnostic status of the participants across time. For contrasts of Wave 1 to Wave 3 diagnostic status, two participants were missing DISC data at Wave 3, reducing the overall sample size from 216 to 214. The clear majority of the comparison group at Wave 1 maintained status as below ADHD symptom criteria at Wave 3 (75/86; 87%); the remainder met symptom criteria (but not necessarily the age-of-onset criterion) for ADHD. Yet only a minority of girls classified as ADHD-I at Wave 1 maintained this classification at Wave 3 (16/41; 39%). The remainder were classified as either below official symptom thresholds (16/41; 39%) or as meeting criteria for ADHD-C (9/41; 22%). Similarly, under half of those initially classified as ADHD-C maintained this status at Wave 3 (33/85; 39%), with the remainder classified as either below official threshold (38/87; 44%), as ADHD-I (15/87; 17%), or as ADHD-HI (2/87; 2%).

Regarding Wave 2–Wave 3 differences, missing DISC data at Wave 2 compounded sample loss (complete $N = 200$). A majority of non-ADHD participants at Wave 2 continued as below symptom criteria at Wave 3 (94/115; 82%); the remainder met symptom (but not necessarily age-of-onset) criteria for ADHD. Of those meeting criteria for ADHD-I at Wave 2, 20/47 (43%) kept this classification by Wave 3, with the rest classified as either below diagnostic threshold (16/47; 34%) or as ADHD-C (11/47; 23%). Finally, of those classified as ADHD-C at Wave 2, a majority maintained this classification (22/38; 58%), with the rest classified as either below threshold (10/38; 26%), as ADHD-I (5/38; 13%), or as ADHD-HI (1/38; 3%).

Table 2
Attention-Deficit/Hyperactivity Disorder Status Compared to Previous Assessment Points

	Wave 1 diagnosis					
	Comparison ($N = 86$)		ADHD inattentive ($N = 41$)		ADHD combined ($N = 87$)	
Wave 3 diagnosis	*n*	%[a]	*n*	%[a]	*n*	%[a]
Below ADHD threshold	75	87	16	39	38	44
ADHD inattentive	8	9	16	39	15	17
ADHD HI	2	2	0	0	2	2
ADHD combined	1	1	9	22	32	37

	Wave 2 diagnosis					
	Below ADHD threshold ($N = 115$)		ADHD inattentive ($N = 47$)		ADHD combined ($N = 38$)	
Wave 3 diagnosis	*n*	%[b]	*n*	%[b]	*n*	%[b]
Below ADHD threshold	94	82	16	34	10	26
ADHD inattentive	13	11	20	43	5	13
ADHD HI	3	3	0	0	1	3
ADHD combined	5	4	11	23	22	58

Note. ADHD = attention-deficit/hyperactivity disorder; HI = hyperactive/impulsive.
[a] Percentages are of diagnostic group at Wave 1. [b] Percentages are of diagnostic group at Wave 2.

Domains of Symptomatology and Impairment

The omnibus MANOVA across all 32 dependent variables was highly significant, $F(64, 132) = 2.27$, $p = .000$, Pillai's trace = 1.05. Of the eight domain-specific MANOVAs, seven were significant (see Table 1); the exception was eating disorder symptoms.

As for the ADHD symptom domain, all four outcomes revealed significant ANOVAs. For parent- and young-adult-rated SNAP Inattention, as well as young-adult-rated SNAP HI, both ADHD types had significantly higher scores than did the comparison girls, with medium (or approaching medium) to extremely large effect sizes, but did not differ from each other. For parent-rated SNAP HI, a stepwise pattern emerged, whereby comparison participants scored lowest, followed by ADHD-I and then ADHD-C; these contrasts were all of medium effect size.

For the externalizing domain, all measures except the SRD yielded significant findings. For ODD/CD diagnoses, both ADHD types had rates of over 40%, far higher than the comparison rate of 5%. For ABCL (reported by parent), both ADHD types had higher scores than the comparison group, with strong effects, and the Wave 1-diagnosed ADHD-C group had higher scores than the Wave 1-diagnosed ADHD-I group, with a nearly medium effect

size. For the ASR, however, both Wave 1-diagnosed ADHD subtypes had higher scores at Wave 3 than did the comparison group (with medium effect sizes) but did not differ significantly from each other (effect size nearly zero).

For internalizing symptomatology, DISC–IV diagnoses at Wave 3 revealed higher rates of anxiety diagnoses in the ADHD groups than in the comparison participants (odds ratio of approximately 4), but the ADHD subtypes did not differ themselves. Similarly, the ABCL Internalizing scale showed effects of both ADHD subtypes above comparison participants, with large or nearly large effect sizes, but no differences between the two subtypes. The ADHD versus comparison contrasts for DISC–IV depression/dysthymia yielded odds ratios of approximately 3, but given the Bonferroni correction, the p value of .024 for the ANOVA is considered nonsignificant. The self-report measures (ASR, BDI–II) showed no significant differences.

Regarding substance use (one outcome measure, thus no MANOVA), no significant differences emerged. With respect to eating disorder symptoms, unlike Wave 2 (Hinshaw et al., 2006), no significant findings were found. Yet regarding global impairment (one outcome measure, thus no MANOVA), both the ADHD-C and ADHD-I subgroups showed significant, large-effect differences from the comparison group but did not differ significantly themselves.

For academic achievement, the ADHD types showed lower scores than the comparison participants (effect sizes large) on WIAT–II Reading and Math but did not differ from each other (effect sizes miniscule). The ADHD-C subgroup had significantly fewer years of education than the comparison group (medium effect); other contrasts were nonsignificant.

In terms of self-perceived well-being, although the MANOVA was significant ($p = .007$) only scholastic self-perceptions differentiated the groups, with both ADHD types showing significantly lower scores than comparisons (effect sizes medium) but not differing themselves.

As for service utilization, both ADHD subgroups were much more likely to have received school-based services since Wave 2 than did the comparison group, with odds ratios approaching 7, but they did not differ themselves. As for non-school-based mental health services, a similar pattern emerged but with odds ratios under 3 for the ADHD versus comparison differences. Here, the base rate of comparison participants' service utilization was over 50%, lowering the odds ratios. Expectedly, girls with childhood (Wave 1) ADHD diagnoses were far more likely (odds ratios of 62.6 and 111.1) than comparison girls to have received ADHD medication between Wave 2 and Wave 3; there was no difference between the ADHD types. Yet no significant differences emerged between girls with and without childhood ADHD regarding their use of nonstimulant psychotropic medications between Wave 2 and Wave 3.

In the key domain of self-harm, for suicide attempts the participants with Wave 1 ADHD-C had a higher rate (22%) than those with ADHD-I (8%) or the comparisons (6%), who did not differ significantly. Odds ratios for the ADHD-C versus (a) ADHD-I and (b) comparison percentages were 3.5 and 4.5, respectively. Self-injury was significantly more likely (OR = 4.4) in the ADHD-C group (51%) than the comparison group (19%); it was also more likely in the ADHD-C group compared to the ADHD-I group (29%; OR = 2.5). Thus, self-harmful behavior predominated in the participants originally diagnosed with ADHD-C.

The sole measure of self-reported driving behavior yielded no significant differences.

Covariates

Our MANCOVAs/ANCOVAs revealed that significant effects of diagnostic status remained for four of the seven domains with significant MANOVAs: ADHD symptoms, $F(8, 282) = 5.73$, $p = .000$; externalizing symptoms, $F(8, 296) = 4.57$, $p = .000$; service utilization, $F(8, 374) = 8.46$, $p = .000$; and self-harm, $F(4, 346) = 2.65$, $p = .033$. Global impairment, a standalone measure, also showed significant diagnostic status differences after strict controls were employed, $F(2, 171) = 3.48$, $p = .033$. For one domain the effect of diagnostic status was reduced to marginal significance: academic, $F(6, 382) = 1.97$, $p = .069$. For internalizing symptoms, $F(10, 294) = 0.81$, $p = .622$, and well-being $F(6, 354) = 0.286$, $p = .943$, the effect of diagnostic status was reduced to nonsignificance. (We did not perform covariance analyses for substance use and eating disorder symptoms because they did not reveal diagnostic group differences in our primary analyses.) As for individual measures surviving ANCOVAs, the most salient were indicators of externalizing behavior, global impairment (Columbia Impairment Scale), school service utilization, suicide attempts, and self-injury.

Discussion

We prospectively followed a large, diverse, and carefully diagnosed sample of preadolescent girls with ADHD 10 years after their ascertainment, which had occurred at 6–12 years of age. The overall retention rate was 95% for the young-adult, Wave 3 assessments. We expand on the findings of Biederman et al. (2010) regarding rates of psychiatric diagnosis in young adult women with ADHD by including core measures of impairment herein. (a) Although the comparison group nearly always maintained their non-ADHD status, a majority of girls with ADHD-I and similar percentages of those with ADHD-C lost (or in some cases gained) enough symptoms that their young-adult diagnostic subgroup status changed. (b) Even so, girls with Wave 1 ADHD in childhood had significantly more symptoms and impairments at the young-adult, Wave 3 follow-up than did the comparison group across most domains, with effect sizes ranging from medium to very large. (c) Significant ADHD-C versus ADHD-I differences at the follow-up were rarely found, with the important exception of self-injury and suicide attempts, which predominated in the childhood-diagnosed combined type. (d) For most domains ADHD-related deficits survived stringent statistical control of age, Wave 1 demographics and comorbidity, IQ, and recent medication status. Overall, our core conclusion is that ADHD in girls portends noteworthy problems 10 years later, at the beginning of adulthood. We note that in alternative analyses, continuous measures of ADHD-related symptomatology also predict key impairments in regression-based and structural equation models (Loya et al., 2012).

We elaborate on key findings as concisely as possible in the remaining space. First, as predicted, ADHD diagnoses (and particularly subtype designations) were not stable over time. When categorical classifications of ADHD are monitored yearly, as in the

careful work of Lahey et al. (2004), measurement error and "regression to the mean" occur; classifications may fluctuate as a function of changes in just one or two symptoms. Our evidence for continued impairment despite the finding that 40% or more of our childhood-diagnosed girls with ADHD no longer met official diagnostic criteria 10 years later places into sharp relief the most accurate means of classifying adult ADHD (American Psychiatric Association, 2000; see also Faraone et al., 2006). Subsequent analyses will examine the validity of alternative conceptions of adult ADHD diagnoses (see Barkley et al., 2008).

Second, our key finding was that childhood-diagnosed girls with ADHD continue to show greater psychiatric symptomatology across multiple symptom areas (ADHD, externalizing, certain aspects of internalizing) and larger functional impairments (global, academic, service utilization rates) than comparison girls. Effect sizes were medium to extremely large, arguing for clinical as well as statistical significance. The sheer range of negative outcomes is noteworthy; the most striking include the high occurrences of suicide attempts and self-injury in the ADHD sample, confined to the childhood-diagnosed combined type. Along with the finding that risk for major depression and dysthymia was also elevated in the participants with ADHD, our results recall those of the small female subsample of Chronis-Tuscano et al. (2010): ADHD in girls and women carries a particularly high risk of internalizing, even self-harmful behavior patterns. Our findings reinforce the contention that female manifestations of ADHD are particularly severe, with public health implications (Hinshaw & Blachman, 2005). On the other hand, self-reported depression symptoms (BDI, ASR) did not significantly differ as a function of diagnostic status, showing the separability of measures of self-harm from psychiatric diagnoses (see relevant discussion in Nock, 2012; Selby, Bender, Gordon, Nock, & Joiner, 2012).

Third, parallel to Wave 1 (Hinshaw, 2002; Hinshaw et al., 2002) and the Wave 2, 5-year follow-up, ADHD-C versus ADHD-I differences were rarely significant and almost always of small effect size. Yet a clear exception is that suicide attempts and self-injury predominated in the combined type compared to the inattentive type. In all, the distinctiveness of ADHD subtypes remains a contentious issue (Milich et al., 2001), and our data clearly suggest that many outcomes for girls with ADHD-C and ADHD-I are highly similar.

Fourth, given the presence of 32 separate outcomes, which we retained in the interest of clinical interpretability (rather than forming composite scores), we examined informant effects. The most consistent findings of continued ADHD-related impairment emanated from parent ratings or objective measures, although self-reports of suicide attempts and self-injury were, as emphasized throughout, highly salient. It will be important for investigators of long-term outcome to avoid relying on any single informant.

Fifth, we demonstrated some evidence of specificity of impairment from Wave 1 ADHD status to young-adult, Wave 3 impairment via our MANCOVAs, which revealed that for ADHD and externalizing symptoms, global impairment, self-harm, and service utilization, diagnostic group differences survived stringent statistical control of Wave 1 age, family income, maternal education, IQ, and comorbid diagnoses, as well as recent medication status. Evidence for complete specificity would entail costly recruitment of control groups matched on Wave 1 comorbidities, which was beyond our intentions and resources. Our findings recall those of Molina et al. (2009) from the MTA Study, in which control of baseline variables revealed the importance of demographic and comorbidity-related factors for long-term adjustment. Furthermore, the stringency of covariates used in the present report, particularly by including IQ—which we performed in parallel with those of Hinshaw (2002) and Hinshaw et al. (2006) at earlier waves of data collection—may have engendered "overcontrol," in which the covariates may have removed some of the variance contributed by ADHD per se (see G. A. Miller & Chapman, 2001). Our covariance findings are therefore extremely conservative.

A key question is why, by young adulthood, young women with ADHD would show markedly high risk for self-harm. The "location" of this finding in the combined type suggests that impulse-control problems may be a central factor. We aim to explore predictors and mediators of the high risk for self-harmful actions in other articles (see Seymour et al., 2012, for data on the mediating role of emotion regulation in ADHD–depression linkages). In all, multifinal, heterotypically continuous outcomes in girls with ADHD over time appear salient.

Yet eating disorder symptoms, which were elevated in our ADHD sample at the adolescent, Wave 2 follow-up, were no longer significantly different from the comparison group by adulthood. Examination of mean scores on our four measures of eating disorder symptoms shows that for all groups on all measures, means rose between adolescence and young adulthood, but the means for the comparison group rose more than they did for the girls with ADHD. Thus, these symptoms among girls with ADHD began earlier but were matched, by young adulthood, among the nondiagnosed girls. Furthermore, our lack of significant differences in the substance use domain echoes Babinski et al. (2011) but differs from Biederman et al. (2010). It may be that boys and men with ADHD show a more consistent pattern of risk for illicit substance use and abuse than females across samples; additional research is sorely needed.

Study limitations include, first, the nonrepresentative nature of the present ADHD sample. As explained in Hinshaw (2002), our goal was not to ascertain a representative, epidemiologically derived sample 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 utilization of multiple recruitment sources and the diversity of the sample suggest that the girls reasonably well reflect the nature of ADHD in the San Francisco Bay area. In addition, the sample did not include the ADHD-HI type at baseline. This category is concentrated in preschoolers (Lahey et al., 1994), and we aimed to preserve statistical power for crucial inattentive versus combined contrasts. Furthermore, our comparison sample was not selected to match the ADHD sample in terms of rates of coexisting disorders but was instead intended to represent girls from similar backgrounds yet without ADHD.

Next, our follow-up efforts yielded a high overall retention rate of participants (95%), but certain measures had more loss of data, and those few participants lost to follow-up were poorer, less intelligent, and more symptomatic than the retained sample. In the interests of space, we did not include other outcomes of interest herein, including peer relations and high-risk sexual behavior (see Flory, Molina, Pelham, Gnagy, & Smith, 2006). Our outcome measure of suicidal behavior did not specify which particular acts participants attempted. Finally, the naturalistic nature of our follow-up meant that continued service utilization and medication

use varied widely. Still, statistical control of recent medication use did not appreciably alter findings.

The overarching conclusion is that ADHD in girls portends continuing problems, through early adulthood, that are of substantial magnitude across multiple domains of symptomatology and functional impairment. Even though male versus female status may not dramatically alter the course of symptoms and correlates over time (see Monuteaux, Mick, Faraone, & Biederman, 2010), our findings argue for the clinical impact of ADHD in female samples, the public health importance of this condition in girls and women, and the need for ongoing examination of underlying mechanisms, especially regarding the high risk of self-harm by young adulthood.

References

Achenbach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA Adult Forms & Profiles.* Burlington: University of Vermont, Research Center for Children, Youth, & Families.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Babinski, D. E., Pelham, W. E., Molina, B. S. G., Gnagy, E. M., Waschbusch, D. A., Yu, J., . . . Karch, K. M. (2011). Late adolescent and young adult outcomes of girls diagnosed with ADHD in childhood: An exploratory investigation. *Journal of Attention Disorders, 15,* 204–214. doi:10.1177/1087054710361586

Barkley, R. A. (2006). *Attention-deficit hyperactivity disorder.* New York, NY: 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. doi:10.1037/0021-843X.111.2.279

Barkley, R. A., Murphy, K. M., & Fischer, M. (2008). *ADHD in adults: What the science says.* New York, NY: Guilford Press.

Barkley, R. A., Murphy, K. M., & Kwasnik, D. (1996). Motor vehicle driving performance and risks in young adults with ADHD. *Pediatrics, 98,* 1089–1095.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories–IA and –II in psychiatric outpatients. *Journal of Personality Assessment, 67,* 588–597. doi:10.1207/s15327752jpa6703_13

Biederman, J., Faraone, S. V., Mick, E., Williamson, S., Wilens, T., Spencer, T. J., . . . Zallen, B. (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 & Adolescent Psychiatry, 38,* 966–975. doi:10.1097/00004583-199908000-00012

Biederman, J., Monuteaux, M. C., Mick, E., Spencer, T., Wilens, T., Klein, K. L., . . . Faraone, S. V. (2006). Psychopathology in females with attention-deficit/hyperactivity disorder: A controlled, five-year prospective study. *Biological Psychiatry, 60,* 1098–1105. doi:10.1016/j.biopsych.2006.02.031

Biederman, J., Petty, C. R., Monuteaux, M. C., Fried, R., Byrne, D., Mirto, T., . . . Faraone, S. V. (2010). Adult psychiatric outcomes of girls with attention deficit hyperactivity disorder: 11-year follow-up in a longitudinal case-control study. *American Journal of Psychiatry, 167,* 409–417. doi:10.1176/appi.ajp.2009.09050736

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, NY: Guilford Press.

Chronis-Tuscano, A., Molina, B. S. G., Pelham, W. E., Applegate, B., Dahlke, A., Overmeyer, M., & Lahey, B. B. (2010). Very early predictors of adolescent depression and suicide attempts in children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry, 67,* 1044–1051. doi:10.1001/archgenpsychiatry.2010.127

Claes, L., Vandereycken, W., & Vertommen, H. (2001). Self-injurious behavior in eating-disordered patients. *Eating Behaviors, 2,* 263–272. doi:10.1016/S1471-0153(01)00033-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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. doi:10.1207/s15374424jccp2104_5

Faraone, S. V., Biederman, J., & Mick, E. (2006). The age-dependent decline of attention-deficit hyperactivity disorder: A meta-analysis of follow-up studies. *Psychological Medicine, 36,* 159–165. doi:10.1017/S003329170500471X

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.

Flory, K., Molina, B. S. G., Pelham, W. E., Gnagy, E., & Smith, B. (2006). Childhood ADHD predicts risky sexual behavior in young adulthood. *Journal of Clinical Child and Adolescent Psychology, 35,* 571–577. doi:10.1207/s15374424jccp3504_8

Garner, D. M. (1991). *Eating Disorder Inventory (EDI–2)* (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. doi:10.1017/S0033291700030762

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. doi:10.1017/S0033291700049163

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. doi:10.1007/BF01447474

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53,* 87–97. doi:10.2307/1129640

Harter, S. (1988). *Manual for the Self-Perception Profile for Adolescents.* Unpublished manuscript, University of Denver.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111,* 127–155. doi:10.1037/0033-2909.111.1.127

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. doi:10.1037/0022-006X.70.5.1086

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, NY: Kluwer Academic/Plenum Press. doi:10.1007/0-306-48674-1_4

Hinshaw, S. P., Carte, E. T., Fan, C., Jassy, J. S., & Owens, E. B. (2007). Neuropsychological functioning of girls with attention-deficit/hyperactivity disorder followed prospectively into adolescence: Evidence for continuing deficits? *Neuropsychology, 21,* 263–273. doi:10.1037/0894-4105.21.2.263

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. doi:10.1037/0022-006X.70.5.1099

Hinshaw, S. P., March, J. S., Abikoff, H., Arnold, L. E., Cantwell, D. P., Conners, C. K., . . . Wigal, T. (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. doi:10.1177/108705479700100403

Hinshaw, S. P., Owens, E. B., Sami, N., & Fargeon, S. (2006). Prospective follow-up of girls with attention-deficit/hyperactivity disorder into adolescence: Evidence for continuing cross-domain impairment. *Journal of Consulting and Clinical Psychology, 74,* 489–499. doi:10.1037/0022-006X.74.3.489

Jessor, R., Donovan, J. E., & Costa, F. M. (1989). *Health Behavior Questionnaire.* Boulder, CO: University of Colorado, Institute of Behavioral Science.

Lahey, B. B., Applegate, B., McBurnett, K., Biederman, J., Greenhill, L., Hynd, G. W., . . . Shaffer, D. (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., . . . Massetti, G. (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. doi:10.1176/appi.ajp.161.11.2014

Loya, F., Guelzow, B. T., & Hinshaw, S. P. (2012). *The developmental trajectory of ADHD in females: Predictive associations between symptom change and young adult outcomes*. Manuscript submitted for publication.

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.

Mikami, A. Y., Hinshaw, S. P., Patterson, K. A., & Lee, J. C. (2008). Eating pathology among adolescent girls with attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology, 117,* 225–235. doi:10.1037/0021-843X.117.1.225

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. doi:10.1093/clipsy.8.4.463

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110,* 40–48. doi:10.1037/0021-843X.110.1.40

Miller, M., Ho, J., & Hinshaw, S. P. (2012). Executive functions in girls with ADHD followed prospectively into young adulthood. *Neuropsychology, 26,* 278–287. doi:10.1037/a0027792

Miller, M., Nevado-Montenegro, A. J., & Hinshaw, S. P. (2012). Childhood executive function continues to predict outcomes in young adult females with and without childhood diagnosed ADHD. *Journal of Abnormal Child Psychology, 40,* 657–668. doi:10.1007/s10802-011-9599-y

Molina, B. S. G., Hinshaw, S. P., Swanson, J. M., Arnold, L. E., Vitiello, B., Jensen, P. S., . . . the MTA Cooperative Group. (2009). The MTA at 8 years: Prospective follow-up of children treated for combined type ADHD in a multisite study. *Journal of the American Academy of Child & Adolescent Psychiatry, 48,* 484–500. doi:10.1097/CHI.0b013e31819c23d0

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. doi:10.1037/0021-843X.112.3.497

Monuteaux, M. C., Mick, E., Faraone, S. V., & Biederman, J. (2010). The influence of sex on the course and psychiatric correlates of ADHD from childhood to adolescence: A longitudinal study. *Journal of Child Psychology and Psychiatry, 51,* 233–241. doi:10.1111/j.1469-7610.2009.02152.x

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.

Nock, M. K. (2012). Future directions for the study of suicide and self-injury. *Journal of Clinical Child and Adolescent Psychology, 41,* 255–259. doi:10.1080/15374416.2012.652001

Owens, E. B., Hinshaw, S. P., Kraemer, H. C., Arnold, L. E., Abikoff, H. B., Cantwell, D. P., . . . Wigal, T. (2003). Which treatment for whom for ADHD? Moderators of treatment response in the MTA. *Journal of Consulting and Clinical Psychology, 71,* 540–552. doi:10.1037/0022-006X.71.3.540

Owens, E. B., Hinshaw, S. P., Lee, S. S., & Lahey, B. B. (2009). Few girls with childhood attention-deficit/hyperactivity disorder show positive adjustment during adolescence. *Journal of Clinical Child and Adolescent Psychology, 38,* 132–143. doi:10.1080/15374410802575313

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 & Adolescent Psychiatry, 39,* 1424–1431. doi:10.1097/00004583-200011000-00017

Rucklidge, J. J. (2010). Gender differences in attention-deficit/hyperactivity disorder. *Psychiatric Clinics of North America, 33,* 357–373. doi:10.1016/j.psc.2010.01.006

Selby, E. A., Bender, T. W., Gordon, K. H., Nock, M. K., & Joiner, T. E., Jr. (2012). Non-suicidal self-injury (NSSI) disorder: A preliminary study. *Personality Disorders: Theory, Research, and Treatment, 3,* 167–175. doi:10.1037/a0024405

Seymour, K. E., Chronis-Tuscano, A., Halldorsdottir, T., Stupica, B., Owens, K., & Sacks, T. (2012). Emotion regulation mediates the relationship between ADHD and depressive symptoms in youth. *Journal of Abnormal Child Psychology, 40,* 595–606. doi:10.1007/s10802-011-9593-4

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 & Adolescent Psychiatry, 39,* 28–38. doi:10.1097/00004583-200001000-00014

Swanson, J. M. (1992). *School-based assessments and interventions for ADD students.* Irvine, CA: K. C. Publishing.

Vanderlinden, J., & Vandereycken, W. (1997). *Trauma, dissociation, and impulse decontrol in eating disorders.* Philadelphia, PA: Brunner/Mazel.

Wechsler, D. (1991). *Wechsler Intelligence Scales for Children* (3rd ed.). New York, NY: Psychological Corporation.

Wechsler, D. (1992). *Wechsler Individual Achievement Test.* New York, NY: Psychological Corporation.

Wechsler, D. (2001). *Wechsler Individual Achievement Test, Second Version.* New York, NY: Psychological Corporation.

Weiss, G., & Hechtman, L. T. (1993). *Hyperactive children grown up* (2nd ed.). New York, NY: 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. doi:10.1111/j.1469-7610.2004.00350.x

Received November 1, 2011
Revision received May 21, 2012
Accepted June 11, 2012 ■