ABSTRACT

Objective: This longitudinal study investigated changes in neurocognitive functioning from childhood to early adolescence in a sample of children diagnosed with DSM-IV attention deficit hyperactivity disorder (ADHD). It also compared the neurocognitive trajectories of children who continued to meet the diagnostic criteria for ADHD at follow-up and and those in partial remission.

Methods: Children diagnosed with ADHD (N = 55) were tested at baseline [*M* = 7.7 years, *SD* = 1.5] and four years later [*M* = 11.7 years, *SD* = 1.5] on measures of intellectual, academic, and executive functioning. Group and individual analyses were used to examine neurocognitive functioning over this period.

Results: Intellectual function was stable over the four-year interval. Reliable change analyses highlighted variability in academic performance. Approximately half the sample showed a reliable decline in at least one academic subject with almost a third showing reliable improvement. Executive functions generally followed a stable or improving course, with significant improvements on measures of information processing, attentional control, cognitive flexibility, and goal setting. There was some evidence of better neurocognitive performance in those with partial symptom remission at follow up.

Conclusion: Study findings emphasize the importance of monitoring academic performance in children with ADHD, including examination of change at the individual level. Declines in academic performance were observed despite stable intellectual and improving executive function. Early cognitive functioning did not predict symptom remission, however reduced symptoms at follow-up were associated with better executive function.

Key terms: ADHD, longitudinal study, neurocognition

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a common neurodevelopmental disorder characterized by elevated, and developmentally inappropriate, levels of inattention, overactivity and impulsivity that impair functioning. It is typically diagnosed in early to middle childhood with an estimated prevalence of 5 to 7% [1] making it one of the most common psychiatric disorders of this developmental period. While some children with ADHD show symptom remission over time, it is estimated 50-65% continue to meet diagnostic criteria for the disorder in adulthood.[2,3]

In addition to the disorders core symptoms, cross-sectional studies often report lower scores on tests of intellectual function and academic underachievement in children with ADHD. Meta-analytic findings indicate children with ADHD display significantly lower full-scale intelligence quotient (FSIQ) scores relative to control groups.[4] Adults with ADHD also display significantly lower scores on IQ tests compared with control groups, suggesting these mild intellectual inefficiencies persist across development.[5] It is estimated as many as 80% of children diagnosed with ADHD have academic difficulties, with approximately one-third presenting with specific learning disabilities in reading, mathematics, or writing.[6,7] Children with ADHD are at increased risk of grade retention, placement into special education classrooms, and high-school dropout (see [8] for a review of academic performance and ADHD).

Cross-sectional studies also show that children with ADHD demonstrate specific cognitive inefficiencies within the executive domain.[9-12] While various definitions of executive function exist, the term is typically used to encompass a collection of "capacities that enable a person to engage successfully in independent, purposeful, self-serving behaviors".[13(p42)]

Anderson [14] proposed a model for pediatric populations that conceptualizes executive function as four distinct subdomains: attentional control (i.e., selective attention, self-regulation and monitoring, inhibition), cognitive flexibility (i.e., divided attention, working memory, conceptual transfer, and feedback utilization), goal setting (i.e., initiative, conceptual reasoning, planning, strategic organization), and information processing (i.e., efficiency, fluency, speed of processing). While numerous studies demonstrate deficits in intellectual, academic, and executive functioning in children with ADHD, research examining the neurocognitive developmental trajectories of these children is still limited. Further research is needed to more fully understand the nature and continuity of these deficits and their relationship to longer term functioning in children with ADHD.

A question of interest in field is whether or not neurocognitive functioning is predictive of symptom outcome over time. Many children with ADHD remain symptomatic into adulthood and researchers have hypothesized that neurocognitive functioning may be a predictor of symptom persistence. Halperin and Schulz [15] proposed that ADHD is primarily a non-cortical disorder arising from basal ganglia and cerebellum dysfunction that persists across the lifespan. They suggest development of the prefrontal cortex, and its associated circuitry, in adolescence may compensate for these non cortical deficits leading to symptom reduction over the course of development. Their model predicts that children who show the greatest neurocognitive improvement, particularly on tasks reliant on the prefrontal cortex (e.g., high level mental effort/executive function tasks), will be those who show a reduction in their ADHD symptomatology over time. Verifying this hypothesis is challenging given that both cortical (including the prefrontal cortex) and subcortical (including the basal ganglia and the cerebellum) regions appear to contribute to executive function test performance. [16-18] Longitudinal neuroimaging approaches will likely be necessary to

investigate the proposed underlying neuroanatomical mechanisms in Halperin and Schulz's model. Irrespective of the underlying anatomical substrates, determining whether neurocognitive functioning can reliably predict ADHD symptom outcome over time has important implications for clinical management.

Cross-sectional studies comparing the executive performance of children continuing to meet criteria for ADHD (i.e., persisters) and those who no longer meet criteria for ADHD (i.e., partial remitters) suggest the former have greater difficulty in a number of executive domains, including attentional control,[19-21] cognitive flexibility,[21-22] goal-setting and information processing. [22] However, in their recent systematic review of the literature, van Lishout and colleagues concluded that neither higher nor lower neurocognitive functions reliably differentiate persistent from remitting ADHD, with both groups showing poorer performance than controls.[23]

When assessing the predictive value of neurocognitive function, few studies have included the same neurocognitive measures at different assessment points, and few comprehensively assess the neurocognitive function of children with ADHD. Amongst those that have, support for Halperin and Schulz's theory is limited. The majority of studies do not support a link between neurocognitive performances over time and symptom change. [24-26] Exceptions include better performance on measures of goal setting i.e., Rey Complex Figure Test [26] and cognitive flexibility i.e., Spatial Working Memory from the CANTAB [24] which are associated symptom reduction across development. Additional longitudinal studies are needed to further delineate the relationship between neurocognitive and symptom trajectories in children with ADHD.

Aims and hypotheses

This longitudinal study examined the intellectual, academic and executive function of children diagnosed with DSM-IV ADHD over a four-year interval. Based on previous research the children's baseline intellectual and [4] academic performance [8] was expected to fall below population norms. Impairment on measures of executive function was also expected. [10-13] The results of cross sectional and longitudinal studies suggested the children's intellectual functioning would be stable over time, [21,27] while their academic performance would follow a stable or declining course. [8,28-29] Executive function test performance was also predicted to improve over time,[30-32] although the literature offers limited guidance regarding which domains are most likely to improve. Given the heterogeneity of ADHD, neurocognitive trajectories were examined at the individual as well as the group level where the data permitted. The longitudinal nature of the data also enabled us to compare the neurocognitive trajectories of children who continued to meet full criteria for ADHD at follow up and those whose symptoms were in partial remission. Results would support Halperin and Schulz's theory if the subgroup showing attenuation of symptoms demonstrated greater improvement in executive functioning. Meta-analytic findings [23] suggest baseline neurocognitive functioning does not predict which children will show symptom remission.

METHODS

Participants

Fifty-five children with a DSM-IV diagnosis of ADHD were followed up over a four-year period[α]. The children were initially recruited through an ADHD Research Clinic at the

[α] Follow-up assessments were scheduled between 44 and 52 months after baseline assessments, i.e., 4 years ± 4 months.

University of Otago between 1997 and 2001. They came primarily from local outpatient health services responsible for assessing and treating disordered behavior in children. Children were included in the longitudinal study if they received a diagnosis of ADHD at baseline, had a FSIQ of at least 70 on initial assessment, and showed no evidence of neurological disorder or psychosis. Demographic and diagnostic characteristics are presented in Table 1.

The participating children were part of a larger sample of 103 children, diagnosed with ADHD, whose parents agreed to contact about future studies. Non-participation at follow-up occurred for a variety of reasons (i.e., no response to contact attempts/no current contact details n = 20/4, declined participation n = 18, neurological complications n = 3, moved outside study locale n = 2, non compliant with assessment n = 1). There were no significant differences in baseline age, IQ, socio-economic status, gender, ADHD subtype, or ethnicity between those children who did and did not take part in the follow-up study (Supplementary Table 1).

Procedure

Children were initially diagnosed with ADHD following comprehensive multi-method multi-informant assessments that included parent, teacher, and child semi-structured interviews and parent- and teacher-completed behavioral questionnaires. Parents completed semi-structured interviews in which they described their children's current difficulties and developmental history and were also interviewed with the Anxiety Disorders Interview for Children [33] (parent version) incorporating sections on ADHD, ODD and CD. Children completed a clinical interview, designed for the study, that assessed perceptions of their

behavior, academic performance and peer relationships, they were also rated according to the Rutter and Graham Interview Schedule.[34] Teachers responded to a series of questions about the children's behavioral, academic and social functioning at school. Parent and child interviews were conducted face-to-face and teacher interviews via telephone by graduate students in clinical psychology. Parents and teachers completed the Disruptive Behavior Disorders Rating Scale (DBD), [35] requiring them to rate the severity of symptoms of DSM-IV ADHD, ODD and CD, with symptoms rated as occurring *often* or *very often* considered present. Data from parent and teacher interviews and DBD rating scales were used in diagnostic decision-making[β]. Observations of child behavior during the child interview and neuropsychological testing were used to supplement parent and teacher reports. This information was particularly useful when parent and teacher reports were discrepant. A diagnosis of ADHD was made if: a child exhibited six or more symptoms of inattention and/or hyperactivity/impulsivity in at least one setting; there was evidence of symptoms in a second setting; and symptoms caused clinically significant impairment in at least two settings; were not accounted for by another mental disorder or medical factors; and were inconsistent with the child's developmental level. Symptoms were not added across informants. Preliminary diagnoses were made by the doctoral-level clinical student responsible for the assessment. All diagnoses were reviewed by a senior clinical psychologist (GT) and any disagreements resolved through discussion and consensus.

At follow-up, parents were interviewed face-to-face and teachers via telephone about the children's current social, academic and behavioral functioning. Children were interviewed

[β] If parent interview data indicated six or more symptoms of inattention and/or hyperactivity-impulsivity, but their DBD data did not, the interview data, during which the researchers had the opportunity to clarify parent responses was accepted for diagnostic decision making. If parent DBD data indicated sufficient symptoms but their interview data did not, all available data was reviewed before accepting the DBD symptom count for diagnostic decision making.

about their friends, hobbies and school performance. Parents and teachers were also asked to complete the DBD rating scale. Data from interviews, the DBD, and observations during testing were used to determine if the children continued to meet DSM-IV criteria for ADHD. Those continuing to meet full criteria were characterized as ADHD persisters (n = 34), those no longer meeting full criteria as having ADHD in partial remission (i.e., less than 6 symptoms of inattention or hyperactivity/impulsivity in any setting n = 10, or displaying 6 or more symptoms of inattention or hyperactivity/impulsivity in one setting but no evidence of symptoms in a second setting n = 11). For the 25 of the 34 persisters and 11 of the 21 partial remitters prescribed methylphenidate at time 2, parents and teachers were not specifically directed to base their DBD ratings on behaviors when off medication. Summary data from parent interviews (time1) and parent and teacher completed questionnaires (time 1 and time 2) is presented in Supplementary Table 2.

Children completed the same neurocognitive assessments at baseline and follow-up. The assessments were conducted by trained clinical psychology graduate students and took approximately three hours to complete. The majority of children were assessed in an ADHD Research Clinic at the University of Otago. Due to distance from the University, one baseline and two follow-up assessments were conducted in other locations. Children prescribed methylphenidate were medication free for at least 24 hours for both neurocognitive assessments.

Ethical approval for the study was obtained from the Otago Ethics Committee. Written consent to participate was obtained from the parents, teachers, and children before each assessment.

Measures

Participants' scores were converted to age-standardized scores to control for age-related improvement over the follow-up period. This was done using standardized scores provided by the measure developers for the WISC-III, WRAT-III and WCST. For Verbal Fluency the standardized scores for semantic Word Generation from the NEPSY-II manual (A Developmental Neuropsychological Assessment 2nd edition, 2007) [36] were used. For the TMT age standardized z-scores were derived from data provided by the developers and for the ACPT and Design Fluency task, data from a typically developing control sample recruited during the course of the current study (n = 157)[χ]. The control sample children completed the measures once only.

Neurocognitive measures administered at baseline and follow-up

The Wechsler Intelligence Scale for Children - Third Edition (WISC-III) [33] is a standardized measure of general intellectual functioning. Performance IQ (PIQ), Verbal IQ (VIQ), and Full Scale IQ (FSIQ) scores were pro-rated based on eight subtests (Similarities, Arithmetic, Vocabulary, Digit Span, Picture Completion, Coding, Picture Arrangement, and Block Design) using the procedures outlined by Wechsler. [37] The Freedom From Distractability (FFD, Arithmetic and Digit Span) and Processing Speed (PS, Coding and Symbol Search) Index scores were also calculated. The IQ and Index scores were analyzed with higher scores indicating better performance.

The Wide Range Achievement Test - Third Edition (WRAT-III) [38] is a standardized screening measure of academic achievement in the domains of word reading, spelling, and arithmetic.

[χ] These data were organized in single year age brackets for all ages except 5 & 6 years and 14 & 15 years. Sample sizes for each age band ranged from n = 10 (7 years) to n = 31 (10 years). Means and standard deviations were calculated for each age bracket and used to calculate age corrected z-scores for the ADHD participants. See Supplementary Table 3.

The tan form was used at baseline and the equivalent blue form at follow-up. The three domain scores were analyzed with higher scores representing better performance.

The *Auditory Continuous Performance Test (ACPT)* [39] assessing attentional control, requires participants to remain vigilant to randomly ordered infrequent words presented aurally over a sustained period of time (10 minutes), indicating each time the target word is heard. The number of errors of omission, commission and the vigilance decrement, i.e., difference in omission errors between the first and last trial, were recorded. Total errors and vigilance decrement scores were analyzed. For both lower scores represent better better performance.

The *Trail Making Test - Children's Version (TMT)* [40] assessing information processing and cognitive flexibility requires participants to draw a line connecting the numbers 1-15 in order (TMT-A), and a line alternating between numbers (1-8) and letters (a to h) in ascending order (TMT-B). Time taken to complete each part, including time to correct mistakes, was recorded and analyzed. On this task negative z scores indicate faster task completion and positive scores slower task completion.

The *Wisconsin Card Sorting Test (WCST)* [41] is a standardized measure assessing cognitive flexibility and goal-setting, requiring participants to match response cards to four key cards. Participants are informed whether each attempt is correct or incorrect only, with the criterion on which cards are matched changing when a pre-arranged level of success is met. Total errors, non-perseverative errors, perseverative responses, and perseverative errors were converted to T scores. Perseverative error T scores, considered useful for documenting problems in concept formation, learning from feedback, and conceptual flexibility [42] were analyzed. Higher T scores represent better performance, i.e., fewer errors.

The *Verbal Fluency Task* [43] assessing information processing requires participants to name as many things meeting a criterion as they can in one minute, in this study "things you can eat" and "animals". The total number of correct responses, summed across the two conditions, were analyzed following conversion to standard scores. Higher scores indicate better performance.

The *Design Fluency Task* [44] is a visual analogue of the Verbal Fluency Task. Participants were asked to generate (draw) as many novel, unnamable designs as possible within 3-minutes. In the free condition there are no restrictions, in the fixed condition the designs must have exactly four lines. The number of novel responses, wrong answers (unacceptable designs), perseverative responses and total errors were recorded. Novel responses and total error scores were analyzed, with higher z scores indicating more novel responses and lower z scores fewer errors.

Statistical analysis

Analyses were conducted on the entire sample and subgroups i.e., ADHD persisters and those in partial remission. Mixed-design ANOVA were used to assess time (baseline to follow-up) and group (persist versus partial remission) effects as appropriate. Paired sample t-tests were used in some instances to compare the performance of the whole ADHD group at baseline and follow-up. One-sample t-tests were used to compare baseline and follow-up performance with population norms. Where data violated the assumptions for the use of parametric statistics, nonparametric procedures were used, i.e., the Wilcoxon-Signed-Rank test for within group comparisons and the Mann-Whitney U test for between group comparisons. For whole sample comparisons the Hochberg method, a step-up modification

of the Bonferroni method, was used to control for family-wise error rates within neurocognitive tests.[45] Given the smaller group sizes for the persist and partial remission comparisons findings significant at $p \leq 0.05$ are reported.

Reliable change analyses were carried out to examine the extent to which individual children's scores changed from baseline to follow-up. The procedures outlined by Chelune and colleagues [46] were followed. Their reliable change index (RCI) determines if the difference between two scores exceeds variation expected from instrument error and change due to practice effects (RCI+PE). Participants whose age standardized scores increased from baseline to follow-up in excess of the RCI+PE were judged to have reliably improved, those whose scores decreased below the RCI+PE cut-off to have reliably declined. Adequate reliability coefficients were available for the WISC-III and WRAT-III only, limiting reliable change analyses to these measures. The chi square goodness of fit test was used to determine if the observed proportions of reliable change for the ADHD group exceeded expected rates based on normative data (i.e., 5% of control subjects improve, 5 % deteriorate, and 90 % remain unchanged). The proportions of children showing reliable change for the ADHD persist and partial remission subgroups were compared directly using chi square.

RESULTS

Summary statistics for intellectual, academic and executive performance measures at baseline and follow-up are presented for the entire sample in Tables 2 and 3, and for the ADHD persist and partial remission subgroups in Supplementary Tables 4 and 5. Reliable change analyses for the intellectual and academic measures are presented in Table 4. Whole sample comparisons are presented first followed by the subgroup comparisons. Only significant findings are described in the text.

Whole sample comparisons

Comparisons of the children's baseline and follow-up neurocognitive performance data against population norms are presented in summary form in Tables 2 and 3.

The children's mean WISC-III IQ, Index, and subtest scores showed little change from baseline to follow-up (see Table 2). Reliable change analyses showed WISC-III summary scores did not change over the 4-year follow-up period for the majority of children (see Table 4). Rates of reliable improvement were higher than expected for FSIQ (21.3%, χ^2 = 26.21, *df* = 1, *p* <.001), PIQ (12.8%, χ^2 = 5.97, *df* = 1, *p* = 0.015) and the PS Index (13.8%, χ^2 = 5.97, *df* = 1, *p* = 0.015).

Mean reading and arithmetic scores on the WRAT-III did not change significantly from baseline to follow-up, while spelling scores declined significantly [*F*(1,53) = 5.071, *p* = 0.028][δ] (see Table 2). However, reliable change analyses indicated that more than 20 percent of children showed a reliable decline in age-standardized reading (23.6%), spelling (27.3%), or arithmetic (21.8%) performance over this period (see table 4). These rates of decline are significantly higher than expected: reading (χ^2 = 40.22, *df* = 1, *p* < 0.001), spelling (χ^2 = 57.44, *df* = 1, *p* < 0.001), and arithmetic (χ^2 = 32.75, *df* = 1, *p* < 0.001). Altogether 28 children (51%) showed a reliable decline in one subject area with 11 (20%) demonstrating reliable decline in two or more areas[ε]. The number of children showing reliable

[δ] When this analysis was run as a simple repeated measures ANOVA the result was significant after controlling for multiple *F*(1,54) = 6.844, *p* = 0.012.

[ε] Raw scores for all children increased over the four-year interval, indicating no loss of academic skills. Declines in standardized scores suggest a slower learning rate compared with controls and a widening of the age corrected performance gap.

improvement in reading (14.5%, $\chi^2 = 10.55$, $df = 1$, $p = 0.001$) and arithmetic (10.9%, $\chi^2 = 4.04$, df = 1, p = 0.044) was also higher than expected. Sixteen children (29%) showed reliable improvement in one subject area with two (3.6%) showing reliable improvement in two areas.

For the most part, age-standardized performance on the measures of executive function was stable or improved over time (see Table 3). On the ACPT children made fewer errors [$Z = -2.593$, N = 47, $p < 0.01$] and showed a smaller vigilance decrement [$Z = -2.392$, N = 47 $p < 0.017$] at follow-up. On the WCST they made fewer perseverative errors [$F(1,33) = 24.125$, $p < 0.001$] and generated more words on the Verbal Fluency test [$F(1,52) = 6.702$, $p = 0.012$] at the second assessment. Conversely, under the fixed condition of the Design Fluency task the children made more errors [$t(41) = -2.983$, $p = 0.005$] and produced fewer novel designs [$t(41) = 2.258$, $p = 0.029$] at follow-up. Time to complete part A of the TMT increased to a level similar to that reported for controls [Z = -2.526, N=39, $p = 0.012$] at follow-up.

Subgroup comparisons

Demographic and diagnostic information for the ADHD subgroups is presented in Table 1. The ADHD persistent and partial remission groups did not differ in age, gender composition or baseline ADHD subtype. Baseline ADHD symptom counts and observer ratings of inattention and hyperactivity were not significantly different for the children in the two subgroups (see Supplementary Table 2). The baseline neurocognitive performance of the subgroups was also similar (see Supplementary Tables 4 and 5). Significant group differences were found for Part B of the TMT [U = 65.0, $p < 0.001$] only, with the partial remission group completing the task more quickly. At follow-up a higher proportion of

children in the persist group (73.5% vs 52.4%) were taking methylphenidate for symptom management ($\chi^2 = 10.53$, $df = 2$, $p = 0.005$).

Significant group differences were found for FSIQ [$F(1,45) = 4.88$, $p = 0.032$] and PIQ [$F(1,45) = 7.918$, $p = 0.007$], with children in the partial remission group obtaining higher scores. Scores of the partial remission group increased over time, while scores for the persist group were similar at both assessments, however the group by time interaction was not significant (see Supplementary Table 4). Rates of reliable change for the two groups were not significantly different (see Table 4).

Analyses of WRAT-III scores indicated a significant group effect for arithmetic [$F(1,53) = 5.811$, $p = 0.019$], with the partial remission group performing better (see Supplementary Table 4). Rates of reliable change did not differ across the two groups, nor did the proportion of children in each group showing improvement or decline in at least one subject area (see Table 4). However, more children in the persistent ADHD group demonstrated a reliable decline in at least two subject areas (Fishers exact test $p = 0.019$).

Descriptive statistics for the executive function tests are presented in Supplementary Table 5. For the ACPT, within group comparisons showed a significant decrease in total errors for children in the partial remission group [Z = -2.154, $p = 0.031$] from baseline to follow-up. At follow-up children in the persist group made more errors than those in the partial remission group [U = 186.5, p = 0.008]. There was a significant group effect [$F(1,33) = 7.303$, $p = 0.011$] for perseverative errors on the WCST, with children in the partial remission group making fewer errors. Under the free condition of the Design Fluency test, children in the persist group made more errors over time [$Z = -3.003$, $p = 0.003$]. At follow-up children in

the partial remission group made more novel responses than those in the persist group [U = 275, p = 0.043]. In the fixed condition, the persist group made significantly fewer novel responses [Z = -2.489, p = 0.013] and the partial remission group more errors [Z = -2.580, p = 0.010] from baseline to follow-up. At follow-up children in the partial remission group made more errors than those in the persist group [U = 126, p = 0.049]. From baseline to follow-up children in the persist group showed a significant increase in time to complete Part A of the TMT [Z = -2.165, p = 0.03], together with a significant decrease in the time to complete Part B [Z=-2.240, p = 0.025].

DISCUSSION

This study comprehensively examined the intellectual, academic, and executive function trajectories of children diagnosed with ADHD over a four-year period. Findings at the group level demonstrated stable intellectual and academic performances, with the exception of a decline in age corrected spelling performance. At both assessments the children's scores were significantly below population means. Performance on measures of executive function mostly followed a stable or improving course. Consistent with the findings of previous cross-sectional and longitudinal studies [4,7,12] improvements were observed for aspects of attentional control, cognitive flexibility and goal setting, and information processing. With the exception of the Design Fluency task, follow-up performance did not fall below population means.

Results from reliable change analyses offer a somewhat different picture of the academic trajectories of children with ADHD. Despite stable intellectual functioning and improvements in executive function, more than 20 percent of children showed a reliable decline in their age standardized reading, arithmetic or spelling performance, with half the

sample demonstrating a decline in at least one academic area. These analyses also highlighted the variability in the academic performance of children with ADHD, identifying more children than expected with reliable improvements in reading and arithmetic scores. Such analyses are not typically undertaken in longitudinal studies, raising questions about the true extent of academic performance changes in children with ADHD. These findings argue strongly for considering individual trajectories when conducting longitudinal research.

The "normalization" of performance on some of the neurocognitive measures over the four-year interval suggest gains in executive function exceeding age related improvement. This may reflect a "catching up" in the executive domain, possibly resulting from fine-tuning of neural connectivity [46] or more efficient use of cognitive resources. [15] Some improvement may also reflect prior exposure to the tasks, i.e., practice effects. [42] Age corrected and standardized scores were derived with data from children who completed the tasks once only, potentially leading to overestimates of improvement on some measures, e.g., the WCST. For other measures performance ceiling effects may contribute to over estimates of improvement, e.g., the ACPT is normed up to age 12 only, necessitating calculation and comparison of age corrected z scores. The absence of adequate test-retest data for the executive function tests prevented us from undertaking reliable change analyses with these data.

Our data indicate continued academic performance deficits, including age standardized performance declines while executive function improved and, on some measures, appeared to normalize. One possible explanation for this is the quality of the neurocognitive measures included. The WRAT-III is well normed across the study age range and provides alternate forms, removing potential practice effects. By contrast, the quality of the executive

performance test norms is variable and practice effects are not addressed. The skill sets assessed by these measures may also be important. The executive function tests measured problem solving capacities assessed under optimal conditions. Whereas the WRAT assessed academic knowledge acquired under variable conditions and levels of engagement. Lost learning opportunities may not be easily overcome even as other skills are "caught up". Furthermore, symptoms of ADHD likely continue to impact learning opportunities irrespective of executive function capabilities. The current study did not systematically assess additional factors that may have affected academic performance over time (e.g., level of resources, academic support, learning disabilities).

An important question for clinicians and researchers is whether neurocognitive performance predicts later outcomes, including symptom remission. At follow-up some children continued to meet the full diagnostic criteria for ADHD while others did not, allowing examination of the predictive value of neurocognitive function for symptom persistence over time. Baseline ADHD symptom counts and questionnaire ratings for the ADHD subgroups were similar, ruling out initial ADHD severity as a source of subgroup differences at baseline or follow-up. Neurocognitive function at baseline did not separate the groups. The partial remission group often evidenced better, but not statistically separable, performance to those with persistent ADHD, consistent with recent meta-analytic findings. [23] Over the two assessments group differences emerged on some measures of intellectual, academic, and executive function. The persistent ADHD group obtained lower FSIQ and PIQ scores. They showed poorer arithmetic performance, a greater decline in spelling scores and were more likely to show a reliable decline in academic performance across multiple subject areas. These data suggest the persistence of ADHD may be associated with more severe academic

difficulties, consistent with reports that children with ADHD under achieve academically in childhood and beyond.

With respect to executive functions, children in partial remission group out performed those in the persistent ADHD group on measures of attentional control, information processing, cognitive flexibility and goal setting. While the executive function trajectories of both groups improved over time, careful review of the data suggests the smaller non significant differences at baseline increased over time. When data from both assessments was pooled in the analyses the group differences reached significance. These findings offer some support for Halperin and Shultz's [15] theory. The executive function test performance of children whose symptoms declined over time is better than that of the children whose symptoms persisted. The current results do not, however, support the predictive value neurocognitive function for symptom persistence.

The subtle group differences we observed in the children's neurocognitive functioning may reflect the modest size of the samples, symptom overlap between the groups, or the quality of the neurocognitive tests, or some combination of these factors. While the children in the partial remission group no longer met diagnostic criteria for ADHD, they continued to demonstrate some symptoms of the disorder. More than half the children in the partial remission group were currently prescribed methylphenidate, suggesting ongoing impairment from these symptoms. Comparisons of the performance of children with persistent ADHD and those in full remission might identify larger differences. The tests of executive function administered likely rely on multiple brain regions including the hypothesized dysfunctional subcortical regions,[16,47] potentially reducing observed differences. As already noted, the normative data for the executive function measures is limited, potentially reducing their

ability to detect differences. Additional longitudinal studies, with larger more diverse samples and using better quality measures of executive function would seem to be justified.

Strengths

This study comprehensively assessed a broad range of neurocognitive functions longitudinally in a well-defined sample of children with ADHD. Diagnostic practices followed international guidelines and the same neurocognitive measures were administered at baseline and follow-up. The findings emphasize the importance of examining performance at the individual as well as the group level, the former revealing important differences not detected at the group level. This was especially true for academic performance, where group comparisons failed to identify the extent of the decline in academic performance over time.

Limitations

As already discussed, the norms for the executive function measures are limited and adequate test-retest data is lacking. They were included to ensure adequate characterization of the executive domain, and because they were the best clinical measures available at the time. While every effort was made to ensure accurate measurement of each subdomain, some of the improvements in executive function may reflect practice effects rather than genuine improvement, particularly on measures such as the WCST. It is also important to acknowledge that executive function measures involve multiple domains, including non-executive processes.[16,47] As cognitive tests purporting to measure executive function become more sophisticated, researchers should attempt to include a measure for each of the individual executive subdomains in their executive batteries to increase clarity regarding the longitudinal trajectories of executive function in individuals with ADHD. Only performance-based measures of executive function, administered under optimal conditions, i.e., structured

settings with clear expectations, were included in the study. Future longitudinal research should consider including rating scales of executive functioning, designed to assess these skills in less structured and more complex everyday settings. [48]

Attrition of the original sample occurred from baseline to follow-up, resulting in modest size samples. We cannot rule out the possibility that families who choose not to participate at follow-up were those in which the children were managing well. Our longitudinal sample may therefore be biased towards children with greater neurocognitive difficulty and/or symptom persistence at follow-up. Importantly, analyses did not reveal any significant differences in demographic characteristics, ADHD subtypes, or comorbidity between participants who did and did not take part in the follow-up assessment.

For children prescribed methylphenidate parents and teachers were not explicitly told to rate the children's behavior off medication, potentially leading to under estimation of symptoms and placement in the partial remission rather than persistent ADHD group. Review of the symptom report patterns for children in the partial remission group who were and were not prescribed medication suggests this did not occur. Moving from the ADHD group (initial assessment) to the partial remission group is more likely a function of developmental/maturational factors, including development of the prefrontal cortex and its associated circuitry.

Conclusions

Here we examined the neurocognitive trajectories of children diagnosed with ADHD from childhood to early adolescence. The data indicate stability in intellectual performance, stability or deterioration in academic performance, and improvement on aspects of executive

function from baseline to follow-up. Children whose ADHD symptoms were in partial remission performed better on measures of neurocognitive function than those with persistent ADHD.

Study findings highlight the importance of monitoring academic performance over time and the need for academic remediation programs. The observed performance declines occurred despite 90% of the children being prescribed stimulant medication at some time during the four-year follow-up period. Results of individual change analyses caution against basing assumptions solely on group level comparisons, and emphasize the need for including both types of analyses when examining heterogeneous clinical populations, including ADHD.

REFERENCES

1. Polanczyk GV, Willcutt EG, Salum GA, et al. ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *Int J Epidemiol*. 2014;43:434-442.
2. Faraone SV, Biederman J, Mick E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. *Psychol Med.* 2006;36:159-165.
3. Lara C, Fayyad J, de Graaf R, et al. Childhood predictors of adult attention-deficit/hyperactivity disorder: results from the World Health Organization World Mental Health Survey Initiative. *Biol Psychiatry.* 2009;65:46-54.
4. Frazier TW, Demaree HA, Youngstrom EA. Meta-analysis of intellectual and neuropsychological test performance in Attention-Deficit/Hyperactivity Disorder. *Neuropsychology*. 2004;18:543-555.
5. Bridgett DJ, Walker ME. Intellectual functioning in adults with ADHD: a meta-analytic examination of full scale IQ differences between adults with and without ADHD. *Psychol Assess.* 2006;18:1-14.
6. Barkley RA. *Attention Deficit Hyperactivity Disorder: A handbook for diagnosis and treatment. 3rd ed.* New York, NY: Guildford Press; 2006.
7. DuPaul GJ, Volpe RJ. ADHD and learning disabilities: research findings and clinical implications. *Curr Attent Disord Reps.* 2009;1:152-155.
8. Daley D, Birchwood J. ADHD and academic performance: why does ADHD impact on academic performance and what can be done to support ADHD children in the classroom? *Child Care Health Dev*. 2010;36:455-464
9. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. *Psychol Bull.* 1997;121:65-94.

10. Brown TE. Executive functions and Attention Deficit Hyperactivity Disorder: implications of two conflicting views. *Int J Disabil Dev Educ.* 2006;53:35-46.
11. Castellanos FX, Sonuga-Barke EJ, Milham MP, et al. Characterizing cognition in ADHD: beyond executive dysfunction. *Trends Cogn Sci.* 2006;10:117-123.
12. Willcutt EG, Doyle AE, Nigg JT, et al. Validity of the executive function theory of Attention-Deficit/Hyperactivity Disorder: a meta-analytic review. *Biol Psychiatry.* 2005;57:1336-1346.
13. Lezak M. *Neuropsychological Assessment. 3rd ed*. New York, NY: Oxford University Press; 1995.
14. Anderson P. Assessment and development of executive function (EF) during childhood. *Child Neuropsychol.* 2002;8:71-82.
15. Halperin JM, Schulz KP. Revisiting the role of the prefrontal cortex in the pathophysiology of Attention-Deficit/Hyperactivity Disorder. *Psychol Bull.* 2006;132:560-581.
16. Chung HJ, Weyandt LL, Swentosky A. The physiology of executive functioning. In Goldstein S, Naglieri JA, eds. *Handbook of Executive Functioning*. New York, NY: Springer; 2014: 13-27.
17. Leh SE, Petrides M, Strafella AP. The neural circuitry of executive functions in healthy subjects and Parkinson's disease. *Neuropsychopharmacology*. 2009;35:70-85.
18. Strick PL, Dunn RP, Fiez JA. Cerebellum and nonmotor function. *Ann Rev Neurosci.* 2009;32:413-434.
19. Bedard AC, Trampush JW, Newcorn JH, et al. Perceptual and motor inhibition in adolescents/young adults with childhood-diagnosed ADHD. *Neuropsychology.* 2010;24:424-434.

20. Fischer M, Barkley RA, Smallish L, et al. Executive functioning in hyperactive children as young adults: attention, inhibition, response perseveration, and the impact of comorbidity. *Dev Neuropsychol.* 2005;27:107-133.
21. Halperin JM, Trampush JW, Miller CJ, et al. Neuropsychological outcome in adolescents/young adults with childhood ADHD: profiles of persisters, remitters and controls. *J Child Psychol Psychiatry.* 2008;49:958-966.
22. Robinson T, Tripp G. Neuropsychological functioning in children with ADHD: symptom persistence is linked to poorer performance on measures of executive and nonexecutive function. *Jpn Psychol Res.* 2013;55:154-167.
23. van Lieshout M, Luman M, Buitelaar J, et al. Does neurocognitive functioning predict future or persistence of ADHD? A systematic review. *Clin Psychol Rev.* 2013;33:539-560.
24. Coghill DR, Hayward D, Rhodes SM, et al. A longitudinal examination of neuropsychological and clinical functioning in boys with attention deficit hyperactivity disorder (ADHD): improvements in executive functioning do not explain clinical improvement. *Psychol Med*. 2014;44:1087-1099.
25. Mick E, Byrne D, Fried R, et al. Predictors of ADHD persistence in girls at 5-year follow-up. *J Atten Disord.* 2011;15:183-192.
26. Miller M, Loya F, Hinshaw S, P. Executive functions in girls with and without childhoo*d ADHD: developmental trajectories and associations with symptom change. J Child Psychol Psychiatry.* 2013;54:1005-1015.
27. Nyden A, Billsteadt E, Hjelmquist E, et al. Neurocognitive stability in Asperger syndrome, ADHD, and reading and writing disorder: a pilot study. *Dev Med Child Neurol*. 2001;43:165-171.

28. Hinshaw SP, Owens EB, Ami N, et al. Prospective follow-up of girls with attention-deficit/hyperactivity disorder into adolescence: evidence for continuing cross-domain impairment. *J Consult Clin Psychol*. 2006;74:489-499.
29. Massetti GM, Lahey BB, Pelham WE, et al. Academic achievement over eight years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4-6 years of age. *J Abnorm Child Psychol*. 2008;36:399-410.
30. Hinshaw SP, Carte ET, Fan C, et al. Neuropsychological functioning of girls with attention-deficit hyperactivity disorder followed prospectively into adolescence: Evidence for continuing deficits? *Neuropsychology*. 2007;21:265-273.
31. Fisher M, Barkley RA, Edelbrock CS, et al. The adolescent outcome of hyperactive children diagnosed by research criteria: II Academic, attentional, and neuropsychological status. *J Consult Clin Psychol.* 1990;58:580-588.
32. Biederman J, Petty CR, Fried R, et al. Stability of executive function deficits into young adult years: A prospective longitudinal follow-up study of grown up males with ADHD. *Act Psychiatr Scand*. 2007;116:129-136.
33. Silverman WK. *Anxiety Disorders Interview Schedule for Children*. Albany, NY: Graywind Publications; 1991.
34. Rutter M, Graham P. The reliability and validity of the psychiatric assessment of the child. I. Interview with the child. *Br J Psychiatry*. 1968;114:S63-S79.
35. Molina BSG, Pelham WE, Blumenthal J, et al. Agreement among teachers' behavior ratings of adolescents with a childhood history of attention deficit hyperactivity disorder. *J Clin Child Psychol.* 1998;27:330-339.
36. Korkman M, Kirk U, Kemp S. *NEPSY-II: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation; 2007.

37. Wechsler D. *Wechsler Intelligence Scale for Children - Third Edition.* San Antonio, TX: The Psychological Corporation; 1991.
38. Wilkinson GS. *The Wide Range Achievement Test: Administration Manual.* Wilmington, DE: Wide Range Inc; 1993.
39. Keith R. *The Auditory Continuous Performance Test.* San Antonio, TX: Psychological Corporation; 1994.
40. Reitan R. *Manual for Administration of Neuropsychological Test Batteries on Adults and Children.* Bloomington, IN: Indiana University Press; 1969.
41. Heaton RK, Chelune GJ, Talley JL, et al. *Wisconsin Card Sorting Test Manual: Revised and Expande*d. Odessa, FL: Psychiatric Assessment Resources; 1993.
42. Strauss E, Sherman EMS, Spreen O. *A Compendium of Tests and Assesment Techniques 3rd Edition.* New York, NY: Oxford University Press; 2006.
43. Elliot C. *British Ability Scales Manual 2: Technical Handbook.* Windsor, UK: NFER-Nelson; 1983.
44. Jones-Gotman M, Milner B. Design fluency: the invention of nonsense drawings after focal cortical lesions. *Neuropsychologia.* 1977;15:653-674.
45. Blakesley RE, Mazumdar S, Dew MA, et al. Comparisons of methods for multiple hypothesis testing in neuropsychological research. *Neuropsychology*. 2009;23:255-264.
46. Chelune GJ, Naugle RI, Lüders H, et al. Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology*. 1993;7:41-52.
47. Casey BJ, Tottenham N, Liston C, et al. Imaging the developing brain: what have we learned about cognitive development. *Trends Cogn Sci*. 2005;9:104-110.

48. Toplak ME, West RF, Stanovich KE. Practioner review: Do performance-based measures and ratings of executive function assess the construct? *J Child Psychol Psychiatry*. 2013;54:131-143.