HHS Public Access

Predicting the course of ADHD symptoms through the integration of childhood genomic, neural and cognitive features.

Gustavo Sudre, PhD[1], **Wendy Sharp, MSW**[2], **Paul Kundzicz, BA**[1], **Marine Bouyssi-Kobar, PhD**[1], **Luke Norman, PhD**[1], **Saadia Choudhury, MA**[1], **Philip Shaw, MD, PhD**[1,3,*]

[1]Neurobehavioral Clinical Research Section, Social and Behavioral Research Branch, National Human Genome Research Institute

[2]Office of the Clinical Director, National Institute of Mental Health.

[3]Intramural Program, National Institute of Mental Health.

Abstract

Childhood attention deficit hyperactivity disorder (ADHD) shows a highly variable course with age: some individuals show improving, others stable or worsening symptoms. The ability to predict symptom course could help individualize treatment and guide interventions. By studying a cohort of 362 youth, we ask if polygenic risk for ADHD, combined with baseline neural and cognitive features could aid in the prediction of the course of symptoms over an average period of 4.8 years. Compared to a never affected comparison group, we find that participants with worsening symptoms carried the highest polygenic risk for ADHD, followed by those with stable symptoms, then those whose symptoms improved. Participants with worsening symptoms also showed atypical baseline cognition. Atypical microstructure of the cingulum bundle and anterior thalamic radiation were associated with improving symptoms while reduction of thalamic volume was found in those with stable symptoms. Machine learning algorithms, trained and tested on independent groups, performed well in classifying those never affected against groups with worsening, stable and improving symptoms (area under the curve > 0.79). We conclude that some measures of polygenic risk, cognition, and neuroimaging show significant associations with the future course of ADHD symptoms and may have modest predictive power. These features warrant further exploration as prognostic tools.

Introduction.

One of the most fascinating aspects of ADHD is the high degree of individual variation in the course of its symptoms from childhood into early adulthood. Prospective studies find that while a majority show some symptomatic improvement, many show stable or worsening symptoms [1, 2]. Understanding the variables that underpin different symptom course might form the basis of prognostic tools to help stratify individuals for treatment. Here, we ask if genomic, neural and cognitive features might be associated with, and predict symptom

*Corresponding author: shawp@mail.nih.gov, Bldg 31, B1 B37, Bethesda, MD 20892, USA.

change, above and beyond the established associations with age and baseline symptom severity [3].

Recent advances in ADHD genomics, particularly the mapping through genome wide associations studies (GWAS) of how common genetic variations are associated with the disorder, prompt us to first consider if genomics may provide prognostic features [4]. Using GWAS data, the polygenic risk score can be calculated which provides an index of an individual's genomic risk for the disorder [5]. Here, we ask if polygenic risk for the diagnosis of ADHD may also predict its symptom course. This is informed by prior longitudinal twin studies that find that around half of the genetic factors contributing to the course of ADHD symptoms overlap with those tied to its onset [6]. Additionally, a population cohort study found the persistence of ADHD symptoms across adolescence was associated with higher polygenic risk, a finding we attempt to replicate in a cohort with detailed, prospective symptom mapping [7].

We also consider neural and cognitive features. Among neuroimaging measures, preliminary associations have been reported between adolescent ADHD symptom trajectories and baseline measures of neuroanatomy (specifically dimensions of the superior cerebellar vermis and anterior cingulate cortex), neurophysiology (altered theta, beta ratios in EEG) and functional connectivity (coupling between activity in the medial and dorsolateral prefrontal cortex) [8–11]. We build on these studies, using multimodal imaging to ask if baseline anatomic and microstructural properties of white matter tracts measured might be tied to future symptom course. We also consider cognition, looking at domains implicated in the pathogenesis of ADHD, specifically intelligence, processing speed, visuomotor integration and working memory [12–14].

In summary, we follow 362 subjects for an average of 4.8 years, drawing contrasts between a group who remained symptom-free and participants showing either stable, improving or worsening symptoms of ADHD. First, using regression-based methods, we characterize baseline differences between these dissimilar symptom course groups in atypical cognitive, genomic and neural features. We then use machine learning approaches for prediction. The machine learning analyses first learn how baseline features relate to group membership and then predict symptom course in a separate test set of participants.

Methods

Overall, 362 participants were recruited from communities surrounding the study site. The inclusion criteria were at least two clinician-led assessments of symptoms of ADHD, with a minimum duration of follow-up of one year (average duration of follow-up was 4.8 [SD 2.3] years). Additionally, all subjects were genotyped, allowing polygenic risk for ADHD to be calculated. The principal exclusion criteria were the presence of any DSM-5 disorder besides ADHD at study entry, IQ <70, and major neurological disorders. The cohort was composed of 155 singletons, and the remaining 207 children came from 85 families. Parental history of ADHD was determined from clinical interview or from self-report of a diagnosis of ADHD- Supplement A1. Age of onset of impairing symptom was ascertained by parental report. The study was approved by the Institutional Review Board of the

National Human Genome Research Institute. Parents or guardians of children and adult participants gave written consent and children gave assent.

Outcome measures.

We defined four groups based on the course of ADHD symptoms: never affected, and those with stable, improving or worsening symptoms. Symptoms of ADHD were ascertained by trained clinicians (PS or WS) using the Diagnostic Interview for Children and Adolescents for parents [15]. Two clinical assessments were obtained on 106 participants, three assessments on 79; and 4 or more assessments on 177 participants. At each clinical assessment, participants were categorized as having none or minimal symptoms (0, 1, or 2 symptoms of either inattention or hyperactivity-impulsivity), moderate symptoms (3, 4 or 5 symptoms) or severe symptoms (6 to 9 symptoms) - Supplement A1. The main analyses considered changes in both symptom categories combined. Results for change in inattention and in hyperactivity-impulsivity separately are given as supplemental material.

The **never affected group** was defined as those with consistently having no or minimal symptoms of both inattention and hyperactivity-impulsivity. The **stable-symptomatic group** showed the same symptom level at each observation; that is, either moderate or severe symptoms of inattention and/or hyperactivity-impulsivity throughout. The **improving group** showed a drop in symptom level, for example moving from severe to moderate or no/ minimal symptoms of hyperactivity-impulsivity, and/or inattention. The **worsening group** showed the opposite pattern, for example increasing from no/minimal symptoms to either moderate or clinical/level symptoms.

Baseline predictors.

Polygenic risk for ADHD.—Genotyping used the Illumina HumanOmniExpressExome array with genome build GRCh37/hg19. Following filtering, 544,897 SNPs were retained for PRS analysis -details in Supplement A2. PRSice calculated polygenic risk scores for ADHD [16] as the sum of disorder-associated alleles, weighted by effect sizes from the multi-ethnic genome wide association study for ADHD from the Psychiatric Genetics Consortium (which did not contain our recently genotyped cohort) [17]. We included a range of PRS at increasingly liberal significance thresholds (p < 0.0001, 0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5). Ten principal components (PCs) were included to control for population stratification along with a random intercept term for familial relatedness. Genotypes are being deposited in dbGAP for consenting individuals.

Neuroanatomic: High resolution T1 weighted anatomic sequences were collected on a 3-T GE Signa scanner. Parameters and quality control measures are detailed in Supplement A3 and included visual inspection by two raters of the 'raw' data and FreeSurfer reconstructions [18]. Following quality control, 267 scans were retained. The anatomic features were volumes of structures that have been most consistently implicated in ADHD by meta-analyses and pathophysiological models of the disorder: namely the striatum, the amygdala, the cerebellum, the cingulate, orbitofrontal and dorsolateral prefrontal cortex [19–22]. We did not conduct anatomic analyses at the voxel/vertex level in view of the sample size and combined right and left hemispheres given high correlations between homologous structures.

White matter tract microstructure: Diffusion tensor imaging (DTI) data were acquired using a single-shot, dual-spin echo, echo-planar imaging sequence on the same 3-T GE Signa scanner- Supplement A3. 166 data set passed quality control measures, including the reacquisition of corrupted data in real time, visual inspection, exclusion of participants with more than two corrupted volumes, and excessive movement. We calculated fractional anisotropy, which summarizes the diffusion of water molecules along and perpendicular to the axon, of the major white matter tracts (corpus callosum, anterior thalamic radiation, superior and inferior longitudinal, uncinate, and inferior frontal-occipital fasciculi, parahippocampal and medial frontal divisions of the cingulum bundle, and the corticospinal tract [23]).

Cognition.—General intelligence was assessed using age-appropriate versions of the Wechsler scales [24, 25]. Working memory spans were defined as the number of correctly recalled digits/tapping patterns in original and reverse order [26]. The ability to integrate visual and motor skills was assessed using the Beery-Butenika Developmental Test [27]. Processing speed used subtests of the Woodcock Johnson III Test of Cognitive Abilities [28]: the visual matching task which requires the matching of identical numbers under time pressure, and the decision speed task which requires the matching of two images under time pressure.

Demographic.—Familial socioeconomic status was measured by the Hollingshead scale which assesses parental educational attainment and occupation [29].

Analyses.

We first determined associations between baseline variables and symptom course groups. The three ADHD symptom course groups were coded as three dummy variables (stable, improving and worsening symptoms) with the never affected group as the reference group. These groups were entered into mixed model regressions, including a random intercept term for family to account for relatedness (using lmertest in R v.3.6.0). This approach provides t and p-values for the differences between each clinical and the never affected (reference) group via Satterthwaite's degrees of freedom method. Group differences were also expressed as effect sizes (using R package effectsize) (d), where $d{\sim}0.2$ is a small, $d{\sim}0.5$ is a medium, and $d{\sim}0.8$ is a large effect size [30]. Parental history of ADHD was dichotomized and its association with outcome groups assessed using mixed effects logistic regression. To adjust for multiple testing, we concatenated all the p values and applied a false rate discovery procedure and highlight findings that survived adjustment at q<0.05. All analyses included gender and age at study entry and exit as covariates. The neuroimaging analyses also included as covariates image quality metrics and the polygenic risk analyses included 10 PCs for population stratification.

We tested for associations between psychostimulant medication history among those with ADHD and symptom course groups- Supplement A6b. Using moderation analyses, we also determined if psychostimulant medication history impacted the associations between baseline characteristics and symptom course.

We next used machine learning approaches to identify the combination of baseline features that best predicts group membership. Unlike the univariate approach, which looks at the

contribution feature independently, machine learning algorithms can detect information conveyed by combinations of features, potentially improving group classification.

Our primary machine learning analysis used conditional random forests (implemented in R, caret_6.0, party_1.3–5 cforest). This approach fits several decision tree classifiers on sub-samples of the training data and uses weighted averaging to maximize prediction. Random forests can handle missing data, thus increasing sample size and reducing the chances of model overfitting. This conditional implementation uses unbiased variable selection to overcome the limitation of more general tree-based algorithms that favor the selection of variables with multiple cut-points [31]. For comparison, we provide results for a more conservative machine learning method (sparse linear discriminant analysis, ipred_0.9–9 slda). This identifies the linear combination of features that best models the differences between groups, and then predicts group membership using a Bayesian estimation of the probabilities that a new set of inputs belongs to each group- Supplement A4.

Prior to machine learning, data were split into training and test sets, such that the data in the test sets were only used to evaluate the trained models. The performance of each model in classifying between all pairwise group combinations are given as the area under the receiver-operating curve (AUC), sensitivity and specificity. The contribution of each feature to group classification was calculated following the permutation principle of the `mean decrease in standard accuracy' importance for conditional random forests.

Results

Symptom course groups.

There were 129 never affected participants; 63 showed stable symptoms; 110 showed overall symptom improvement and 60 showed worsening symptoms- details in Table 1 and Supplement A1. The symptom course groups did not differ in gender composition (χ^2=4.51, p=0.21), in number of observations (χ^2=4.98, p=0.55) nor the interval between observations (F=1.42, p=0.24). Pairwise contrasts showed that a group differences in age at study entry (F=2.49, p=0.06) arose as the worsening symptom group were younger at study entry than the never affected (p=0.008) with no other significant pairwise group differences. A group age difference at the last observation (F=6.75, p=0.001) arose as the improvers and never affected were older than the stable and worsening groups. Thus, age at study entry and the age of last observation were both entered as covariates in analyses.

Associations between baseline variables and symptom course groups.

Differences in baseline features between the symptom course groups and the never affected group varied by domain – Figure 1. For the genomic variables, the worsening symptom group showed a significantly higher polygenic risk for ADHD than the never affected group, that survived adjustment for multiple testing. Increased polygenic risk in the worsening symptom group held across multiple risk thresholds (at p<0.01, p<0.05, p<0.1, p<0.2, p<0.3, p<0.4, p<0.5), with a maximum effect of 0.52 (at threshold of p<0.2, with t=4.04, FDR adjusted p=0.003). By contrast, the groups with improving or stable symptoms showed generally non-significant increases in polygenic risk of small effect sizes.

The group with worsening symptoms also showed atypical cognition. The cognitive deficits in the worsening symptom group were associated with medium effects for IQ [in contrast against never affected, t=3.72,adjusted p=0.004, d=0.43]; spatial working memory [t=2.8,adjusted p=0.03, d=0.34] and visuomotor integration, [t=2.83,adjusted p=0.03, d=0.34]). The groups with stable and improving symptoms both showed significant atypical performance in processing speed, visuomotor integration and IQ. Socioeconomic status was not significantly associated with symptom course following adjustment for multiple testing.

Among neuroanatomic features, the stable symptom group showed the most atypical structure at baseline, the thalamus (t=2.47, nominal p=0.01, adjusted p=0.07, d=0.31). A nominally significant decrease in cerebellar white matter (t=2.2, nominal p=0.03) was also found in the stable symptom group, and no other significant group differences were noted.

Among white matter microstructural features, fractional anisotropy differed significantly between the group showing symptom improvement and the never affected for both the parahippocampal division of cingulum bundle (t=3.5, adjusted p=0.006, d=0.63) and the anterior thalamic radiation (t=2.86, adjusted =0.03, d=0.51). No other white matter microstructure anomalies survived adjustment for multiple testing.

Results for the groups defined by change in inattention and change in hyperactivity-impulsivity are given in Supplement A5. Higher polygenic risk and atypical cognition were significantly associated only with worsening inattention, not hyperactivity-impulsivity. Anomalies in thalamic volume and white matter microstructure of the cingulum were also more prominent among the groups defined on the basis of changing inattention.

In summary, those with worsening symptoms carried the highest polygenic risk for ADHD and had cognitive deficits at study entry. By contrast, those with improving symptoms had atypical microstructure of the cingulum bundle, and a group with stable symptom had atypical reduction in thalamic volume at baseline. These baseline associations were stronger with future change in inattention than change in hyperactivity-impulsivity.

Robustness analyses:

The pattern of results was largely unchanged when we confined analyses to those who met criteria for ADHD during the study - Supplement A6a. The proportions of those with ADHD who never, intermittently or continuously took psychostimulant medication did not differ between the three symptom course groups (χ^2=3.97, p=0.41). Additionally, the associations between baseline features and symptom course groups were not significantly moderated by history of psychostimulant medication– Supplemental A6b. The association between polygenic risk for ADHD and a worsening symptom course also held when analyses were confined to the largest subpopulation (white, non-Hispanic)– Supplement A6c. Results were also similar when the typically developing children who had affected siblings were excluded- A6d.

Machine learning analyses.

The performance of the random forest machine learning algorithm on a test set, not used in the training phase, is shown in Table 2. The algorithm performed well in using baseline

features to classify the never affected group from each of the three symptom course groups (all AUC > 0.79). The algorithm also performed well classifying the clinical groups (worsening against stable, AUC 0.81; worsening against improving symptom groups, AUC 0.94; and improving against stable AUC 0.77). A linear machine learning approach based on discriminant analysis differentiated the improving from worsening and never affected groups well (AUCs > 0.75) but performed less well in discriminating the worsening group from stable and never affected groups - Supplement A8.

The relative importance of the different variable types (cognitive, polygenic, anatomic and white matter microstructure) in the random forest classification of group membership is shown in Figure 2 and Supplement A8. Cognitive features combined contributed most to classification, particularly in distinguishing the never affected from each of the three affected groups. Polygenic risk provided the second most important set of features, contributing prominently to the discrimination of the worsening symptom group, followed by the neural features.

Discussion.

Three key findings emerge in this study of the relationships between baseline genomic, neural and cognitive features and the later course of ADHD symptoms. First, those with worsening symptoms, particularly of inattention, had the highest polygenic risk for ADHD and most atypical cognition. Second, neural features showed different associations with outcome groups. Specifically, baseline reduction of the thalamic volume was found in those who later showed stable symptoms, whereas atypical microstructure of the cingulum bundle and anterior thalamic radiation was prominent among those who later showed symptom improvement. Finally, using a machine learning algorithm we find these objective measures, obtained at study entry, can classify with reasonable accuracy groups defined on future symptoms course. Overall, in the machine learning analyses polygenic risk and cognitive performance emerged as the most important features in predicting future symptom course, followed by neural features.

We find that those with worsening symptoms, particularly of inattention, had the highest polygenic risk for ADHD. Earlier reports find that high polygenic risk for ADHD is associated with symptom persistence but did not consider the possibility that polygenic risk may be associated with a worsening symptom course, as we report [7]. This finding extends previous reports of cross-sectional associations between polygenic risk and clinical severity, cognition and academic impairment[32–35]. The potential prognostic utility of polygenic risk is likely to improve as increasingly large GWAS allow more refined measures of polygenic risk through stratification based on age. For example, the polygenic risk for ADHD derived from GWAS of adult ADHD, which by definition persists from childhood, might better predict adolescent symptom course compared to polygenic risk determined from childhood cohorts, which contain both those destined to persist or remit.

While all the outcome groups showed some atypical cognition at baseline, deficits were slightly more prominent in the worsening symptom groups (at FDR significant levels for IQ, working memory and visuomotor integration). This finding may inform the debate on

whether cognitive deficits drive later symptoms or if there is a reverse direction of effects, whereby children with higher symptoms of ADHD engage less in activities that require, and therefore, foster executive functioning skills (Carr et al., 2006). We find that weaker cognitive performance at study entry was associated with later symptoms in the worsening group, even among those with no or minimal symptoms at baseline, which would not be expected if weaker cognitive performance was merely a downstream consequence of earlier symptoms. Some other studies similarly find childhood cognitive measures, specifically intact working memory and low response time variability, are associated with later symptom improvement [36–38]. We extend this literature by delineating a cognitive profile that may indicate risk for symptom deterioration.

White matter microstructural anomalies as measured by fractional anisotropy were most pronounced among the group showing improvement. The anomalies localized to the parahippocampal division of the cingulum bundle which connects the cingulate gyrus with medial temporal regions. The cingulum bundle is pivotal in executive functions pertaining to attentional control and memory and cingulum anomalies are among the earliest reported white matter findings in ADHD [39, 40]. The cingulum is also the latest maturing white matter tract, suggesting a prolonged window for its plasticity, in keeping with a role in the processes underpinning symptom resolution [41].

Findings were robust to a range of potential confounders including psychostimulant medication use although we relied on parent recall for medication histories which can be prone to recall error [42]. While there were no significant gender differences, the worsening symptom course contained 73% males whereas the never affected group had 60% males. Larger cohorts or more prolonged follow-up could reveal gender differences in clinical outcomes [43, 44]. This study focused on those who had ADHD without comorbid disorders at study entry which enhances the specificity of findings, but limits generalizability. The three outcome groups were not equated in total symptom severity over childhood, and thus while the outcome groups capture change they may not characterize the lifetime burden of symptoms. We note however that the pattern of findings held when we confined analyses to those who met criteria for ADHD during the study and thus had more equivalent overall symptom experience. Additionally, the mean final age of observation was 13.1 years, meaning that our cohort was mostly too young to capture the emergence of some important comorbidities particularly substance misuse.

The utility of machine learning techniques in diagnosis and predicting treatment response has been demonstrated [45–47]; here we add a potential prognostic application. A machine learning algorithm based on conditional random forests performed well in using baseline features to distinguish between unaffected and affected groups, and between those with worsening as opposed to improving and stable symptoms (all AUC >= 0.79). Prognostic tools that leverage routinely available clinical data have been developed but their accuracy is modest, limiting their clinical utility[48]. Here we find that measures of genetic risk, neural features and cognition are associated with later symptom trajectories, and begin to evaluate the impact of incorporating such objective features into prognostic tools. While promising, these neurocognitive and genomic measures are not ready for routine clinical use. Further validation on independent cohorts is needed, particularly cohorts that include youth with

comorbid disorders followed over longer periods of time. The costs of these measures also need to be weighed against gain in predictive power. Nonetheless, the findings suggest that neurocognitive and genomic tools warrant further exploration as we work towards individualized treatments informed by likely prognosis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References

1. Sibley MH, Mitchell JT, Becker SP. Method of adult diagnosis influences estimated persistence of childhood ADHD: a systematic review of longitudinal studies. The Lancet Psychiatry 2016; 3(12): 1157–1165. [PubMed: 27745869]

2. Faraone SV, Biederman J, Mick E. The age-dependent decline of attention deficit hyperactivity disorder: a meta-analysis of follow-up studies. Psychological Medicine 2006; 36(2): 159–165. [PubMed: 16420712]

3. Caye A, Spadini AV, Karam RG, Grevet EH, Rovaris DL, Bau CH et al. Predictors of persistence of ADHD into adulthood: a systematic review of the literature and meta-analysis. European Child & Adolescent Psychiatry 2016; 25(11): 1151–1159. [PubMed: 27021056] :

4. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. bioRxiv 2017.

5. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide. Genome Research 2007; 17: 1520–1528. [PubMed: 17785532] :

6. Pingault JB, Viding E, Galera C, Greven C, Zheng Y, R. P et al. Genetic and environmental influences on the developmental course of attention-deficit/hyperactivity disorder symptoms from childhood to adolescence. JAMA Psychiatry 2015; **doi:**10.1001/jamapsychiatry.2015.0469 **T.doi:T**

7. Riglin L, Collishaw S, Thapar AK, Dalsgaard S, Langley K, Smith GD et al. Association of genetic risk variants with attention-deficit/hyperactivity disorder trajectories in the general population. JAMA psychiatry 2016; 73(12): 1285–1292. [PubMed: 27806167] :

8. Mackie S, Shaw P, Lenroot R, Pierson R, Greenstein DK, Nugent TF, 3rd et al. Cerebellar development and clinical outcome in attention deficit hyperactivity disorder.[see comment]. American Journal of Psychiatry 2007; 164(4): 647–655.:

9. Shaw P, Lerch J, Greenstein D, Sharp W, Clasen L, Evans A et al. Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention deficit/hyperactivity disorder. Archives of General Psychiatry 2006; 63(5): 540–549. [PubMed: 16651511] :

10. Clarke AR, Barry RJ, Dupuy FE, McCarthy R, Selikowitz M, Heaven PC. Childhood EEG as a predictor of adult attention-deficit/hyperactivity disorder. Clinical Neurophysiology 2011; 122(1): 73–80. [PubMed: 20598939] :

11. Whitfield-Gabrieli S, Wendelken C, Nieto-Castañón A, Bailey SK, Anteraper SA, Lee YJ et al. Association of Intrinsic Brain Architecture With Changes in Attentional and Mood Symptoms During Development. JAMA psychiatry 2019.

12. Martinussen R, Hayden J, Hogg-Johnson S, Tannock R. A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. Journal of the American Academy of Child & Adolescent Psychiatry 2005; 44(4): 377–384. [PubMed: 15782085] :

13. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biological Psychiatry 2005; 57(11)**:** 1336–1346. [PubMed: 15950006] **:**

14. Cook NE, Braaten EB, Surman CB. Clinical and functional correlates of processing speed in pediatric Attention-Deficit/Hyperactivity Disorder: a systematic review and meta-analysis. Child Neuropsychology 2018; 24(5)**:** 598–616. [PubMed: 28345402] **:**

15. Reich W, Welner Z, Herkanic B. Diagnostic Interview for Children and Adolescents (DICA-IV). Multi-Health Systems: North Tonawanda, NY, 1997.

16. Euesden J, Lewis CM, O'Reilly PF. PRSice: polygenic risk score software. Bioinformatics 2014; 31(9)**:** 1466–1468. [PubMed: 25550326] **:**

17. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nature Genetics Advance online publication 2018.

18. Fischl B FreeSurfer. Neuroimage 2012; 62(2)**:** 774–781. [PubMed: 22248573] **:**

19. Shaw P, Ishii-Takahashi A, Park MT, Devenyi GA, Zibman C, Kasparek S et al. A multicohort, longitudinal study of cerebellar development in attention deficit hyperactivity disorder. Journal of Child Psychology and Psychiatry 2018.

20. Hoogman M, Muetzel R, Guimaraes JP, Shumskaya E, Mennes M, Zwiers MP et al. Brain imaging of the cortex in ADHD: a coordinated analysis of large-scale clinical and population-based samples. American Journal of Psychiatry 2019; 176(7)**:** 531–542.**:**

21. Hoogman M, Buitelaar JK, Faraone SV, Shaw P, Franke B. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults–Authors' reply. The Lancet Psychiatry 2017; 4(6)**:** 440–441. [PubMed: 28495548] **:**

22. Rubia K Cognitive Neuroscience of Attention Deficit Hyperactivity Disorder (ADHD) and Its Clinical Translation. Frontiers in human neuroscience 2018; 12**:** 100. [PubMed: 29651240] **:**

23. Hua K, Zhang J, Wakana S, Jiang H, Li X, Reich DS et al. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. Neuroimage 2008; 39(1)**:** 336–347. [PubMed: 17931890] **:**

24. Wechsler D Wechsler Abbreivated Scale of Intelligence- Second Edition. PsychCorp2011.

25. Wechsler D Wechsler Preschool and Primary Scale of Intelligence- Fourth Edifition. Peason, PsychCorp2012.

26. Wechsler D Wechsler Intelligence Scale for Children- Fourth Edition. Pearson2003.

27. Beery KE. Beery VMI: The Beery-Buktenica developmental test of visual-motor integration. Minneapolis, MN*:* Pearson 2004.

28. Woodcock RW, McGrew KS, Mather N. Woodcock-Johnson III : Tests of Achievement. Riverside publishing: Houghton Mifflin Harcourt2001.

29. Hollingshead A Four Factor Index of Social Status. Yale University Department of Sociology: New Haven, 1975.

30. Rosnow RL, Rosenthal R. Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. Psychological methods 1996; 1(4)**:** 331.**:**

31. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: A conditional inference framework. Journal of Computational and Graphical statistics 2006; 15(3)**:** 651–674.**:**

32. Vuijk PJ, Martin J, Braaten EB, Genovese G, Capawana MR, O'Keefe SM et al. Translating Discoveries in Attention-Deficit/Hyperactivity Disorder Genomics to an Outpatient Child and Adolescent Psychiatric Cohort. Journal of the American Academy of Child & Adolescent Psychiatry 2019.

33. Nigg JT, Gustafsson HC, Karalunas SL, Ryabinin P, McWeeney SK, Faraone SV et al. Working memory and vigilance as multivariate endophenotypes related to common genetic risk for attention-deficit/hyperactivity disorder. Journal of the American Academy of Child & Adolescent Psychiatry 2018; 57(3)**:** 175–182. [PubMed: 29496126] **:**

34. Sudre G, Frederick J, Sharp W, Ishii-Takahashi A, Mangalmurti A, Choudhury S et al. Mapping associations between polygenic risks for childhood neuropsychiatric disorders, symptoms of attention deficit hyperactivity disorder, cognition, and the brain. Molecular psychiatry 2019**:** 1.**:**

35. Martin J, Hamshere ML, Stergiakouli E, O'Donovan MC, Thapar A. Genetic risk for attention-deficit/hyperactivity disorder contributes to neurodevelopmental traits in the general population. Biological psychiatry 2014; 76(8)**:** 664–671. [PubMed: 24673882] **:**

36. van Lieshout M, Luman M, Twisk JWR, Faraone SV, Heslenfeld DJ, Hartman CA et al. Neurocognitive Predictors of ADHD Outcome: a 6-Year Follow-up Study. Journal of Abnormal Child Psychology 2017; 45(2)**:** 261–272. [PubMed: 27395390] **:**

37. Sjöwall D, Bohlin G, Rydell A- M, Thorell LB. Neuropsychological deficits in preschool as predictors of ADHD symptoms and academic achievement in late adolescence. Child Neuropsychology 2017; 23(1)**:** 111–128. [PubMed: 26212755] **:**

38. Karalunas SL, Gustafsson HC, Dieckmann NF, Tipsord J, Mitchell SH, Nigg JT. Heterogeneity in development of aspects of working memory predicts longitudinal attention deficit hyperactivity disorder symptom change. Journal of abnormal psychology 2017; 126(6)**:** 774. [PubMed: 28782975] **:**

39. Makris N, Buka SL, Biederman J, Papadimitriou GM, Hodge SM, Valera EM et al. Attention and executive systems abnormalities in adults with childhood ADHD: A DT-MRI study of connections. Cerebral Cortex 2008; 18(5)**:** 1210–1220. [PubMed: 17906338] **:**

40. Chiang H-L, Chen Y-J, Lo Y-C, Tseng W-YI, Gau SS-F. Altered white matter tract property related to impaired focused attention, sustained attention, cognitive impulsivity and vigilance in attention-deficit/hyperactivity disorder. Journal of psychiatry & neuroscience: JPN 2015; 40(5)**:** 325. [PubMed: 25871496] **:**

41. Lebel C, Gee M, Camicioli R, Wieler M, Martin W, Beaulieu C. Diffusion tensor imaging of white matter tract evolution over the lifespan. Neuroimage 2012; 60(1)**:** 340–352. [PubMed: 22178809] **:**

42. Kuriyan AB, Pelham WE Jr, Molina BS, Waschbusch DA, Sibley MH, Gnagy EM. Concordance between parent and physician medication histories for children and adolescents with attention-deficit/hyperactivity disorder. Journal of Child and Adolescent Psychopharmacology 2014; 24(5)**:** 269–274. [PubMed: 24945086] **:**

43. Owens EB, Cardoos SL, Hinshaw SP. Developmental progression and gender differences among individuals with ADHD. In: Barkley RA (ed). Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment The Guilford Press2015.

44. Dalsgaard S, Mortensen PB, Frydenberg M, Thomsen PH. Conduct problems, gender and adult psychiatric outcome of children with attention-deficit hyperactivity disorder. The British Journal of Psychiatry 2002; 181(5)**:** 416–421. [PubMed: 12411268] **:**

45. Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF. From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. Neuroscience & Biobehavioral Reviews 2015; 57**:** 328–349. [PubMed: 26254595] **:**

46. Kim J-W, Sharma V, Ryan ND. Predicting methylphenidate response in ADHD using machine learning approaches. International Journal of Neuropsychopharmacology 2015; 18(11).

47. Luo Y, Alvarez TL, Halperin JM, Li X. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. NeuroImage: Clinical 2020**:** 102238. [PubMed: 32182578] **:**

48. Caye A, Agnew-Blais J, Arseneault L, Gonçalves H, Kieling C, Langley K et al. A risk calculator to predict adult attention-deficit/hyperactivity disorder: generation and external validation in three birth cohorts and one clinical sample. Epidemiology and psychiatric sciences 2020; 29.
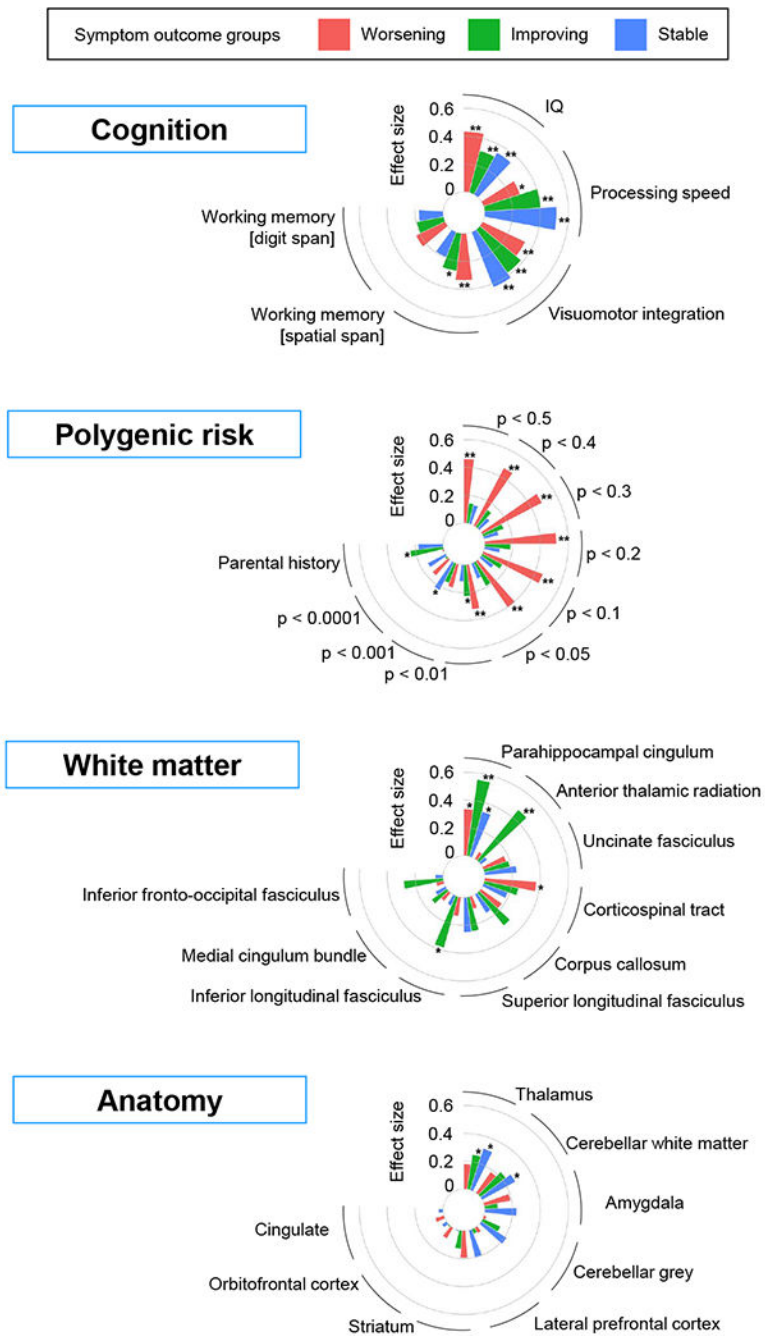
Figure 1:
Associations between symptom course groups and cognitive, polygenic risk white matter and anatomic features. ** = adjusted p-value < 0.05; * = unadjusted p-value < 0.05
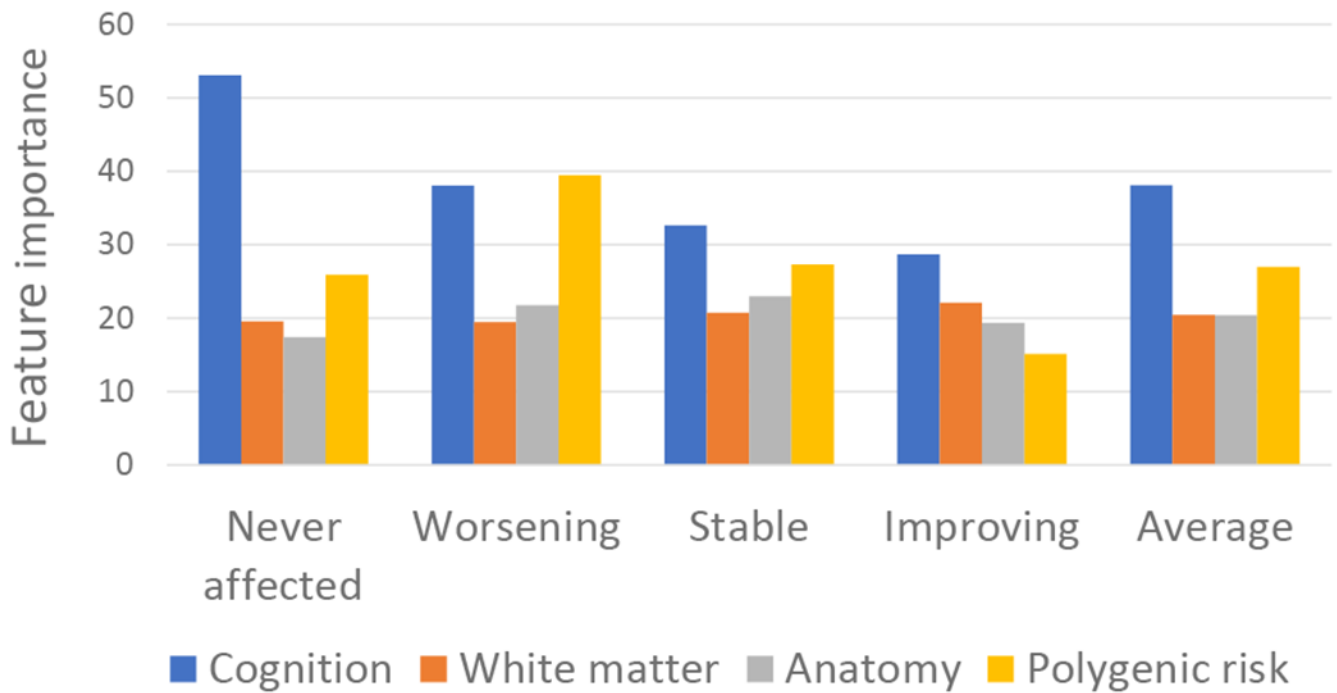
Figure 2.
The importance of each feature type in classification. The average for the importance scores for each type of feature (cognition, polygenic risk, anatomic, white matter microstructural) is shown. Values for individual features are given in Supplement A7. The first four bar groups (never-affected, worsening, stable, and improving) indicate the average feature importance in all pairwise comparisons involving that group. The final bar group ('average') indicates the average importance of feature type across groups.

Table 1.

The gender composition, age at study entry and end, number and interval between clinical observations and psychostimulant medication use for each symptom course group.

	Symptom course groups				
	Never affected	**Improvers**	**Stable**	**Worsening**	**Tests of group differences**
Total number	129	110	63	60	-
Male Female	79 (61%) 50 (39%)	81 (74%) 29 (26%)	44 (70%) 19 (30%)	42 (70%) 18 (30%)	χ^2 =4.52 p = 0.21
Age at baseline: mean (SD) years	8.7 (3.0)	8.2 (2.0)	8.1 (2.4)	7.6 (2.8)	$F_{(3,358)}$=2.49, p=0.06
Age at last observation	13.3 (3.1)	13.8 (2.7)	12.6 (2.9)	12.0 (3.1)	$F_{(3,358)}$=6.75, p=0.0001 [Improvers=never affected]> [stable=worsening]
Clinical 2 assessments 3 4 or more	38 (30%)	25 (23%)	24 (38%)	19 (32%)	χ^2=6.64, p=0.36
	29 (22%)	22 (20%)	13 (21%)	15 (25%)	
	62 (48%)	63 (57%)	26 (41%)	26 (43%)	
Inter-observation interval Mean (SD) years	1.97 (0.9)	2.16 (1.22)	2.15 (1.08)	2.02 (0.97)	$F_{(3,358)}$=0.85, p=0.47
Baseline inattention	0.1 (0.5)	6.2 (2.2)	6.3 (2.9)	2.3 (2.5)	$F_{(3,358)}$=281, p<0.00001, Never affected <worsening <[stable=improvers]
Baseline hyperactivity-impulsivity	0.1 (0.4)	5.26 (2.7)	4.6 (2.1)	3.0 (3.1)	$F_{(3,358)}$=109, p<0.00001, Never affected <worsening <[stable=improvers]
Intelligence quotient	115 (16)	108 (13)	107 (16)	106 (15)	$F_{(3,358)}$=8.1, p<0.00001, Never affected >[worsening=stable=improvers]
Psychostimulant treatment during study					
None Intermittent Constant	NA	23 (24%)	19 (33%)	15 (33%)	χ^2=3.97, p=0.41
	NA	54 (56%)	24 (41%)	22 (49%)	
	NA	19 (20%)	15 (26%)	8 (18%)	

Table 2.

Performance of a conditional random forest machine learning algorithm. Metrics are given for classification performance between all possible pairwise group combinations on a test set not used in the training phase. AUC= area under the curve of the receiver operating curve.

	Never affected			Stable			Improving		
	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC	Sensitivity	Specificity
Worsening	0.79	0.58	0.83	0.81	0.83	0.83	0.94	0.91	1
Improving	0.8	0.73	0.58	0.77	0.91	0.67			
Stable	0.83	0.75	0.67						