Genomic Structural Equation Modeling Offers Insight Into Expressive Differences Between Childhood and Late-Diagnosed ADHD: An Analysis of Genetic Overlap with External Traits

Hannah J. Gebhardt

Department of Psychology and Neuroscience, University of Colorado Boulder

Defense Date: March 23, 2023

Honors Thesis Advisor: Dr. Andrew Grotzinger

Defense Committee:

Dr. Andrew Grotzinger, Department of Psychology and Neuroscience, Thesis Advisor

Dr. Cinnamon Bidwell, Department of Psychology and Neuroscience, Honors Representative

Dr. Christine Brennan, Department of Speech, Language, and Hearing Sciences, Outside Reader

Abstract

Recent literature on attention-deficit hyperactivity disorder (ADHD) has identified differences in the genetic influence of age at first diagnosis. In particular, genome-wide association studies (GWAS) have found that childhood, persistent, and late-diagnosed categories of ADHD exhibit divergent comorbidity patterns. Based on these findings, we hypothesized that there would be significant differences in genetic risk sharing across childhood and late-diagnosed ADHD with other external traits. The traits selected for analysis included 3 dimensions of risk-taking behavior, *Automobile Speeding Propensity, Drinks Per Week,* and *Number of Sexual Partners*, and 4 dimensions of diet-related behavior, *Protein Intake, Fat Intake, Carb Intake, and Sugar Intake*. Results from Genomic SEM applied to investigate differences in genetic risk sharing with these external traits and the two ADHD subgroups revealed significant differences for number of sexual partners ($p = 0.003$) and automobile speeding propensity ($p = 0.001$). Practical implications of these findings are discussed and offer reason to reevaluate the current treatment of ADHD.

Genomic Structural Equation Modeling Offers Insight Into Expressive Differences Between Childhood and Late-Diagnosed ADHD: An Analysis of Genetic Overlap with External Traits

Attention-deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder in which affected individuals experience symptoms of inattention, hyperactivity, and impulsivity that onset in childhood (Khan & Faraone, 2006). As diagnoses have increased in recent years, ADHD has become one of the more common disorders experienced by the American population today. The American Psychiatric Association (APA) estimates that between 5 percent and 10 percent of U.S. children have been diagnosed, with two-thirds of cases persisting into adulthood (Asherson & Gurling, 2011). The diagnosis rate of ADHD is also significantly higher in males than females, with a ratio of up to 10:1, and females are often diagnosed later than males (Davies, 2014).

In diagnosis, ADHD can be specified by subtype and age of diagnosis. Subtypes are characterized with an inattentive or hyperactive emphasis, while age of diagnosis is determined by the age of an individual at the time of their first ADHD diagnosis. While many individuals with ADHD are diagnosed before the age of 18, some may have persisting symptoms from childhood to adulthood, or receive a diagnosis later in life. Importantly, this latter, late-diagnosed subgroup is conceptualized as something that still onsets in childhood but was not diagnosed until later in life. These diagnoses result in childhood-diagnosed, persistent, and late-diagnosed categorizations of the disorder, respectively.

Although increasing rates of diagnosis have drawn attention to ADHD, there is still insufficient data on its etiology more generally and, in particular, what distinguishes these subtypes. While previous research has provided evidence for a strong genetic influence on

ADHD development, specific aspects of the disorder remain unexplored (Rajagopal et al, 2022). In particular, the current literature lacks salient information on genetic differences between subtypes and ages of diagnosis (Rajagopal et al, 2022). Furthermore, current treatment of ADHD is universalized and fails to consider potential influence of these factors. Identifying these differences will be critical in obtaining a holistic understanding of ADHD, as well as increasing efficacy of treatment.

Like several other developmental disorders, ADHD has a relatively high comorbidity rate. Interestingly, patterns of comorbidity can change throughout the lifespan. In childhood, affected individuals seem to be at the greatest risk for developing additional neurodevelopmental disorders like autism spectrum disorder, learning disorders, oppositional defiant disorder, and conduct disorder (Demontis et al, 2019). As they move into adulthood, individuals with ADHD show a greater prevalence, relative to the general population, of substance use disorders, personality disorders, sleep disorders, obesity, and asthma (Demontis et al, 2019). Anxiety, mood disorders, hypertension, infections, migraines, and epilepsy are also common throughout the lifespan, as well as a heightened risk for premature death (Demontis et al, 2019). These findings are especially relevant to the lack of research on etiology and outcomes of childhood, persistent, and late-diagnosed ADHD, as they imply that age of diagnosis may impact presentation, as well as correlations with external traits.

To further investigate potential genetic influence on disorders like ADHD, the genome-wide association study (GWAS) has become a powerful research tool. A GWAS aims to obtain a cumulative examination of the genome, often by identifying genetic variants that differ across individuals, referred to as single-nucleotide polymorphisms (SNPs), that are relevant to a trait (Grimm et al, 2018). A GWAS conducted by Rajagopal et al. in 2022 evaluated the genetic

architecture of childhood, persistent, and late-diagnosed ADHD, identifying subsets of loci that were significantly associated with only one of the subgroups. Because these results suggest structural genetic differences between these ADHD subgroups, it is plausible that these differences might influence presentation and related aspects of the disorder.

In order to contribute to the limited research on the implications of diagnosing ADHD at different ages, the present study investigates genetic correlations of childhood and late-diagnosed ADHD with relevant external traits. We perform a multivariate genomic analysis using Genomic Structural Equation Modeling (Genomic SEM) in R to identify differences in genetic correlations (r_g) between childhood-diagnosed ADHD and late-diagnosed ADHD, with respect to risk-taking and diet-related traits. Genomic SEM utilizes GWAS summary statistics to synthesize their genetic correlations and single-nucleotide polymorphism (SNP) heritabilities to identify genetic divergence and convergence across genetically overlapping traits (Grotzinger et al, 2019). The risk-taking and diet-related traits were selected based on their potential relevance to symptoms of ADHD, such as impulsivity. Our results revealed significant findings for two of the selected risk-taking traits. These results offer insight into genomic differences of childhood and late-onset ADHD. Thus, they provide a basis for reevaluating modern treatment of the disorder, as all ages of diagnosis are currently treated the same.

Method

ADHD Sample Characteristics

The ADHD genome-wide association study (GWAS) data was derived from a nationwide population-based case-cohort sample through iPSYCH, and initially consisted of a population sample of 133,296 genotyped individuals (Rajagopal et al, 2022). Among this sample of

individuals, 14,878 were diagnosed with ADHD prior to the age of 18, resulting in the categorization of childhood-diagnosed ADHD. Within this same sample, 6,961 separate individuals were diagnosed with ADHD after the age of 18, and were identified as having late-diagnosed ADHD. We do not consider the persistent ADHD subgroup due to limited genetic signal for this group. These cases were evaluated alongside 38,303 controls, and all individuals were of Danish (European) descent. The sex distributions also differed for childhood-diagnosed and late-diagnosed ADHD groups, with 23% of childhood-diagnosed individuals and 41% of late-diagnosed individuals being female.

External Trait Characteristics

The external trait data for both risk-taking and diet were selected from samples within the UK Biobank (UKB; Table 1). The UKB is a large biomedical data resource with genotype, physiological, and medical information from ~500,000 individuals living in the UK in the age range of 40-69. We specifically utilize the Pan-UK Biobank public data resource (https://pan.ukbb.broadinstitute.org/) that provides the GWAS summary statistics for all UKB phenotypes. As all genetic analyses in the current research must be undertaken within a single ancestry group, and ADHD child and late-diagnosed GWAS were only available in European ancestry individuals, the risk-taking and diet GWAS were strictly analyzed for individuals of European descent.

For risk-taking, we specifically utilize the continuous phenotypes assessing "*automobile speeding propensity," "drinks per week," and "number of sexual partners,"* containing 404,291, 414,343, 370,711 participants, respectively. The use of these particular phenotypes was guided by prior work on externalizing traits (Karlsson Linnér et al, 2019). Diet was analyzed through

intake of the four macronutrients, protein, fat, carbohydrates, and sugar. The protein, fat, and carbohydrate samples each contained 268,922 individuals, while the sugar sample contained 235,391 individuals. The selection of these phenotypes was again guided by prior work, in this case in the macronutrient space (Meddens et al, 2020).

Genomic Structural Equation Modeling (Genomic SEM)

In order to analyze the genetic influence of childhood and late-diagnosed ADHD on the selected risk-taking and diet traits, our research team utilized the GenomicSEM software in R (Grotzinger et al, 2019). While preparing the GWAS summary statistics for analysis, the effective sample sizes for childhood and late-diagnosed ADHD were computed. A sum of effective sample size calculation was used to ensure an unbiased estimation of liability scale heritability within the cohorts (Grotzinger et al, 2023). Because of the spectral, polygenic nature of ADHD, this calculation acknowledges that a binary line of diagnosis is an arbitrary distinction on a continuous distribution of risk (liability). Additionally, the sum of effective sample size calculation corrects for ascertainment bias, reflecting the fact that cases with the disease outcome, in this case ADHD, are sampled at a rate higher than the population base rate.

Analysis began with munging the summary statistics in order to format them correctly for the LD-score regression function (on a Z-statistic metric, aligned to the same reference allele). The *munge* function requires SNP, allele, regression, and p-value information to create an LDSC-compatible file. The childhood and late-diagnosed ADHD traits, containing this information, were then run through 6 arguments. These arguments included naming the summary statistic file, defining the reference file used to align the alleles, naming the trait, listing the sample size, defining imputation quality and minor allele frequency (MAF) filters to exclude

extremes and potentially corrupt data, and ultimately running the *munge* function. Following the childhood and late-diagnosed ADHD traits, this process was replicated for the risk-taking traits (automobile speeding propensity, drinks per week, and number of sexual partners) as well as the diet traits (protein, fat, carbohydrate, and sugar intake).

After all traits were run through the *munge* function, they were provided as input to the *ldsc* function. LD-score regression (Bulick-Sullivan et al, 2015) uses the formatted GWAS summary statistics to estimate the genetic covariance matrix across included traits. This matrix includes the SNP-based heritabilities on the diagonal and genetic co-heritabilities (covariances) on the off-diagonal. This process included creating a vector of the munged summary statistics, entering a sample prevalence of 0.5 for the ADHD traits to reflect the munging using sum of effective sample size, creating a vector of the population prevalences, identifying the folder containing the LD scores and weights, naming the traits, and ultimately running the *ldsc* function.

After estimating LDSC, the matrices were loaded in and prepared for modeling. The usermodel function was used to identify the genetic correlations of the childhood and late-diagnosed ADHD traits, and two models were specified. The initial model was unconstrained, estimating how much childhood and late-diagnosed ADHD overlapped with genetic variance in each external trait. The second model was constrained and assumed that childhood and late-diagnosed ADHD have equal genetic overlap with each external trait. In order to test our initial hypothesis that there would be genetic differences between childhood and late-diagnosed ADHD and the external traits they overlap with, the fit of the unconstrained model was compared to the constrained model for each of the risk-taking and diet traits. We use a strict Bonferroni threshold to define significance ($p < 0.007$) to account for multiple testing.

Significant results established that the constrained model fits worse than the unconstrained model in some instances, meaning that childhood and late-diagnosed ADHD correlate differently for the given external trait.

Results

The GenomicSEM software in R was used to determine whether a relationship existed between age of diagnosis for ADHD and expression of external traits. Analyses focused on differences in fit of a constrained model, in which childhood and late-diagnosed ADHD were assumed to correlate equally with external traits, versus an unconstrained model, in which these genetic correlations were freely estimated. The comparison of these models revealed a stronger, positive correlation between late-diagnosed ADHD ($r_g = 0.435$) and number of sexual partners (NSP) relative to childhood-diagnosed ADHD and NSP ($r_g = 0.326$; $p = 0.003$). We also identify a significantly larger negative correlation between automobile speeding propensity (ASP) and late-diagnosed ADHD ($r_g = -0.109$) relative to childhood-diagnosed ADHD ($r_g = 0.018$; $p = 0.001$). The remaining diet and risk-taking traits did not meet thresholds of significance. These results offer support for our initial hypothesis that age of diagnosis for ADHD is differentially associated with the genetic signal for clinically relevant external traits.

Discussion

Based on population-specific data from European cohort samples, we identified significant differences between genetic correlations of childhood and late-onset ADHD with specific external traits. Of the 4 risk-taking traits analyzed, NSP and ASP demonstrated significantly different correlations with childhood and late-diagnosed ADHD. Both childhood

and late-diagnosed ADHD yielded positive correlations with NSP, with the late-diagnosed ADHD correlation being slightly stronger. Late-diagnosed ADHD also correlated negatively with ASP, although childhood-diagnosed ADHD did not significantly correlate with ASP at all. The significant model comparison found when comparing the ADHD traits and NSP implies that, when unconstrained, childhood and late-diagnosed ADHD correlate differently with NSP. Specifically, late-diagnosed ADHD is correlated with NSP, while childhood-diagnosed ADHD is slightly less correlated with NSP. Similarly, the significant value found when comparing the ADHD traits and ASP implies that, when unconstrained, childhood and late-diagnosed ADHD correlate differently with ASP. However, for this trait, late-diagnosed ADHD was negatively correlated with ASP, and childhood-diagnosed ADHD was not correlated with ASP.

These results indicate opposite findings. The positive correlations between the ADHD traits and NSP suggest that both categories of ADHD may be risk factors for increased sexual behavior. Alternatively, the negative correlation between late-diagnosed ADHD and ASP suggests that having late-diagnosed ADHD may be a protective factor against risky driving behavior. These significant differences in genetic overlap between childhood and late-diagnosed ADHD were unsurprising, as they are consistent with our initial hypothesis. However, it was unexpected that only two out of the 7 traits examined would yield significant results. As for the diet-related external traits, it may be worth noting that each sample size was nearly half that of the risk-taking traits. Although each sample contained a large amount of data, it is possible that these differences influenced the final effects. Furthermore, it is fair to assume that the risk-taking traits are more closely aligned with characteristics of impulsivity typically found in ADHD. While diet and macronutrient intake involve decision making, they may have more of a distant relationship with the symptoms of ADHD. Finally, this study was constrained to a single

population in order to avoid population stratification effects. Thus, these effects may not be generalizable to individuals of another, non-European population.

In order to continue the line of research on this topic, future studies might examine specific single nucleotide polymorphism differences between risk-related and protective external traits associated with ADHD. Although it appears that age of diagnosis for ADHD has a genetic impact on the expression of external traits, it is still unclear what the architectural genetic differences are. Moreover, further directions for research may also investigate performance-related differences in the current treatment of ADHD. Because age of diagnosis could impact external behaviors in individuals with ADHD, it is likely that one group may benefit more or less from the current treatment protocols.

In summary, our results reflect different correlations between ADHD categories, dependent on age of first-diagnosis. These results are population-based, as all data was collected within populations of European descent. Both late and childhood-diagnosed ADHD were found to positively and significantly correlate with NSP, although the late-diagnosed ADHD correlation was slightly stronger. Further, a negative correlation between late-diagnosed ADHD and ASP was identified, while childhood-diagnosed ADHD did not appear to correlate with significance. These results suggest that both childhood and late-diagnosed ADHD may be risk factors for an increased number of sexual partners, and late-diagnosed ADHD may be a protective factor against automobile speeding. In all, these results contradict the current image of equivalence that surrounds age of diagnosis for ADHD. Differences in patterns of genetic overlap with clinically relevant traits based on age of diagnosis for this disorder imply that when a person is diagnosed with ADHD matters behaviorally. Thus, the current belief that all age-diagnosis categories of ADHD should be treated the same is refuted by the findings in this study. Instead, these findings

offer support for differential treatment of ADHD subgroups, based on differential genetic architecture.

References

Asherson, P., & Gurling, H. (1970, January 1). *Quantitative and molecular genetics of ADHD*. SpringerLink. Retrieved May 27, 2022, from https://link.springer.com/chapter/10.1007/7854_2011_155

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015, February 2). *LD score regression distinguishes confounding from polygenicity in genome-wide association studies*. Nature Genetics. Retrieved March 3, 2023, from https://www.nature.com/articles/ng.3211

Davies, W. (2014, March 26). *Sex differences in attention deficit hyperactivity disorder: Candidate genetic and endocrine mechanisms*. Science Direct. Retrieved May 27, 2022, from https://www.sciencedirect.com/science/article/abs/pii/S0091302214000375?casa_token=V47CSGQyyY8AAAAA%3Art7KK_E4r8RXBwviHLW05XW9E3L5VlC7y19iGG9uMnVeziuqgF-_1DVdYeWWvD6cm0Ph4bUZFg

Demontis D., Walters RK., Martin J., Mattheisen M., Als TD., Agerbo E., Baldursson G., Belliveau R., Bybjerg-Grauholm J., Bækvad-Hansen M., Cerrato F., Chambert K., Churchhouse C., Dumont A., Eriksson N., Gandal M., Goldstein JI., Grasby KL., Grove J., Gudmundsson OO., Hansen CS., Hauberg ME., Holl. (2019, November 26). *Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder*. PubMedCentral. Retrieved July 12, 2022, from https://pubmed.ncbi.nlm.nih.gov/30478444/

Grimm, O., Kittel-Schneider, S., & Reif, A. (2018, May 2). *Recent developments in the genetics of attention-deficit hyperactivity disorder*. Wiley Online Library. Retrieved June 13, 2022, from https://onlinelibrary.wiley.com/doi/10.1111/pcn.12673

Grotzinger, A.D., Rhemtulla, M., de Vlaming, R. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits.*Nat Hum Behav* 3, 513–525 (2019). https://doi.org/10.1038/s41562-019-0566-x

Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, Lebreton M, Tino SP, Abdellaoui A, Hammerschlag AR, Nivard MG, Okbay A, Rietveld CA, Timshel PN, Trzaskowski M, Vlaming R, Zünd CL, Bao Y, Buzdugan L, Caplin AH, Chen CY, Eibich P, Fontanillas P, Gonzalez JR, Joshi PK, Karhunen V, Kleinman A, Levin RZ, Lill CM, Meddens GA, Muntané G, Sanchez-Roige S, Rooij FJV, Taskesen E, Wu Y, Zhang F; 23and Me Research Team; eQTLgen Consortium; International Cannabis Consortium; Social Science Genetic Association Consortium; Auton A, Boardman JD, Clark DW, Conlin A, Dolan CC, Fischbacher U, Groenen PJF, Harris KM, Hasler G, Hofman A, Ikram MA, Jain S, Karlsson R, Kessler RC, Kooyman M, MacKillop J, Männikkö M, Morcillo-Suarez C, McQueen MB, Schmidt KM, Smart MC, Sutter M, Thurik AR, Uitterlinden AG, White J, Wit H, Yang J, Bertram L, Boomsma DI, Esko T, Fehr E, Hinds DA, Johannesson M, Kumari M, Laibson D, Magnusson PKE, Meyer MN, Navarro A, Palmer AA, Pers TH, Posthuma D, Schunk D, Stein MB, Svento R, Tiemeier H, Timmers PRHJ, Turley P, Ursano RJ, Wagner GG, Wilson JF, Gratten J, Lee JJ, Cesarini D, Benjamin DJ, Koellinger PD, Beauchamp JP. Genome-wide association

analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet. 2019 Feb;51(2):245-257. doi: 10.1038/s41588-018-0309-3. Epub 2019 Jan 14. PMID: 30643258; PMCID: PMC6713272.

Khan, S. A., & Faraone, S. V. (2006). *The genetics of ADHD: A literature review of 2005* . SpringerLink. Retrieved May 27, 2022, from https://link.springer.com/article/10.1007/s11920-006-0042-y

Meddens SFW;de Vlaming R;Bowers P;Burik CAP;Linnér RK;Lee C;Okbay A;Turley P;Rietveld CA;Fontana MA;Ghanbari M;Imamura F;McMahon G;van der Most PJ;Voortman T;Wade KH;Anderson EL;Braun KVE;Emmett PM;Esko T;Gonzalez JR;Kiefte-de Jong JC;Langenberg C;Luan J; (2020, May 11). *Genomic analysis of diet composition finds novel loci and associations with health and lifestyle*. Molecular psychiatry. Retrieved February 20, 2023, from https://pubmed.ncbi.nlm.nih.gov/32393786/

Rajagopal, V.M., Duan, J., Vilar-Ribó, L. *et al.* Differences in the genetic architecture of common and rare variants in childhood, persistent and late-diagnosed attention-deficit hyperactivity disorder. *Nat Genet* 54, 1117–1124 (2022). https://doi.org/10.1038/s41588-022-01143-7

Thapar, A., & Stergiakouli, E. (2008, August). *An overview on the genetics of ADHD*. PubMedCentral. Retrieved May 27, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2854824/

Figure 1.

Sample sizes of each external trait.

Trait	DPW	ASP	NSP	PRO	FAT	CAR	SUG
Sample Size	414,343	404,291	370,711	268,922	268,922	268,922	235,391

Figure 2.

Genetic correlations between childhood and late-diagnosed ADHD with risk-taking and diet-related external traits. Figure depicts the point estimates for the genetic correlations for childhood and adult (late - diagnosed) ADHD. Error bars depict 95% confidence intervals obtained directly from the *GenomicSEM* software package. Stars are shown above bars that surpassed the Bonferroni corrected significance threshold when performing the model comparisons.