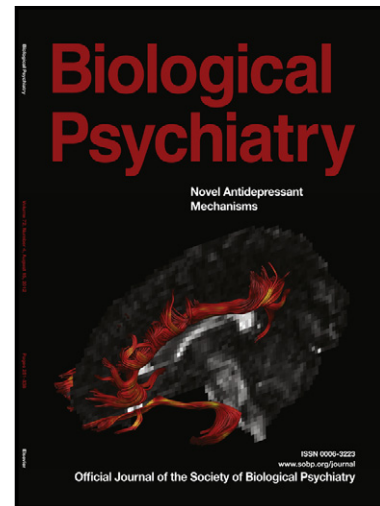


Author's Accepted Manuscript

Genetic Risk for Attention Deficit Hyperactivity Disorder Contributes to Neurodevelopmental Traits in the General Population

Joanna Martin BSc (Hons), Marian L. Hamshere Ph.D., Evangelia Stergiakouli Ph.D., Michael C. O'Donovan F.R.C.Psych., Ph.D., Anita Thapar F.R.C.Psych., Ph.D.



www.elsevier.com/locate/bps

PII: S0006-3223(14)00108-5

DOI: <http://dx.doi.org/10.1016/j.biopsych.2014.02.013>

Reference: BPS12148

To appear in: *Biological Psychiatry*

Cite this article as: Joanna Martin BSc (Hons), Marian L. Hamshere Ph.D., Evangelia Stergiakouli Ph.D., Michael C. O'Donovan F.R.C.Psych., Ph.D., Anita Thapar F.R.C.Psych., Ph.D., Genetic Risk for Attention Deficit Hyperactivity Disorder Contributes to Neurodevelopmental Traits in the General Population, *Biological Psychiatry*, <http://dx.doi.org/10.1016/j.biopsych.2014.02.013>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Abstract: 245**Article Body: 3998****Tables: 4****Figures: 2****Appendices: 7**

Genetic Risk for Attention Deficit Hyperactivity Disorder Contributes to Neurodevelopmental Traits in the General Population

Joanna Martin, BSc (Hons)^{1*}, Marian L. Hamshere, Ph.D.¹, Evangelia Stergiakouli, Ph.D.², Michael C. O'Donovan, F.R.C.Psych., Ph.D.¹, Anita Thapar, F.R.C.Psych., Ph.D.^{1*}

¹MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff UK

² Avon Longitudinal Study of Parents and Children, School of Social and Community Medicine, University of Bristol, Bristol UK

*Address for correspondence: Joanna Martin: Email: martinjm1@cardiff.ac.uk, Tel: +44 2920688390. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, UK

Key words: Avon Longitudinal Study of Parents and Children (ALSPAC), Attention Deficit Hyperactivity Disorder, Autism Spectrum Disorder, Social-Communication, Pragmatic Language, Genetics

Abstract

Background: Attention-deficit hyperactivity disorder (ADHD) can be viewed as the extreme end of traits in the general population. Epidemiological and twin studies also suggest that ADHD frequently co-occurs with and shares genetic susceptibility with autism spectrum disorder (ASD)/ASD-related traits. The aims of this study were to determine whether a composite of common molecular genetic variants, previously found to be associated with clinically-diagnosed ADHD, predicts ADHD and ASD-related traits in the general population.

Method: Polygenic risk scores were calculated in the Avon Longitudinal Study of Parents and Children (ALSPAC) population sample (N=8,229), based on a discovery case-control genome-wide association study of childhood ADHD. Regression analyses were used to assess whether polygenic scores predicted ADHD traits and also ASD-related measures (pragmatic language abilities and social cognition) in ALSPAC. Polygenic scores were also compared in males and females endorsing any (≥ 1) ADHD item (N=3,623).

Results: ADHD polygenic risk showed a positive association with ADHD (hyperactive-impulsive: $p=0.0039$; inattentive: $p=0.037$) traits. ADHD polygenic risk was also negatively associated with pragmatic language abilities ($p=0.037$), but not with social cognition ($p=0.43$). In children with a rating ≥ 1 for ADHD traits, females had a higher polygenic score than males ($p=0.003$).

Conclusions: These findings provide molecular genetic evidence that risk alleles for the categorical disorder of ADHD influence hyperactive-impulsive and attentional traits in the general population. The results further suggest that common genetic variation that contributes to ADHD diagnosis may also influence ASD-related traits, which at their extreme are a characteristic feature of ASD.

Introduction

Attention-deficit hyperactivity disorder (ADHD) is a highly heritable neurodevelopmental disorder, characterised by early onset and developmentally inappropriate inattentive, hyperactive and impulsive behaviours (1). The disorder occurs more frequently in males, with a male:female ratio of about 3-7:1 (2,3). Similarly to other common disorders, the genetic architecture of ADHD is complex, with rare and common variants involved (4). Whilst clinical diagnoses are defined categorically, ADHD psychopathology can also be viewed dimensionally, with inattentive and hyperactive-impulsive symptoms distributed continuously in the general population (5). Twin and epidemiological studies find that heritability estimates for dimensional ADHD are similar across a variety of cut-off points (6,7). This indicates that genetic factors act throughout the full distribution of ADHD symptoms. However, the postulated relationship between dimensional measures of ADHD in the population and clinical diagnoses has not yet been confirmed at the level of molecular genetics.

In recent years, it has become clear that the boundaries between different neurodevelopmental and psychiatric disorders are not clear cut, as exemplified by the observed clinical and genetic overlap between ADHD and other disorders. Rates of co-occurrence are especially high for ADHD and autism spectrum disorder (ASD), another highly heritable neurodevelopmental disorder, characterised by social communication/interaction deficits, as well as restrictive and repetitive behaviours (8). Studies of children with clinical diagnoses have found that large (>500kb), rare (<1% frequency) copy number variants (CNVs) in ADHD show significant overlap with CNV loci previously implicated in ASD (9,10), although a recent collaborative cross-phenotype analysis found no clear common genetic overlap in diagnosed ADHD and ASD cases (11). ASD can also be viewed dimensionally (12) and twin studies find that ADHD and ASD traits share common genetic influences in the general population, as well as at the quantitative extreme (13-19). This suggests that genetic variants associated with ADHD diagnosis might also contribute to population variation in ASD-related trait measures.

Previous research suggests that children clinically diagnosed with ADHD (N=452) differ from controls (N=5,081) on the basis of a polygenic risk score, an aggregate score of thousands of common alleles of very small effect which together index genetic risk for ADHD (20). In this paper, we test the hypothesis that, *en masse* common genetic variants that confer risk for a clinical ADHD diagnosis are associated with ADHD traits in the general population. Moreover, given the established clinical and genetic overlap between ADHD and ASD (13,14,16), we undertake analysis of the secondary hypothesis that, *en masse* ADHD common genetic variants are also associated with ASD-related/social-communication traits in the general population.

Method

Target population sample – ALSPAC

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a large, well-characterised longitudinal dataset (21,22). ALSPAC originally recruited N=14,541 pregnant women resident in Avon, England, with expected delivery dates of 01.04.91-31.12.92. An additional 713 eligible children whose mothers did not enrol during pregnancy were enrolled after age 7, resulting in a total sample of N=14,701 of children alive at age 1 year. Full data (both phenotypic and genotypic) were available for up to N=5,661 children, depending on the outcome variables. Children with more than 30% missing items on any outcome variable were excluded from analyses of that variable. The study website (<http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/>) contains details of all available data. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and Local Research Ethics Committees.

Phenotypic measures

Data on ADHD traits were collected when participants were aged approximately 7 years, 7 months old, using the parent Development and Well-Being Assessment (DAWBA)(23). For each ADHD item, parents marked boxes to say whether their child showed the behaviour; these were coded: 0 for

“no”, 1 for “a little more than others” and 2 for “a lot more than others”. A total ADHD trait score was calculated by summing these responses to give a possible range of 0-36. Scores were also calculated for inattentive and hyperactive-impulsive ADHD traits, separately (with a possible range of 0-18 each).

Social-communication traits were assessed using the Social and Communication Disorders Checklist (SCDC)(24) and the pragmatic language scales of the Children’s Communication Checklist (CCC)(25). A quantitative measure of restricted, repetitive behaviours was not available. Both the CCC and the SCDC have been shown to have good predictive reliability for a clinical diagnosis of ASD in the ALSPAC sample (26). The CCC shows good inter-rater reliability (0.80), internal consistency (0.80-0.87) and validity for language problems (25) and the SCDC shows good internal consistency (0.93), high test-retest reliability (0.81) and validity for a diagnosis of ASD (24). The SCDC assesses social cognition and understanding, whereas the CCC pragmatic language scales measure ability to use language in a social context. Previous research has shown that children with ADHD or ASD have lower pragmatic language ability scores than typically developing controls, but those with ASD have lower scores than those with ADHD (27).

The SCDC was assessed at the same time as the DAWBA ADHD measures. Parents were asked to judge how much 12 descriptions applied to their child’s behaviour. The responses were coded: 0 for “not true”, 1 for “quite/sometimes true” and 2 for “very/often true”. A total SCDC score was calculated by summing these responses (with a possible range of 0-24).

An abridged version of the CCC was used to assess language abilities at the approximate age of 9 years, 7 months. Parents were asked to rate whether statements about their child were “certainly true”, “somewhat true” or “not true”, which were coded as 0, 1 and 2, respectively. The following sub-scales were summed to generate a pragmatic language abilities score: inappropriate initiation, coherence, stereotyped conversation, conversational context and conversational rapport. Sub-scale scores were based on 6-8 items each. The pragmatic language total score was obtained for children

with data available for each subscale. As the CCC measures language abilities, lower scores suggest pragmatic language deficits.

Information on DSM-IV ADHD diagnoses is available based on the DAWBA at approximately age 7.

Data on ASD diagnoses are available based on clinical records, utilising a clinician's diagnosis of ASD (28). Scores on measures with <30% missing items were mean-imputed.

Genetic data

After quality control (QC), genome-wide data for 500,527 single nucleotide polymorphisms (SNPs) were available for N=8,229 of the children, of whom N=4,213 (51.2%) were male. Details of QC procedures can be found in Supplement 1.

Discovery clinical sample for generating ADHD polygenic risk scores

The analytic method described by the International Schizophrenia Consortium (ISC)(29) was used to identify ADHD risk alleles in a discovery GWAS from which polygenic risk scores were derived in the ALSPAC individuals. A published GWAS of British and Irish children with a confirmed DSM-IV research diagnosis of ADHD (N=727) and population controls (N=5,081) was used to define risk alleles. This clinical sample was selected as the primary discovery sample as it is similar to the ALSPAC general population in ethnicity and underwent similar diagnostic assessment processes. The ascertainment of DNA samples, QC procedures and GWAS results have been described in detail previously (4). This GWAS was based on 502,702 SNPs after strict QC. Following the ISC study, alleles that were more common in cases than controls at SNPs showing evidence for association at the very relaxed threshold $p < 0.5$ were considered risk alleles.

Generating polygenic scores

Full details are available in Supplement 1. In brief, SNPs in approximate linkage equilibrium in the ALSPAC genome-wide data were identified using PLINK (30). From this set of SNPs, we retained alleles which showed evidence for association ($p < 0.5$) in the discovery ADHD GWAS and used those

to calculate a polygenic score for each individual in ALSPAC, using PLINK (30). The polygenic scores were standardised using z-score transformations.

Data analysis strategy

In the ALSPAC sample, children with ADHD or ASD diagnoses were compared with each other and with the remainder of the sample on ADHD, SCDC and CCC traits, using Student's T-test. Females and males were also compared. Analyses were conducted on the 8,229 ALSPAC children with full genetic data available after all QC.

Due to a strongly negatively-skewed distribution of the CCC pragmatic language data, variables were transformed ($\ln x + 1$) and linear regression analyses were performed to test for association with ADHD polygenic score. ADHD and SCDC traits were highly positively-skewed, contained an excess of zero values and could not be transformed to normality (see Figure 1 for variable distributions). Analysing such data using standard linear regressions may yield biased estimates of parameters and increases Type I & II error rates (31,32). The distribution of data was better explained by a negative binomial than a Poisson distribution of simulated data with the same mean and N (see Supplementary Figure S1). Therefore, these data were analysed using zero-inflated negative binomial (ZINB) regression models. Gender was included as a covariate in all models.

The ZINB model consists of two sub-models that allow for a distribution with an inflated number of individuals with values of zero: a) logistic regression model of an unobserved dichotomous outcome to predict who has a score=0 and who has a score>0 and b) negative binomial model of the continuous outcome in those having a score \geq 0. Likelihood ratio tests were used to determine an overall p-value for each ZINB model in comparison to a null model, which included gender but not polygenic score. ZINB analyses were performed using Mplus version 7 (33).

For each association test, the amount of variance explained was calculated as the difference of Nagelkerke pseudo- R^2 in the full model, as compared with the null model. Given the non-

independence of the outcome variables, all results are interpreted using a significance threshold of $p < 0.05$.

Given that previous analysis of polygenic scores for ADHD in a clinical sample of children with ADHD showed that females had higher polygenic scores than males (20), a Student's T-test was used to test whether polygenic scores in children rating positive for any (≥ 1) ADHD trait in the target sample were significantly higher in females than males.

Where significant associations were observed, secondary analyses were run to determine whether the same associations could be detected for ADHD traits at a later time point (approximate age 10 years, 8 months). Replication was sought using a second ADHD GWAS discovery sample, that of the Psychiatric Genomics Consortium (PGC) (34). It contained 2,064 trios, 896 cases and 2,455 control individuals from four individual studies. A total of 54 cases (2% of the cases in this second sample) overlapped with the main discovery sample but could not be removed as only the summary statistics were available for this analysis.

Results

Sample phenotypic characteristics

Figure 2 presents descriptive statistics of the trait measures in children with no ADHD/ASD ($N=5,585$), those diagnosed with ADHD ($N=105$), ASD ($N=35$) or both ($N=8$). Of the children with a diagnosis of ADHD, 7.1% also had a diagnosis of ASD and 36.4% of those with ASD also had ADHD; this overlap was greater than would be expected by chance ($\chi^2=136.0$, $p < 0.001$).

As expected, ADHD traits were higher in children with a diagnosis of ASD than in those without ADHD or ASD (hyperactive-impulsive: $t=13.03$, $p < 0.001$; inattentive: $t=13.12$, $p < 0.001$). Children with ASD had lower levels of inattentive traits than children with ADHD ($t=-3.50$, $p < 0.001$) but did not differ significantly in terms of hyperactive-impulsive traits ($t=-1.70$, $p=0.09$).

Children with an ADHD diagnosis had significantly higher SCDC scores ($t=26.71$, $p<0.001$) and lower CCC pragmatic language scores ($t=-11.45$, $p<0.001$) than those without ADHD or ASD, but had lower SCDC scores ($t=-2.45$, $p=0.016$) and higher pragmatic language ability scores ($t=6.17$, $p<0.001$) than children with ASD. The ADHD and social-communication outcomes were moderately correlated (see Table 1).

As compared with males, females had significantly lower scores for ADHD (hyperactive-impulsive: $t=-12.48$, $p<0.001$; inattentive: $t=-13.06$, $p<0.001$) and SCDC ($t=-9.50$, $p<0.001$) and higher CCC pragmatic language ability scores ($t=6.44$, $p<0.001$).

Polygenic score analysis of ADHD and ASD-related/social-communication traits

The ADHD polygenic scores were based on 49,595 SNPs and were normally distributed in the ALSPAC sample ($N=8,229$). Among children with any ADHD traits (≥ 1 ; $N=3,623$), females had a higher polygenic score than males ($t=2.94$, $p=0.003$, Cohen's $d=0.098$). This is not attributable to an overall population difference on polygenic score by gender ($t=1.59$, $p=0.11$; $N=8,229$). Gender was included as a covariate in all further analyses.

Results of associations of ADHD polygenic score with the ADHD and social-communication outcomes are shown in Table 2. ZINB models show that ADHD polygenic risk predicted ADHD total scores ($R^2=0.005$, $p=0.0026$), hyperactive-impulsive traits ($R^2=0.002$, $p=0.0039$) and inattentive traits ($R^2=0.002$, $p=0.037$). The ZINB models indicate that the association signal comes from the zero-inflated part (part a) of the model for all ADHD outcomes.

To further explore the contribution of polygenic scores to ADHD trait levels in those with non-zero scores, the population was split into three arbitrary groups, based on increasing trait score: children who scored zero ($N=2038$), those with low levels of ADHD (score=1-11; $N=2817$) and those with moderate-to-high levels of ADHD (score ≥ 12 ; $N=806$). ANOVA shows a significant group difference ($F=4.66$, $p=0.010$) and *post-hoc* tests revealed that children with no ADHD traits had a lower mean

polygenic score than those with ADHD scores of 1-11 ($p=0.022$) and also those with scores ≥ 12 ($p=0.037$). The difference between the two other groups was not significant ($p=0.80$).

ADHD polygenic scores showed significant association with lower CCC pragmatic language scores ($\beta=-0.028$, $p=0.037$). Exploration of whether findings were attributable to specific CCC subscales showed association with lower scores on the 'inappropriate initiation' and 'conversational context' subscales ($\beta=-0.034$, $p=0.009$, $\beta=-0.034$, $p=0.010$, respectively) but not with 'coherence', 'stereotyped conversation' and 'conversational rapport' (all $p>0.05$). No association was found between polygenic score and SCDC total score ($p>0.05$).

Structural equation modelling with ADHD and pragmatic language as correlated outcomes, confirmed that both constructs are independently predicted by polygenic score (see Supplementary Figure S2). The amount of variance explained (R^2) for all models was very small, although this estimate does not reflect the true magnitude of the genetic overlap as it is highly sensitive to sample size (29). Including the 10 EIGENSTRAT principal components as covariates in the analyses did not affect the results (see Supplementary Table S2).

Testing associations at age 10

The observed association between polygenic score and ADHD (at ~age 7.5 years) could also be seen at the later time point (~age 10.5 years, $N \geq 5,495$) for total ADHD traits ($R^2=0.004$, $p=0.012$) and hyperactive-impulsive traits ($R^2=0.003$, $p=0.039$), with weak association with inattentive traits ($R^2=0.002$, $p=0.055$). See Table 3 for details. Among children with any ADHD traits at age 10 (≥ 1 ; $N=3,316$), females had a higher polygenic score than males ($t=2.35$, $p=0.019$, Cohen's $d=0.082$).

Replication using second discovery sample

Polygenic scores based on the second discovery sample (34) were not significantly associated with ADHD traits at age 7 ($p>0.05$) but did show an association at age 10 with total ADHD traits ($R^2=0.001$, $p=0.019$) and hyperactive-impulsive traits ($R^2<0.001$, $p=0.018$), with weak association with inattentive traits ($R^2<0.001$, $p=0.055$) (see Table 4). Polygenic scores based on the second discovery

sample also showed an association with the CCC 'conversational context' subscale ($\beta=-0.031$, $p=0.017$) but showed no association with the CCC 'inappropriate initiation' subscale ($\beta=-0.006$, $p=0.37$).

In children with ADHD trait scores ≥ 1 at age 7, there was a trend for females to have a higher polygenic score than males, calculated using this second discovery sample ($t=1.80$, $p=0.071$, Cohen's $d=0.060$). At age 10, females had significantly higher polygenic scores than males ($t=2.18$, $p=0.029$, Cohen's $d=0.076$).

Discussion

As hypothesised, this study found that ADHD polygenic score, based on common genetic variants previously found to be associated with risk of a clinical diagnosis of ADHD, was also associated with ADHD traits measured at ages 7 and 10 years, in the general population. The importance of this finding is that it provides support at the level of molecular genetics for the hypothesis that ADHD represents the extreme end of traits present in the general population (6,7). The results also support the relevance of common genetic variants to ADHD (4), extending findings by showing they also act on non-clinical ADHD traits in a community sample.

The exploratory ANOVA results show that polygenic score, which is derived from common genetic variants relevant to clinical (i.e. severe) ADHD, predicted both low and high levels of ADHD traits in the general population. The ZINB analysis suggested that the association signal between polygenic score and ADHD traits originates from the zero-inflated part of the model (i.e. whether ADHD trait score was zero or non-zero). This result might be due to greater power at the lower end of ADHD traits, as progressively fewer children have higher levels of ADHD traits.

Consistent with previous literature in clinical and general population samples (15,16,35), children with diagnoses of ADHD had more ASD-related/social-communication problems than those without a diagnosis of ADHD or ASD, while children with ASD had more ADHD traits than those without either diagnosis. Interestingly, although children with ADHD had higher inattentive traits than those

with ASD, levels of hyperactive-impulsive traits in these two groups did not differ significantly. However, this could have been due to low power as few children in ALSPAC had a clinical ASD diagnosis.

Results of the genetic analysis also suggest that risk alleles for ADHD may contribute to phenotypic traits in the general population, beyond core ADHD features. Polygenic risk scores previously found to be associated with ADHD diagnosis were also nominally associated with pragmatic language abilities in the general population, but not with social cognition traits, as indexed by SCDC scores.

Secondary exploratory analyses suggested that the association of ADHD polygenic risk with pragmatic language score was driven by scores on the ‘inappropriate initiation’ and ‘conversational context’ subscales of the CCC. Some items in the ‘inappropriate initiation’ subscale may tap into impulsive ADHD behaviours (in particular, the CCC item “he/she talks too much”) but items in the ‘conversational context’ subscale (e.g. “he/she can understand sarcasm” or “he/she says things which are tactless or socially inappropriate”) have no apparent link with ADHD features. Overall, the findings suggest that risk variants for ADHD may have pleiotropic effects on closely-related but conceptually different neurodevelopmental traits in the general population. These findings also support those from a twin study, which found that ADHD traits at age 8 shared genetic effects and were most associated with ASD communication difficulties, rather than ASD social difficulties or stereotyped behaviours (17).

One possible advantage of the primary discovery ADHD sample used to derive risk alleles, over the replication sample, is its similarity to the ALSPAC cohort in terms of ancestry and geography, but nevertheless, the sample was relatively small (4). Analyses using a second, larger ADHD sample (34) show a partial replication of the primary analysis. Polygenic scores based on this sample predicted ADHD traits at age 10, though not at age 7. Similarly, although polygenic scores derived from the second ADHD dataset predicted pragmatic language problems, as assessed using the CCC ‘conversational context’ subscale, they did not predict variation on the CCC ‘inappropriate initiation’

subscale. These replication results suggest that the associations of ADHD polygenic score with ADHD traits and pragmatic language problems are robust. However, further replication is necessary to conclusively rule out possible type I error. They also further highlight the fact that absence of clear individually associated loci in current GWAS studies of ADHD reflects inadequate power of the GWAS samples rather than an absence of common susceptibility variants.

Although we found an association between ADHD polygenic score and pragmatic language abilities, there was no association with social cognition, as measured by the SCDC. A recent collaborative cross-phenotype analysis suggests that common GWAS variants do not contribute to the overlap in diagnoses of ADHD and ASD (11). Nevertheless, twin evidence is consistent in finding high heritability for neurodevelopmental trait measures and in showing shared genetic influences on ADHD and ASD (6,7,16). Thus, it is too early to discount the contribution of common variants to the overlap of ADHD and ASD, particularly in terms of continuously distributed traits. The current study points to a possible overlap between susceptibility to clinically diagnosed ADHD and pragmatic language difficulties at a trait level in the general population.

As expected, male children in ALSPAC had higher ADHD trait scores than females (16,36,37). However, a novel observation was that in the group of children with any ADHD symptoms at either age, females had higher polygenic scores than males. For polygenic scores based on the second discovery sample, there was a trend towards similarly higher scores in females at age 7 and significantly higher scores at age 10 years. These results support the previous observation that in children diagnosed with ADHD, females have higher polygenic scores than males (20). One limitation of the earlier study is that it was based on a clinical sample, so the gender difference may have reflected referral bias (i.e. referred females may have on average had a more severe phenotype). The present finding in an epidemiological sample argues against that, and suggests a different liability threshold for females than males, with females requiring a more extreme load of risk factors to manifest ADHD. This is consistent with non-molecular based studies; for example, one study

observed that siblings of females with ADHD have more ADHD symptoms than siblings of males with ADHD (38). Similar findings have been reported in non-identical twin children with ASD (39).

A limitation of this study was that although the SCDC and CCC measures of social cognition and pragmatic language are predictive of a clinical diagnosis of ASD in the sample (26), they are not strictly measures of the specific deficits required for an ASD diagnosis. Moreover, there was no reliable quantitative measure of restrictive and repetitive behaviours available. The finding of an association between ADHD polygenic score and pragmatic language deficits is potentially also relevant to the new DSM-5 category of 'social communication disorder' (40).

As the ALSPAC cohort is longitudinal, the sample suffers from attrition. Previous studies have determined that predictors of attrition include socioeconomic and pregnancy factors, as well as presence of behavioural difficulties, including ADHD, in the study child (41). Assuming that attrition results from the behavioural manifestation of genetic risk, resultant attrition bias is likely to reduce the correlation between risk scores and traits. Multiple imputation methods have been used previously for missing ALSPAC data but do not appear to alter association patterns (42).

Due to the relatively small ADHD GWAS discovery sample sizes, power to detect susceptibility variants is low and aggregate scores based on GWAS are likely to be based on a poor signal-to-noise ratio (4,34). This is a possible explanation for the relatively small amount of phenotype variance explained by polygenic scores in the current study, estimates of explained variance in this form of analysis being strongly affected by discovery sample size. Another limitation of the current study is that a small number (N=54) of cases overlapped in both discovery samples. Although $p < 0.5$ is frequently used as a threshold for calculating polygenic scores (29,43-45), this is largely a convention established on the basis of the optimal threshold in the study of schizophrenia which inspired the wider application of polygenic score analysis (29). As shown by modelling in that study, the optimal threshold depends on both genetic architecture and sample size and therefore other thresholds have the potential to show greater effects. A sensitivity analysis in the present study using a variety

of p-value thresholds for calculating polygenic scores demonstrated that observed effects are fairly consistent across various thresholds (see Supplementary Figure S3).

Conclusions

In the current study, polygenic risk previously found to be associated with clinical ADHD diagnosis predicted inattentive and hyperactive-impulsive traits in a general population sample. The study also indicates that common genetic variants associated with ADHD may also be associated with pragmatic language ability in the general population, a trait measure that is distinct from the core deficits of ADHD. The approach of testing genetic risks that contribute to dimensions that cut-across diagnostic categories, rather than using DSM diagnoses, is in line with the Research Domain Criteria (RDoC) framework (46), and is likely to be a valuable approach for future neurodevelopmental and psychiatric research. As the power of GWAS increases, this method has the potential to further explore the biological overlap of these traits.

Acknowledgements

We are grateful to Dr Stephan Collishaw for advice and comments on a draft of the manuscript. We are very grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 092731) and the University of Bristol provide core support for ALSPAC. The UK Medical Research Council also supports the authors. We acknowledge Peter Holmans, Michael Owen, Kate Langley, Nigel Williams, Lindsey Kent and Michael Gill for their contributions to the original clinical diagnostic data from which polygenic risk scores were derived and which was funded by the Wellcome Trust. We also acknowledge the Psychiatric Genomics Consortium (PGC) ADHD group.

Conflicts of Interest

All the authors report no biomedical financial interests or potential conflicts of interest.

References

1. Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, et al. Molecular genetics of attention-deficit/hyperactivity disorder. *Biol Psychiatry*. 2005 Jun 1;57(11):1313-23.
2. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and meta-regression analysis. *Am J Psychiatry*. 2007;164(6):942.
3. Lahey BB, Applegate B, McBurnett K, Biederman J. DMS-IV field trials for attention deficit hyperactivity disorder in children and adolescents. *The American Journal of Psychiatry*. 1994.
4. Stergiakouli E, Hamshere M, Holmans P, Langley K, Zaharieva I, Hawi Z, et al. Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. *Am J Psychiatry*. 2012;169(2):186-94.
5. Rodriguez A, Järvelin M-R, Obel C, Taanila A, Miettunen J, Moilanen I, et al. Do inattention and hyperactivity symptoms equal scholastic impairment? Evidence from three European cohorts. *BMC Public Health*. 2007;7(1):327.
6. Levy F, Hay DA, McStephen M, Wood C, Waldman I. Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. *J Am Acad Child Adolesc Psychiatry*. 1997;36(6):737-44.
7. Larsson H, Anckarsater H, Råstam M, Chang Z, Lichtenstein P. Childhood attention-deficit hyperactivity disorder as an extreme of a continuous trait: a quantitative genetic study of 8,500 twin pairs. *J Child Psychol Psychiatry*. 2011;53(1):73-80.
8. Rommelse NNJ, Franke B, Geurts HM, Hartman CA, Buitelaar JK. Shared heritability of attention-deficit/hyperactivity disorder and autism spectrum disorder. *Eur Child Adolesc Psychiatry*. 2010;19(3):281-95.
9. Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet*. 2010;376(9750):1401-8.
10. Williams NM, Franke B, Mick E, Anney RJL, Freitag CM, Gill M, et al. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13.3. *Am J Psychiatry*. 2012;169(2):195-204.
11. Smoller JW, Craddock N, Kendler K, Lee PH, Neale BM, Nurnberger JI, et al. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*. 2013;381:1371-9.
12. Constantino JN, Todd RD. Autistic traits in the general population: a twin study. *Arch Gen Psychiatry*. 2003 May;60(5):524-30.
13. Reiersen AM, Constantino JN, Volk HE, Todd RD. Autistic traits in a population based ADHD twin sample. *J Child Psychol Psychiatry*. 2007;48(5):464-72.
14. Ronald A, Simonoff E, Kuntsi J, Asherson P, Plomin R. Evidence for overlapping genetic influences on autistic and ADHD behaviours in a community twin sample. *J Child Psychol Psychiatry*. 2008;49(5):535-42.
15. Lundström S, Chang Z, Kerekes N, Gumpert CH, Råstam M, Gillberg C, et al. Autistic-like traits and their association with mental health problems in two nationwide twin cohorts of children and adults. *Psychol Med*. 2011;41(11):2423-33.

16. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. *Am J Psychiatry*. 2010;167(11):1357-63.
17. Taylor MJ, Charman T, Robinson EB, Plomin R, Happé F, Asherson P, et al. Developmental associations between traits of autism spectrum disorder and attention deficit hyperactivity disorder: a genetically informative, longitudinal twin study. *Psychol Med*. 2012;1(1):1-12.
18. Ronald A, Edelson LR, Asherson P, Saudino KJ. Exploring the relationship between autistic-like traits and ADHD behaviors in early childhood: findings from a community twin study of 2-year-olds. *J Abnorm Child Psychol*. 2010;38(2):185-96.
19. Polderman T, Hoekstra R, Vinkhuyzen A, Sullivan P, van der Sluis S, Posthuma D. Attentional switching forms a genetic link between attention problems and autistic traits in adults. *Psychol Med*. 2012:1-12.
20. Hamshere ML, Langley K, Martin J, Agha SS, Stergiakouli E, Anney RJ, et al. High Loading of Polygenic Risk for ADHD in Children With Comorbid Aggression. *Am J Psychiatry*. 2013 Apr 19.
21. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2012.
22. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Smith GD, et al. Cohort profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2012.
23. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: Description and Initial Validation of an Integrated Assessment of Child and Adolescent Psychopathology. *J Child Psychol Psychiatry*. 2008;41(5):645-55.
24. Skuse DH, Mandy WPL, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry*. 2005;187(6):568-72.
25. Bishop DVM. Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *J Child Psychol Psychiatry*. 1998;39(6):879-91.
26. Steer CD, Golding J, Bolton PF. Traits contributing to the autistic spectrum. *PLoS ONE*. 2010;5(9):e12633.
27. Geurts HM, Broeders M, Nieuwland MS. Thinking outside the executive functions box: Theory of mind and pragmatic abilities in attention deficit/hyperactivity disorder. *Eur J Dev Psychol*. 2010;7(1):135-51.
28. Williams E, Thomas K, Sidebotham H, Emond A. Prevalence and characteristics of autistic spectrum disorders in the ALSPAC cohort. *Dev Med Child Neurol*. 2008;50(9):672-7.
29. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009 Aug 6;460(7256):748-52.
30. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep;81(3):559-75.
31. Karaszia BT, van Dulmen MHM. Regression models for count data: Illustrations using longitudinal predictors of childhood injury. *J Pediatr Psychol*. 2008;33(10):1076-84.

32. Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. Zero-truncated and zero-inflated models for count data. *Mixed effects models and extensions in ecology with R*: Springer; 2009. p. 261-93.
33. Muthén LK, Muthén BO. *Mplus user's guide: The comprehensive modeling program for applied researchers*. Los Angeles: Muthen & Muthen. 1998.
34. Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2010;49(9):884-97.
35. Gadow KD, DeVincent CJ, Pomeroy J. ADHD symptom subtypes in children with pervasive developmental disorder. *J Autism Dev Disord*. 2006 Feb;36(2):271-83.
36. Keen D, Ward S. Autistic spectrum disorder. *Autism*. 2004;8(1):39-48.
37. Meltzer H, Gatward R, Goodman R, Ford T. *Mental health of children and adolescents in Great Britain*. London: London, The Stationery Office; 2000.
38. Rhee SH, Waldman ID. Etiology of sex differences in the prevalence of ADHD: An examination of inattention and hyperactivity-impulsivity. *Am J Med Genet B Neuropsychiatr Genet*. 2004;127(1):60-4.
39. Robinson EB, Lichtenstein P, Anckarsäter H, Happé F, Ronald A. Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences*. 2013.
40. Skuse DH. DSM-5's conceptualization of autistic disorders. *J Am Acad Child Adolesc Psychiatry*. 2012;51(4):344.
41. Wolke D, Waylen A, Samara M, Steer C, Goodman R, Ford T, et al. Selective drop-out in longitudinal studies and non-biased prediction of behaviour disorders. *Br J Psychiatry*. 2009 Sep;195(3):249-56.
42. Langley K, Heron J, Smith GD, Thapar A. Maternal and Paternal Smoking During Pregnancy and Risk of ADHD Symptoms in Offspring: Testing for Intrauterine Effects. *Am J Epidemiol*. 2012;176(3):261-8.
43. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA, et al. Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*. 2011 Oct;43(10):969-76.
44. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N, et al. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics*. 2011 Oct;43(10):977-83.
45. Hamshere ML, O'Donovan MC, Jones IR, Jones L, Kirov G, Green EK, et al. Polygenic dissection of the bipolar phenotype. *Br J Psychiatry*. 2011;198(4):284-8.
46. Insel TR, Cuthbert BN, Garvey MA, Heinssen RK, Pine DS, Quinn KJ, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167(7):748-51.

Tables

Table 1: Pearson correlation coefficients of ADHD and social-communication outcome measures

	ADHD HI	ADHD I	ADHD total	SCDC
ADHD I	0.71			
ADHD total	0.92	0.93		
SCDC	0.65	0.58	0.66	
CCC PL	-0.51	-0.48	-0.53	-0.51

ADHD: attention-deficit hyperactivity disorder; H-I: hyperactive-impulsive; I: inattentive; SCDC: Social and Communication Disorders Checklist; CCC PL: Children's Communication Checklist pragmatic language; All associations significant at $p < 0.001$

Table 2: Associations of polygenic score with ADHD and ASD-related phenotypes in ALSPAC (all analyses using gender as a covariate)

Outcome	N	ZINB count outcome			ZINB zero-inflated outcome			ZINB overall p	ZINB overall R ²	Linear regression*			
		β	SE	p	β	SE	p			β	SE	R ²	
ADHD total traits	5661	0.11	0.10	0.30	-0.06	0.02	0.005	0.0026	0.005	0.032	0.013	0.013	0.001
ADHD hyperactive-impulsive traits	5661	0.15	0.13	0.24	-0.05	0.02	0.024	0.0039	0.002	0.037	0.013	0.005	0.001
ADHD inattentive traits	5656	0.05	0.13	0.71	-0.05	0.02	0.019	0.037	0.002	0.023	0.013	0.084	0.001
SCDC total score	5653	0.15	0.19	0.45	0.02	0.04	0.67	0.43	<0.001	0.012	0.013	0.35	0.0002
CCC pragmatic language score	5641	N/A											

* Linear regression results of ADHD and SCDC phenotypes included only for ease of interpretation; main results shown in bold. ZINB: zero-inflated negative binomial; ADHD: attention-deficit hyperactivity disorder; SCDC: Social and Communication Disorders Checklist; CCC: Children's Communication Checklist. Polygenic scores derived using a threshold of $p < 0.5$ in the discovery sample GWAS results (see text).

Table 3: Secondary analysis – Associations of polygenic score with ADHD at age 10 years (all analyses using gender as a covariate)

Outcome	N	ZINB count outcome			ZINB zero-inflated outcome			ZINB overall p	ZINB overall R ²	Linear regression*			
		β	SE	p	β	SE	p			β	SE	p	R ²
ADHD total traits	5500	-0.05	0.12	0.68	-0.06	0.02	0.003	0.012	0.004	0.087	0.086	0.31	0.0002
ADHD hyperactive-impulsive traits	5505	-0.15	0.25	0.53	-0.06	0.02	0.012	0.039	0.003	0.019	0.043	0.66	3.4E-05
ADHD inattentive traits	5495	0.02	0.14	0.90	-0.04	0.02	0.021	0.055	0.002	0.076	0.051	0.14	0.0004

* Linear regression results included only for ease of interpretation; main results shown in bold. ZINB: zero-inflated negative binomial; ADHD: attention-deficit hyperactivity disorder. Polygenic scores derived using a threshold of p<0.5 in the discovery sample GWAS results (see text).

Table 4: Replication analyses – Associations of polygenic score based on second discovery sample with ADHD at both time points (all analyses using gender as a covariate)

Time	Outcome	N	ZINB count outcome			ZINB zero-inflated outcome			ZINB overall p	ZINB overall R ²	Linear regression*			
			β	SE	p	β	SE	p			β	SE	p	R ²
age 7	ADHD total traits	5661	0.11	0.11	0.30	-0.02	0.02	0.338	0.20	0.001	0.023	0.13	0.052	0.001
	ADHD hyperactive-impulsive traits	5661	0.05	0.10	0.58	-0.03	0.02	0.20	0.26	<0.001	0.020	0.013	0.12	0.0004
	ADHD inattentive traits	5656	0.18	0.20	0.39	-0.02	0.02	0.44	0.17	<0.001	0.027	0.013	0.043	0.001
age 10	ADHD total traits	5500	0.27	0.24	0.26	-0.02	0.02	0.45	0.019	<0.001	0.26	0.087	0.003	0.002
	ADHD hyperactive-impulsive traits	5505	0.30	0.39	0.44	-0.01	0.02	0.56	0.018	<0.001	0.13	0.043	0.003	0.002
	ADHD inattentive traits	5495	0.29	0.33	0.38	-0.01	0.02	0.65	0.055	<0.001	0.13	0.052	0.015	0.001

* Linear regression results included only for ease of interpretation; main results shown in bold. ZINB: zero-inflated negative binomial; ADHD: attention-deficit hyperactivity disorder. Polygenic scores derived using a threshold of p<0.5 in the discovery sample GWAS results (see text).

Figure Titles & Legends

Figure 1 – Histograms of ADHD & social-communication traits

ADHD: attention-deficit hyperactivity disorder; SCDC: Social and Communication Disorders Checklist; CCC: Children's Communication Checklist

Figure 2 – Mean z-scores of ADHD & social-communication outcomes, displayed by diagnostic group

*Scores reversed; ADHD: attention-deficit hyperactivity disorder; H-I: hyperactive-impulsive; I: inattentive; SCDC: Social and Communication Disorders Checklist; CCC PL: Children's Communication Checklist pragmatic language; Error bars represent standard errors of the mean

Fig 1

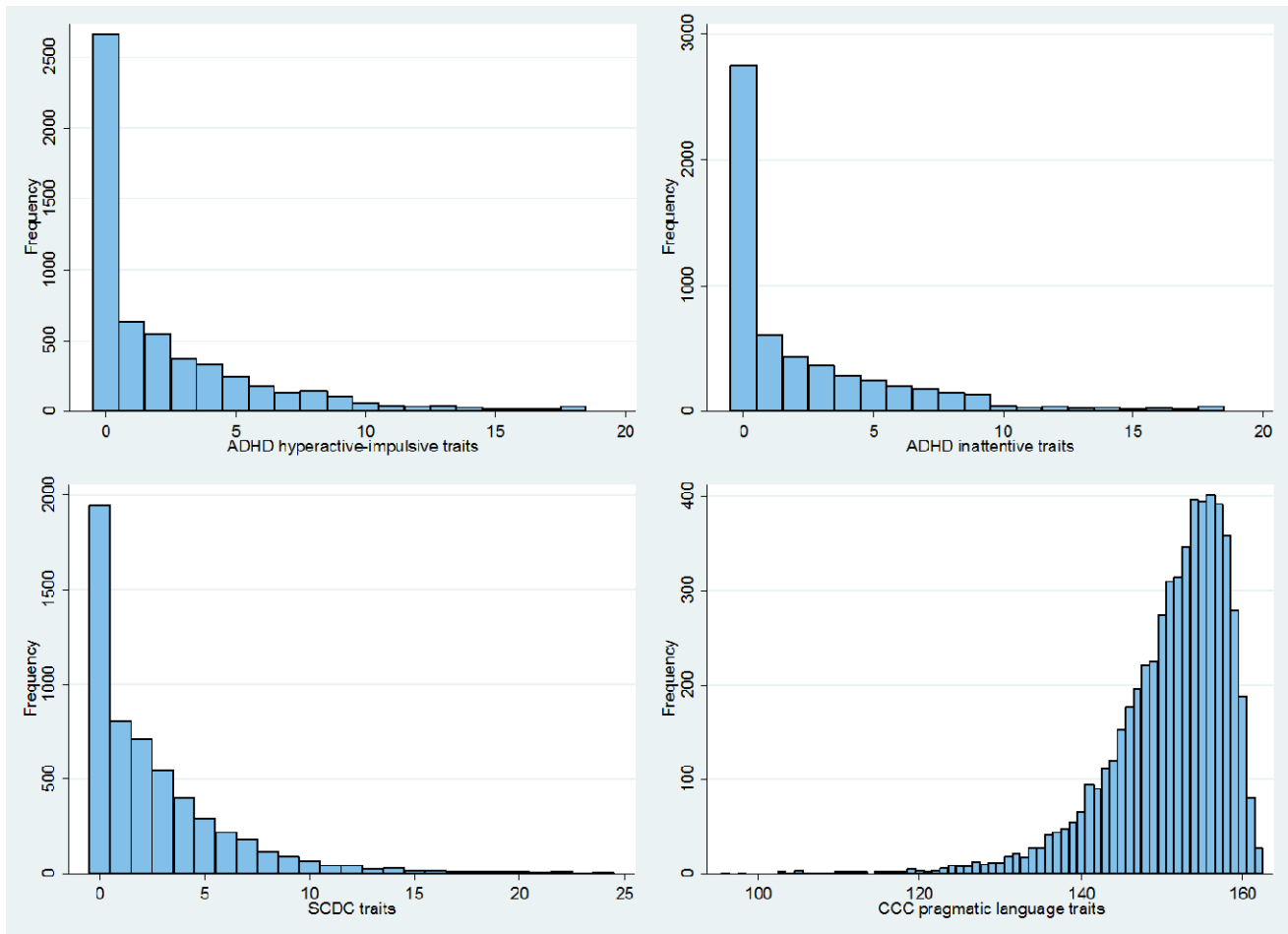


Fig 2

