Early manifestations of neurodevelopmental copy number variants in children: A population-based investigation

Charlotte A Dennison PhD[1,2], Joanna Martin PhD[1,2], Amy Shakeshaft PhD[1,2], Lucy Riglin PhD[1,2], Victoria Powell PhD[1,2], George Kirov PhD[2], Michael J Owen PhD[2], Michael C O'Donovan PhD[2], Anita Thapar PhD[1,2]

[1]Wolfson Centre for Young People's Mental Health, Cardiff University

[2]Centre for Neuropsychiatric Genetics and Genomics, Cardiff University

Corresponding author:

Prof Anita Thapar

E-mail: thapar@cardiff.ac.uk

Address: Hadyn Ellis Building, Cardiff University, Maindy Road, Cardiff, CF24 4HQ, UK

Running title: Neurodevelopmental CNVs in childhood

Key words:

ALSPAC; Millennium Cohort Study; Copy number variants; ADHD; ASD; neurodevelopmental conditions

Abstract

Background

There is clinical interest in recognising copy number variants (CNVs) in children as many have immediate and long-term health implications. Neurodevelopmental CNVs are associated with intellectual disability, autism spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD), conditions typically diagnosed by medical practitioners. However, neurodevelopmental CNVs may have additional, early developmental impacts that have yet to be examined in unselected populations.

Methods

Carriers of known ND CNVs were identified in two UK birth cohorts: the Avon Longitudinal Study of Parents and Children (ALSPAC) (carriers=144, controls=6217) and the Millennium Cohort Study (MCS) (carriers=151, controls=6559). In ALSPAC, we assessed associations between CNV carrier status and: birth complications, preschool development, cognitive ability, neurodevelopmental conditions (ASD, ADHD, reading, language, and motor difficulties), psychiatric, social and educational outcomes. Corresponding phenotypes were identified in MCS and meta-analysed, where available.

Results

In ALSPAC, neurodevelopmental CNVs were associated with low cognitive ability, ADHD and ASD. Neurodevelopmental CNV carriers showed greater likelihood of preterm birth, fine and gross motor delay, difficulties in motor coordination, language, and reading, and special educational needs (SEND). Meta-analysis with available measures in MCS identified elevated likelihood of ASD, ADHD, low birthweight, reading difficulties, SEND, and peer problems.

Discussion

Neurodevelopmental CNVs are associated with a broad range of developmental impacts. While clinicians who see children with intellectual disability, ASD, or ADHD may be aware of the impacts of CNVs and consider genetic testing, our investigation suggests that this training and awareness may need to extend to other professional groups (e.g. speech and language therapists).

Introduction

Copy number variants (CNVs) are a type of genetic variation that include deletions and duplications of chromosomal segments. Known neurodevelopmental (ND) CNVs, although rare with a population frequency <1%, are potentially important to recognise clinically, due to large effect sizes on risk for intellectual disability (1,2), ASD (3,4), ADHD (5,6), schizophrenia (7,8) and physical ill-health (9,10). Accumulating insights into the health impacts of CNVs, as well as the benefits to patients of disclosing genetic findings (11), have led to questions around the appropriateness of early genetic screening in high-risk groups. This raises the question of who should be considered at high-risk of carrying an ND-CNV. Many high-income countries already recommend screening individuals with developmental delay/intellectual disability for ND CNVs. A recent report further recommended that people with schizophrenia also ought to be offered the opportunity for CNV testing in certain circumstances (12).

To date, most studies have focused on adult or highly selected clinical (13,14) or volunteer (15,16) samples, and defined neurodevelopmental CNVs in diverse ways (17). These studies suggest elevated rates of cognitive impairment, learning problems, ADHD, ASD, and psychiatric diagnoses among rare CNV carriers compared to controls (15,16,18–21). The early developmental indicators of ND CNVs are yet to be fully characterised in unselected birth cohorts. Guidance on appropriate referral to clinical genetics services and screening require this evidence. Another uncertainty is whether risk estimates for outcomes from CNVs may have been overestimated by selection bias (21).

This study set out to examine associations between known ND CNVs and a range of developmental outcomes spanning birth and childhood, specifically: birth complications,

early motor and communication development, cognitive ability, broadly defined DSM-5 neurodevelopmental conditions, and social and educational outcomes, in two large UK birth cohorts.

Method

Participants

Full details of the cohort are provided in the supplementary methods. Pregnant women residing in Avon, UK with an expected delivery date between 1st April 1991 and 31st December 1992, were invited to participate in the Avon Longitudinal Study of Parents and Children (ALSPAC). The total sample size for analyses using any data collected after the age of seven is 15,447 pregnancies, of these, 14,901 children were alive at 1 year of age (22,23). The study website contains details of all the data that are available through a fully searchable data dictionary and variable search tool https://www.bristol.ac.uk/alspac/researchers/our-data. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time.

Families with a child born in the UK between 1st September 2000 and 11th January 2002 were invited to participate in the Millennium Cohort Study (MCS) (24). Children living in recruitment areas across the UK at 9 months of age whose family were eligible to receive child benefit were able to participate, which was almost all families in the UK at the time. The total number of enrolled children was 19,870, of whom 10,757 remained in the cohort

at age 17. Ethical approval was obtained from the NHS Research Ethics Committee. Parents provided informed consent for their child to participate and children provided assent. Data availability is described in the supplement.

CNV Calling and Annotation

Details of genotyping are provided in the supplementary methods. Genetic ancestry was inferred using the GenoPred pipeline (https://github.com/opain/GenoPred)(25), as described previously (26). CNV calling, quality control, and annotation were performed using the Cardiff Pathfinder pipeline (https://github.com/CardiffMRCPathfinder) in both cohorts. CNV calling is described in the supplement.

CNV annotation was performed in R. CNVs in 54 regions have been robustly associated with neurodevelopmental conditions as grouped in DSM-5 (defined as intellectual disability, ASD, developmental delay, and ADHD), are widely accepted to be pathogenic (1,27,28), and have been investigated extensively in the literature (9–11,29). We defined these CNVs according to previously published criteria (Table S1)(30). To meet our criteria for an ND CNV, CNVs spanning multiple genes were required to cover >50% of the critical interval and known key genes (defined in Table S1). For CNVs spanning one gene, deletions were required to cover at least one exon, whilst duplications were required to cover the whole gene. The presence of an ND CNV was confirmed through visual inspection of LRR and BAF plots for each identified ND CNV.

Outcomes

All outcome measures were dichotomised, using previously published cut-points, to facilitate clinical interpretation. For measures derived from multiple timepoints, individuals were included in the analysis if they had data for at least one timepoint. Individuals with data over multiple timepoints were coded as having the phenotype if they met the threshold/criteria at least once.

Birth complications

In ALSPAC, preterm birth (<37 weeks), low birthweight (<2500g), and Apgar score <7 at 5 minutes were identified through obstetric records, or parent-report where records could not be identified. In MCS, preterm birth (<37 weeks) and low birthweight (<2500g) were defined by parent-report. Preterm birth and low birthweight were defined only in singleton births.

Early motor and language development

In ALSPAC, preschool development items included: communication, fine motor, and gross motor delays. Each item was defined using the corresponding scores from the developmental milestones questions at approximately 18 months, which were residualised against age and the lowest scoring 5% defined as having delay. Measures of preschool development were unavailable in MCS.

Cognition

Cognitive measures in ALSPAC were: low performance cognitive ability, low verbal cognitive and low general cognitive ability. Low cognitive ability was defined as the lowest scoring 5% after age-correcting the scores from the total IQ scale of the Weschler Intelligence Scale for Children (WISC)(31) at age 8. Low performance and low verbal cognitive ability were defined as the lowest scoring 5%, after age-correction, of the WISC performance and verbal IQ scales, respectively. We could not derive measures of cognitive ability in MCS.

Child neurodevelopmental conditions

Detailed information on the definition of each neurodevelopmental condition is provided in the supplementary methods and described briefly below. In ALSPAC, a DSM-IV diagnosis of ADHD was generated via the Development and Wellbeing Assessment (DAWBA)(32), a research diagnostic interview conducted with the parent at ages 7, 10, and 13. Individuals meeting criteria during at least one timepoint were defined as having ADHD. Probable ASD was defined categorically as a score >12 on the parent-reported Social Communication Disorders Checklist (SCDC)(33), at ages 7, 10, and/or 13. Children scoring below the 5th centile on the reading subtest of the WORD (Wechsler Objective Reading Dimensions)(34) were defined as having reading difficulties (35). Similarly, children scoring below the 5th centile of the structural language score of the Children's Communication Checklist (CCC) (36) were defined as having structural language difficulties and those below the 5th centile of the pragmatic scale of the CCC were defined as having pragmatic language difficulties (35). Children were defined as having motor coordination difficulties if they scored below the 5th centile on a composite score from the Movement Assessment Battery for Children (MABC)

(35). Tics were defined as the presence of motor and/or vocal tics occurring more than once per week, and were measured by parent-report at age 13.

ADHD, ASD, and reading difficulties were assessed in MCS. Probable ADHD was defined by the parent-rated Strengths and Difficulties Questionnaire (SDQ)(37) hyperactivity subscale as a score >8 at any of the following timepoints: 5, 7, 11, or 14 years. ASD was defined as a parent-report of a clinician diagnosis between ages 5 and 14 years. Children were defined as having reading difficulties if they scored below the 5th centile for the reading subscale of the British Ability Scales (38) at age 7.

Psychiatric conditions

In ALSPAC, emotional problems were defined using the parent-reported emotional symptoms subscale of the SDQ (37). Individuals scoring >6 at any point from ages 4-17 years were defined as having emotional problems. Conduct problems were defined as scoring >5 at any time point on the parent-reported conduct problems subscale of the SDQ, measured at the same ages as above. In MCS, emotional problems and conduct problems were defined using the parent-rated SDQ subscale cut-points as above, measured at ages 5-17.

Social and educational outcomes

Special educational needs (SEND) in ALSPAC were defined by parent report at ages 8 or 11 that the child had 'ever been recognised as having special educational needs'. Social difficulties were defined as having peer problems – classified as a score >4 on the parent-reported SDQ peer subscale at any point between ages: 4-17 years. In MCS, SEND was

defined by parent-report at ages 7, 11, and 14. Peer problems were defined by SDQ peer problems subscale, as in ALSPAC, at ages 5-17.

Analyses

Primary

Logistic regressions were used to test for association between ND CNV carrier status and each outcome in ALSPAC. CNV status was coded with non-carriers as the reference category (0) and carriers as 1. Replication was sought in MCS where corresponding measures were available, followed by meta-analysis of ALSPAC and MCS using random-effects models from the R package 'meta' (39). False-discovery rate ($p<0.05$) was used to correct for multiple comparisons within each phenotype category.

Infrequent vs ultra-rare CNVs

Whilst all the ND CNVs are rare, some occur more commonly than others, in particular 15q11.2 deletion and duplication and 16p11.2 duplication. To investigate whether effect sizes in the primary analyses were driven by the more common CNVs, we tested for association between these three specific CNVs, which we term 'infrequent' CNVs, and all outcomes, in comparison to non-carriers. We also tested for association between the remaining 'ultra-rare' CNVs and all outcomes, in comparison to non-carriers.

Deletions vs duplications

For some loci, deletions may have a more severe impact than their reciprocal duplication (40), thus we compared all duplication carriers to CNV non-carriers for association with each neurodevelopmental phenotype, and all deletion carriers compared to CNV non-carriers separately.

Continuously measured outcomes

For ease of clinical interpretation and translation, the continuous measures were dichotomized to create binary traits. We repeated the above analyses in ALSPAC and MCS using continuous traits where appropriate. A detailed description of the continuous measures is provided in the supplementary methods.

Excluding ASD, ADHD, and low cognitive ability

We tested ND CNV carrier status for association with each phenotype in our primary sample ALSPAC, after excluding individuals with ASD, ADHD, and low cognitive ability from the sample, to assess whether associations were driven by comorbidity with these phenotypes.

Sex differences

In line with recommended best practice (41), we explored sex differences as a secondary analysis. Logistic regressions were repeated in male-only and female-only subsamples of ALSPAC, and effect sizes and confidence intervals compared. Interactions between each

phenotype and sex were examined to investigate whether effect sizes significantly differed between males and females.

Missing data

To investigate the impact of missing data on our results, we used logistic regressions to assess association between i) missing exposure (i.e. CNV) data and developmental phenotype, and ii) missing outcome (i.e. developmental phenotype) data and ND CNV carrier status. Additionally, we tested for associations between ND CNVs and non-participation in ALSPAC clinic and questionnaire assessments at a variety of ages (42) and MCS non-participation at age 17, as DNA was not collected until age 14 in this cohort.

Results

8721 participants in ALSPAC had genetic data available, of whom 6361 (73%) passed CNV quality control. In MCS, 8117 participants had genetic data, of whom 6710 (83%) passed CNV quality control. 144 (2.3%) people in ALSPAC carried an ND CNV, whilst 151 (2.3%) of people in MCS carried an ND CNV. The frequencies of each ND CNV are displayed in Table S2; we identified carriers of 23 out of the 54 known ND CNVs. The frequencies of demographic and outcome measures for ND CNV carriers compared to non-carriers are displayed in Table 1. ND CNV carrier status was not associated with female sex in either sample (ALSPAC: Odds ratio=0.91 [95% C=0.65-1.27] p=0.58; MCS: OR= 1.04 [0.75-1.44], p=0.83).

The percentage of people with an ND CNV by number of neurodevelopmental conditions and the percentage of neurodevelopmental conditions by ND CNV carrier status are shown in Figures S1 and S2.

Primary analysis

Full results of the primary, replication, and meta-analysis are presented in Table 2 and Figures 1-3.

After correcting for multiple comparisons in ALSPAC, carrying an ND CNV was significantly associated with greater likelihood of preterm birth, gross and fine motor delay, and low performance, verbal, and general cognitive ability. ND CNVs were also associated with ADHD, ASD, reading difficulties, motor coordination difficulties, and pragmatic and structural language difficulties, but not with tics. ND CNV carrier status was not associated with child emotional or conduct problems. ND CNVs were associated with SEND but not with peer problems (Figure 1).

Consistent with findings in ALSPAC, carrying an ND CNV in MCS was associated with ADHD, reading difficulties, and SEND, and was not associated with low birthweight or emotional problems. In contrast to ALSPAC, ND CNVs in the MCS cohort were not associated with preterm birth or ASD (parent reported diagnosis), but were associated with conduct and peer problems (Table 2 and Figure 1).

In the meta-analysis, ND CNVs were associated with low birthweight, ADHD, ASD, reading difficulties, SEND, and peer problems (Table 2 and Figure 2). We did not find evidence of

association between ND CNVs and preterm birth, conduct or emotional problems in the meta-analysis.

'Infrequent' vs 'ultra-rare' CNVs

In ALSPAC, effect sizes were generally stronger for the 'ultra-rare' CNVs in comparison to the relatively 'infrequent' CNVs - i.e., 15q11.2 deletion or duplication, or 16p11.2 duplication (Table S3 and Figure S3). These differences were particularly pronounced for low cognitive ability, reading difficulties, coordination difficulties, and SEND. In MCS, effect sizes between 'infrequent' and 'ultra-rare' CNV carriers were more comparable, with the exception of peer problems which had a noticeably stronger association with ultra-rare CNV carriers (Table S4 and Figure S3).

Deletions and duplications

Effect sizes were generally consistent between deletions and duplications, with a few exceptions. In ALSPAC, stronger effects were observed between deletions and pragmatic and structural language difficulties and SEND, than were seen with duplications and these items (Table S5 and Figure S4). In MCS, stronger associations were observed between deletions and conduct problems and SEND than for duplications and these items (Table S6 and Figure S4).

Continuously measured outcomes

In ALSPAC, when defining outcomes on a continuous scale, ND CNVs were associated with lower gross and fine motor skills, lower communication skills, lower performance, verbal, and general cognitive ability, and lower reading, coordination, and pragmatic and structural language abilities (Table S7). In MCS, when using continuous outcomes, ND CNVs were associated with lower birthweight, higher ADHD traits, and poorer reading ability (Table S8). These findings were consistent with those observed using our primary outcomes, which were defined as binary items.

Results of analyses investigating 'infrequent' compared to 'ultra-rare' CNVs and deletions compared to duplications against continuously-measured outcomes were consistent with the findings using binary outcomes (Tables S9-S12 and Figures S5-S6).

Excluding ASD, ADHD, and low cognitive ability

Patterns of association were consistent when excluding individuals with ADHD, ASD, and low cognitive ability (Figure S7 and Table S13).

Sex differences

Sex differences were examined as a secondary analysis and results presented in Figures S8-S9 and Table S14. A significant interaction with sex was observed only for low cognitive ability in the direction of low cognitive ability being associated with ND CNVs in males but not females.

Missing data

Carrying an ND CNV was associated with ALSPAC clinic non-attendance only at age 7 (OR=1.62 [1.14-2.31], p=0.007). Full missingness results are displayed in Figure S10 and Tables S15-S16.

Discussion

Using two UK population birth cohorts, we found that children carrying a known neurodevelopmental CNV are at a higher chance of birth complications, lower cognitive ability (tested in ALSPAC only), neurodevelopmental conditions, and special educational needs. In ALSPAC, which contained a wider range of neurodevelopmental measures, all broadly defined DSM-5 neurodevelopmental conditions, except for tics, were associated with carrying an ND CNV. Our study provides an extensive population-wide assessment of the early impacts of CNVs in childhood and highlights a broad range of developmental indicators that may be useful for clinicians to evaluate when suspecting a CNV or considering genetic testing.

It is well established that ND CNVs are enriched in those who are autistic, or have ADHD or intellectual disability (43,44). To date there have been no population-based studies of ND CNVs that have focused on other DSM-5 child neurodevelopmental conditions including specific learning difficulties (e.g. reading difficulties) and speech and language or communication difficulties, and only a limited number of studies investigating developmental motor coordination disorder and difficulties in ND CNV carriers (45,46).

However, one volunteer population cohort of children, examining CNVs defined differently, suggested association with learning problems as well as ASD and ADHD (15). We observed ND CNV effect sizes in ALSPAC for communication, motor coordination, and reading difficulties (OR range 2.81-3.40) that were similar to those for ADHD, ASD, and low cognitive ability (OR range 1.77-4.00), and persisted even after excluding individuals with the latter three conditions. Yet little research has focused on the genetic underpinnings of these neurodevelopmental difficulties (35). In many countries, reading, coordination and language difficulties are typically assessed and managed by therapists, psychologists, and other specialists in educational or other contexts where the possibility of examining genetic contributions, including by genetic screening, may not be considered or available. Our findings suggest that clinicians working with children and young people with a recognised CNV should consider screening for a broad range of neurodevelopmental difficulties, beyond ADHD, ASD, and cognitive difficulties.

Whilst all the CNVs included in our study are rare, with a population frequency <1%, some CNVs are relatively more common, namely 15q11.2 deletion, 15q11.2 duplication, and 16p11.2 duplication. Sensitivity analyses comparing these three relatively more common CNVs, i.e. 'infrequent' CNVs, to non-carriers, and the remaining 'ultra-rare' CNVs to non-carriers suggested that the ultra-rare CNVs had stronger effects on low cognitive ability, reading difficulties, coordination difficulties, and SEND in ALSPAC, and peer problems in MCS. These findings also suggest that associations in the primary analysis were not driven by the more prevalent CNVs. Similarly, deletions were associated with greater effect sizes for language difficulties in ALSPAC, conduct problems in MCS, and SEND in both cohorts. Our findings are consistent with previous research identifying stronger associations with

neuropsychiatric conditions in deletions compared to their reciprocal duplications (14,40), although others have not observed this effect (19).

ASD and preterm birth were significantly associated with ND CNVs in ALSPAC and not MCS, whilst conduct and peer problems were significantly associated with ND CNVs in MCS and not ALSPAC. These disparities may in part be due to differences in measurement and power. For instance, ASD was measured using the SCDC questionnaire in ALSPAC, capturing a broader phenotype, whilst in MCS parents were asked if their child had ever been diagnosed with ASD, to which only a small number of parents responded 'yes'. These differences in power are reflected in the wider confidence intervals of the estimate in MCS compared to ALSPAC, despite having similar effect sizes. Other differences may reflect variation in the sociodemographic make up of each cohort and/or secular trends in these phenotypes (47).

The meta-analysis of two large birth cohorts afforded us greater power to detect associations between CNV carrier status and several of the phenotypes, allowing us to overcome some limitations of study design in the literature. However, only ALSPAC included the full range of neurodevelopmental assessments, and some conditions (e.g. ASD, tics) were not assessed as rigorously as they are in clinical studies. Nevertheless, previous research has focused predominantly on either clinically-ascertained childhood samples of those who carry a specific CNV, those with ADHD, ASD or developmental delay, or adult-based population cohorts. Clinical samples, while invaluable for studying the clinical presentation of specific CNVs, are typically biased towards selecting individuals with more overt physical phenotypes (e.g. cleft lip), more severe neurodevelopmental conditions (e.g. intellectual disability), males, and specific socio-economic and ethnic groups. Therefore,

clinical studies may over-estimate prevalence rates of neurodevelopmental conditions in CNV carriers. This was suggested by a large Danish study that examined associations between recurrent CNVs and ADHD, ASD and schizophrenia (19). Similarly, we found that in ALSPAC, carrying an ND CNV increased the risk of ASD by an odds ratio of 1.77 (CI=1.06-2.95), whereas research in a UK clinical sample, which used similar questionnaire methods to define ASD, reported an odds ratio of 44.1 (15.3-127.5) (18). Adult population samples overcome certain biases but some, e.g. UK Biobank, are subject to strong ascertainment bias (e.g. higher socioeconomic status) and all adult studies are subject to survivor bias. Only a few studies have investigated non-clinical childhood samples (15–17) with limited investigation of known ND loci (17). By utilizing birth cohorts, we were able to minimise the impact of survivor bias in adult samples, without the selective recruitment bias seen in clinical or volunteer samples. Nevertheless, ALSPAC is over-representative of individuals from higher socio-economic classes and genetic data are available only in individuals of European genetic ancestry. MCS is more representative of the UK population due to targeted recruitment of individuals from ethnic minority and economically-deprived backgrounds, and thus provides a less biased estimate of effect size. However, fewer neurodevelopmental and psychiatric measures were available in MCS than in ALSPAC. Additionally, MCS did not collect DNA until age 14, and individuals with ND CNVs may have dropped out of the sample by this age, although we note that in ALSPAC ND CNVs did not appear to strongly predict attrition beyond age 7 clinic attendance.

We observed associations between various developmental phenotypes and missing genetic data, and between CNV carrier status and missing data for coordination and reading

difficulties. Complete case analysis may bias estimates when missingness is associated with both the outcome and exposure (48). This was the case for coordination difficulties in ALSPAC, and reading difficulties in MCS, suggesting that the association we observed for these phenotypes may be attenuated. Additionally, in MCS not having emotional problems and not having peer problems were associated with missing genetic data. For these results, non-random missingness may have led to inflated effect sizes.

Another limitation is that associations with clinical and social/educational phenotypes likely vary by the specific CNV. Due to limitations of sample size, in the current study we pooled all ND CNVs into a combined measure to analyse as a group. Whilst we conducted sensitivity analyses comparing deletions and duplications, and relatively more common compared to ultra-rare CNVs, we did not have the power to examine the effect of individual CNVs, nor the ability to determine whether CNVs were de novo or inherited. Thus, the findings of our study should be considered alongside clinical studies of specific CNVs. Furthermore, we defined neurodevelopmental CNVs on the basis of evidence from the literature (1) and the DECIPHER database (https://www.deciphergenomics.org/disorders/syndromes/list), and also to maintain consistency with other studies (10,30). It is possible that other loci are associated with neurodevelopmental conditions, and the strength of evidence will increase with higher powered studies. Future research should aim to integrate new loci into our list of neurodevelopmental CNVs, using other databases such as ClinVar to evaluate the robustness of the evidence.

Overall, our findings highlight widespread early impacts of ND CNVs on development that encompass nearly all DSM-5 neurodevelopmental conditions. They suggest it may be helpful

to evaluate a range of developmental indicators when considering genetic testing. Our results further highlight the need for genetics training to be extended to other types of professional groups who may work with children with CNVs, such as speech and language therapists, physiotherapists, and psychologists.

Tables

Table 1. Frequency of demographic and outcome measures by ND CNV carrier status in ALSPAC and MCS. Percentages are calculated vertically to indicate the percentage of people with/without a CNV who endorse each phenotype. ADHD – attention deficit hyperactivity disorder, ASD – autism spectrum disorder, SEND – special educational needs. *Cell counts less than 5, and corresponding percentages, are not specified to maintain anonymity as required by ethics guidelines.

Category	Phenotype	Group	ALSPAC		MCS	
			Non-carriers (n=6217)	**ND CNV carriers** (n=144)	**Non-carriers** (n=6559)	**ND CNV carriers** (n=151)
Demographics	Sex	Males	3000 (48.3%)	73 (50.7%)	3244 (51.2%)	73 (50.3%)
		Females	3206 (51.7%)	71 (49.3%)	3088 (48.8%)	72 (49.7%)
	Genetic ancestry	European	6217 (100%)	5057 (86%)	5057 (86%)	124*
		South Asian	NA		667 (11.3%)	11*
		African			154 (2.6%)	<5*
Birth complications	Preterm birth	No	3289 (93.9%)	76 (87.4%)	5735 (93.7%)	137 (95.8%)
		Yes	213 (6.1%)	11 (12.6%)	386 (6.3%)	6 (4.2%)
	Low birthweight	No	3289 (95.3%)	79 (90.8%)	5805 (94.1%)	131 (90.3%)
		Yes	161 (4.7%)	8 (9.2%)	364 (5.9%)	14 (9.7%)
	Apgar score <7	No	3456 (98.8%)	85*	NA	
		Yes	43 (1.2%)	<5*		
Early motor and language development	Gross motor delay	No	4929 (95.4%)	101 (89.4%)	NA	
		Yes	236 (4.6%)	12 (10.6%)		
	Fine motor delay	No	4991 (96.6%)	101 (89.4%)	NA	
		Yes	174 (3.4%)	12 (10.6%)		
	Communication delay	No	4996 (96.6%)	106 (93.8%)	NA	
		Yes	174 (3.4%)	7 (6.2%)		
Cognition	Low performance cognitive ability	No	4096 (95.4%)	80 (87%)	NA	
		Yes	196 (4.6%)	12 (13%)		
	Low verbal cognitive ability	No	4113 (95.8%)	79 (84.9%)	NA	
		Yes	182 (4.2%)	14 (15.1%)		
	Low cognitive ability	No	4089 (95.5%)	81 (88%)	NA	
		Yes	191 (4.5%)	11 (12%)		
Neuro-developmental conditions	ADHD	No	4962 (97%)	109 (92.4%)	6094 (96.9%)	138 (93.9%)
		Yes	154 (3%)	9 (7.6%)	192 (3.1%)	9 (6.1%)
	ASD	No	4618 (90.7%)	99 (84.6%)	6063 (96.5%)	140 (95.2%)
		Yes	474 (9.3%)	18 (15.4%)	223 (3.5%)	7 (4.8%)
	Reading difficulties	No	4537 (95.9%)	86 (88.7%)	5678 (95.8%)	118 (91.5%)
		Yes	193 (4.1%)	11 (11.3%)	250 (4.2%)	11 (8.5%)
	Motor coordination difficulties	No	3559 (95.8%)	65 (89%)	NA	
		Yes	156 (4.2%)	8 (11%)		
		No	4149 (96%)	89 (89%)	NA	

Category	Phenotype	Group	ALSPAC		MCS	
			Non-carriers (n=6217)	**ND CNV carriers** (n=144)	**Non-carriers** (n=6559)	**ND CNV carriers** (n=151)
	Pragmatic language difficulties	Yes	174 (4%)	11 (11%)		
	Structural language difficulties	No	4250 (96.5%)	90 (89.1%)	NA	
		Yes	153 (3.5%)	11 (10.9%)		
	Tics	No	3167 (82.9%)	75 (89.3%)	NA	
		Yes	654 (17.1%)	9 (10.7%)		
Psychiatric conditions	Emotional problems	No	5222 (94.4%)	120 (93.8%)	5403 (82.4%)	121 (80.1%)
		Yes	312 (5.6%)	8 (6.3%)	1154 (17.6%)	30 (19.9%)
	Conduct problems	No	5228 (94.5%)	122 (95.3%)	5783 (88.2%)	118 (78.1%)
		Yes	307 (5.5%)	6 (4.7%)	772 (11.8%)	33 (21.9%)
Social and educational outcomes	SEND	No	3210 (74.2%)	53 (52%)	5374 (87.8%)	117 (81.3%)
		Yes	1115 (25.8%)	49 (48%)	747 (12.2%)	27 (18.8%)
	Peer problems	No	4873 (90.3%)	113 (89%)	5452 (83.2%)	114 (75.5%)
		Yes	523 (9.7%)	14 (11%)	1103 (16.8%)	37 (24.5%)

Table 2 Results of the analysis in ND CNV carriers vs non-carriers in ALSPAC, MCS, and the meta-analysis. P-values are corrected for multiple comparisons using false-discovery rate.

Category	Phenotype	ALSPAC				MCS				Meta-analysis			
		OR	Lower CI	Upper CI	P-value	OR	Lower CI	Upper CI	P-value	OR	Lower CI	Upper CI	P-value
Birth complications	Preterm birth	2.23	1.17	4.27	0.04	0.65	0.29	1.48	0.31	1.24	0.37	4.15	0.73
	Low birthweight	2.07	0.98	4.35	0.08	1.70	0.97	2.99	0.13	1.83	1.17	2.86	0.02
	Apgar score <7	0.95	0.13	6.95	0.96	NA				NA			
Early development	Gross motor delay	2.48	1.34	4.58	0.01	NA				NA			
	Fine motor delay	3.41	1.84	6.32	3.0E-04	NA				NA			
	Communication delay	1.90	0.87	4.13	0.11	NA				NA			
Cognition	Low performance cognitive ability	3.13	1.68	5.85	4.9E-04	NA				NA			
	Low verbal cognitive ability	4.00	2.23	7.21	1.1E-05	NA				NA			
	Low cognitive ability	2.91	1.52	5.55	1.2E-03	NA				NA			
Neuro-developmental conditions	ADHD	2.66	1.32	5.35	0.01	2.09	1.33	3.28	4.4E-03	2.24	1.53	3.28	9.1E-05
	ASD	1.77	1.06	2.95	0.03	1.36	0.63	2.94	0.43	1.63	1.07	2.50	0.02
	Reading difficulties	3.01	1.58	5.72	2.4E-03	2.15	1.20	3.85	0.03	2.51	1.60	3.94	9.1E-05
	Coordination difficulties	2.81	1.32	5.95	0.01	NA				NA			
	Pragmatic language difficulties	2.95	1.55	5.61	2.4E-03	NA				NA			
	Structural language difficulties	3.40	1.78	6.48	1.5E-03	NA				NA			

Category	Phenotype	ALSPAC				MCS				Meta-analysis			
		OR	Lower CI	Upper CI	P-value	OR	Lower CI	Upper CI	P-value	OR	Lower CI	Upper CI	P-value
	Tics	0.58	0.29	1.17	0.13	NA				NA			
Psychiatric conditions	Emotional problems	1.12	0.54	2.30	0.77	1.16	0.78	1.74	0.47	1.15	0.81	1.64	0.44
	Conduct problems	0.84	0.37	1.92	0.77	2.09	1.41	3.10	4.6E-04	1.43	0.59	3.46	0.44
Social and educational outcomes	SEND	2.66	1.79	3.95	2.3E-06	1.66	1.08	2.54	0.02	2.12	1.33	3.36	3.0E-03
	Peer problems	1.15	0.66	2.03	0.62	1.60	1.10	2.34	0.02	1.45	1.06	1.98	0.02

Figure legends

Figure 1. Results of primary analyses in ALSPAC and MCS.

Odds ratio and 95% confidence intervals of the association between each outcome and ND CNV carrier status in A) ALSPAC and B) MCS. Asterisks indicate FDR-corrected $p<0.05$. ADHD – attention deficit hyperactivity disorder, ASD – autism spectrum disorder, SEND – special educational needs.

Figure 2. Results of the meta-analysis of ALSPAC and MCS.

Odds ratio and 95% confidence intervals of the association between each phenotype and ND CNV carrier status. Asterisks indicate FDR-corrected $p<0.05$. ADHD – attention deficit hyperactivity disorder, ASD – autism spectrum disorder, SEND – special educational needs.

Acknowledgements

We are extremely 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.

CD, AS, LR, VP, and AT are supported by funding from the Wolfson Foundation. JM is funded by the Welsh Government through Health and Care Research Wales via an NIHR Advanced Fellowship (NIHR-FS(A)-2022). MJO was supported by a Medical Research Council Centre grant MR/L010305/1 and programme grant MR/P005748/1.

The UK Medical Research Council and Wellcome (Grant ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. For ALSPAC, genome-wide genotyping data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf)

This publication is the work of the authors and CD and AT will serve as guarantors for the contents of this paper.

Disclosures

MJO reported receiving grants from Akrivia Health outside the submitted work. MJO reported receiving grants from Takeda Pharmaceutical Company Ltd outside the submitted work. Takeda and Akrivia played no part in the conception, design, implementation, or interpretation of this study. All other authors report no biomedical financial interests or potential conflicts of interest.

References

1. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, *et al.* (2014): Refining analyses of copy number variation identifies specific genes associated with developmental delay [no. 10]. *Nat Genet* 46: 1063–1071.

2. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, *et al.* (2011): A copy number variation morbidity map of developmental delay. *Nat Genet* 43: 838–846.

3. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, *et al.* (2010): Functional impact of global rare copy number variation in autism spectrum disorders [no. 7304]. *Nature* 466: 368–372.

4. Vicari S, Napoli E, Cordeddu V, Menghini D, Alesi V, Loddo S, *et al.* (2019): Copy number variants in autism spectrum disorders. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 92: 421–427.

5. Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, *et al.* (2010): Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *The Lancet* 376: 1401–1408.

6. Gudmundsson OO, Walters GB, Ingason A, Johansson S, Zayats T, Athanasiu L, *et al.* (2019): Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Transl Psychiatry* 9: 1–9.

7. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, *et al.* (2017): Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 49: 27–35.

8. Rees E, Walters JTR, Georgieva L, Isles AR, Chambert KD, Richards AL, *et al.* (2014): Analysis of copy number variations at 15 schizophrenia-associated loci. *The British Journal of Psychiatry* 204: 108–114.

9. Crawford K, Bracher-Smith M, Owen D, Kendall KM, Rees E, Pardiñas AF, *et al.* (2019): Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *Journal of Medical Genetics* 56: 131–138.

10. Birnbaum R, Mahjani B, Loos RJF, Sharp AJ (2022): Clinical Characterization of Copy Number Variants Associated With Neurodevelopmental Disorders in a Large-scale Multiancestry Biobank. *JAMA Psychiatry* 79: 250–259.

11. Martin CL, Wain KE, Oetjens MT, Tolwinski K, Palen E, Hare-Harris A, *et al.* (2020): Identification of Neuropsychiatric Copy Number Variants in a Health Care System Population. *JAMA Psychiatry* 77: 1276–1285.

12. Royal College of Psychiatrists (2023): College Report CR237 - The role of genetic testing in mental health settings. The Royal College of Psychiatrists. Retrieved January 11, 2023, from https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/College-report-CR237---Genetic-testing-in-mental-health-settings.pdf

13. Chawner S, Watson CJ, Owen MJ (2021): Clinical evaluation of patients with a neuropsychiatric risk copy number variant. *Current Opinion in Genetics & Development* 68: 26–34.

14. Niarchou M, Chawner SJRA, Doherty JL, Maillard AM, Jacquemont S, Chung WK, *et al.* (2019): Psychiatric disorders in children with 16p11.2 deletion and duplication. *Translational Psychiatry* 9: 8.

15. Zarrei M, Burton CL, Engchuan W, Higginbotham EJ, Wei J, Shaikh S, *et al.* (2023): Gene copy number variation and pediatric mental health/neurodevelopment in a general population. *Human Molecular Genetics* ddad074.

16. Martin J, Tammimies K, Karlsson R, Lu Y, Larsson H, Lichtenstein P, Magnusson PKE (2019): Copy number variation and neuropsychiatric problems in females and males in the general population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 180: 341–350.

17. Guyatt AL, Stergiakouli E, Martin J, Walters J, O'Donovan M, Owen M, *et al.* (2018): Association of copy number variation across the genome with neuropsychiatric traits in the general population. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 177: 489–502.

18. Chawner S, Owen MJ, Holmans P, Raymond FL, Skuse D, Hall J, Bree MBM van den (2019): Genotype–phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. *The Lancet Psychiatry* 6: 493–505.

19. Vaez M, Montalbano S, Calle Sánchez X, Georgii Hellberg K-L, Dehkordi SR, Krebs MD, *et al.* (2024): Population-Based Risk of Psychiatric Disorders Associated With Recurrent Copy Number Variants. *JAMA Psychiatry*. https://doi.org/10.1001/jamapsychiatry.2024.1453

20. Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S, *et al.* (2019): Association of Rare Copy Number Variants With Risk of Depression. *JAMA Psychiatry* 76: 818–825.

21. Calle Sánchez X, Helenius D, Bybjerg-Grauholm J, Pedersen C, Hougaard DM, Børglum AD, *et al.* (2022): Comparing Copy Number Variations in a Danish Case Cohort of Individuals With Psychiatric Disorders. *JAMA Psychiatry* 79: 59–69.

22. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, *et al.* (2013): Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology* 42: 111–127.

23. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, *et al.* (2013): Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol* 42: 97–110.

24. Connelly R, Platt L (2014): Cohort Profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology* 43: 1719–1725.

25. Pain O, Al-Chalabi A, Lewis CM (2024, June 13): The GenoPred Pipeline: A Comprehensive and Scalable Pipeline for Polygenic Scoring. medRxiv, p 2024.06.12.24308843.

26. Dennison CA, Martin J, Shakeshaft A, Riglin L, Rice F, Lewis CM, *et al.* (2023): Stratifying early-onset emotional disorders: using genetics to assess persistence in young people of European and South Asian ancestry. *Journal of Child Psychology and Psychiatry* 65: 42–51.

27. Dinneen TJ, Ghrálaigh FN, Walsh R, Lopez LM, Gallagher L (2022): How does genetic variation modify ND-CNV phenotypes? *Trends in Genetics* 38: 140–151.

28. Dittwald P, Gambin T, Szafranski P, Li J, Amato S, Divon MY, *et al.* (2013): NAHR-mediated copy-number variants in a clinical population: Mechanistic insights into both genomic disorders and Mendelizing traits. *Genome Res* 23: 1395–1409.

29. Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V, *et al.* (2019): Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *The British Journal of Psychiatry* 214: 297–304.

30. Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J, *et al.* (2017): Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biological Psychiatry* 82: 103–110.

31. Wechsler D (1991): *The Wechsler Intelligence Scale for Children-Third Edition (WISC-III)*. San Antonio, Texas: The Psychological Corporation.

32. Goodman R, Ford T, Richards H, Gatward R, Meltzer H (2000): The Development and Well-Being Assessment: Description and Initial Validation of an Integrated Assessment of Child and Adolescent Psychopathology. *Journal of Child Psychology and Psychiatry* 41: 645–655.

33. Skuse DH, Mandy WPL, Scourfield J (2005): Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *The British Journal of Psychiatry* 187: 568–572.

34. Rust J, Golombok S, Trickey G (1993): WORD, Wechsler objective reading dimensions manual. Psychological Corporation.

35. Eyre O, Hughes RA, Thapar AK, Leibenluft E, Stringaris A, Davey Smith G, *et al.* (2019): Childhood neurodevelopmental difficulties and risk of adolescent depression: the role of irritability. *Journal of Child Psychology and Psychiatry* 60: 866–874.

36. Bishop D (1998): Development of the Children's Communication Checklist (CCC): A Method for Assessing Qualitative Aspects of Communicative Impairment in Children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 39: 879–891.

37. Goodman R (1997): The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry* 38: 581–586.

38. Elliott CD, Smith P, McCulloch K (1996): *British Ability Scales II*. Windsor, Berkshire: NFER-Nelson Publishing Company.

39. Balduzzi S, Rücker G, Schwarzer G (2019): How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health* 22: 153–160.

40. Lin A, Vajdi A, Kushan-Wells L, Helleman G, Hansen LP, Jonas RK, *et al.* (2020): Reciprocal Copy Number Variations at 22q11.2 Produce Distinct and Convergent Neurobehavioral Impairments Relevant for Schizophrenia and Autism Spectrum Disorder. *Biological Psychiatry* 88: 260–272.

41. Khramtsova EA, Wilson MA, Martin J, Winham SJ, He KY, Davis LK, Stranger BE (2023): Quality control and analytic best practices for testing genetic models of sex differences in large populations. *Cell* 186: 2044–2061.

42. Martin J, Tilling K, Hubbard L, Stergiakouli E, Thapar A, Davey Smith G, *et al.* (2016): Association of Genetic Risk for Schizophrenia With Nonparticipation Over Time in a Population-Based Cohort Study. *American Journal of Epidemiology* 183: 1149–1158.

43. Mollon J, Almasy L, Jacquemont S, Glahn DC (2023): The contribution of copy number variants to psychiatric symptoms and cognitive ability. *Mol Psychiatry* 28: 1480–1493.

44. Rees E, Kirov G (2021): Copy number variation and neuropsychiatric illness. *Current Opinion in Genetics & Development* 68: 57–63.

45. Cunningham AC, Delport S, Cumines W, Busse M, Linden DEJ, Hall J, *et al.* (2018): Developmental coordination disorder, psychopathology and IQ in 22q11.2 deletion syndrome. *The British Journal of Psychiatry* 212: 27–33.

46. Cunningham AC, Hall J, Owen MJ, Bree MBM van den (2021): Coordination difficulties, IQ and psychopathology in children with high-risk copy number variants. *Psychological Medicine* 51: 290–299.

47. Esposti MD, Matijasevich A, Collishaw S, Martins-Silva T, Santos IS, Menezes AMB, *et al.* (2023): Secular trends and social inequalities in child behavioural problems across

three Brazilian cohort studies (1993, 2004 and 2015). *Epidemiology and Psychiatric Sciences* 32: e23.

48. Hughes RA, Heron J, Sterne JAC, Tilling K (2019): Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology* 48: 1294–1304.