The phenotypic expression of neuropsychiatric copy number variants

Dr Kimberley Marie Kendall

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy.

MRC Centre for Neuropsychiatric Genetics and Genomics
Division of Psychiatry and Clinical Neurosciences
School of Medicine
Cardiff University
2021

Summary

Large, rare copy number variants (CNVs) are associated with neurodevelopmental disorders but there is limited evidence on their phenotypic effects in individuals who do not develop these phenotypes. The aim of this thesis was to establish the effects of 53 neurodevelopmental disorder associated CNVs on individuals with and without psychiatric disorders, in psychiatric, cognitive and physical health domains.

In analyses of individuals without neurodevelopmental disorders in UK Biobank, carriers of 53 neurodevelopmental CNVs performed more poorly than CNV noncarriers across seven cognitive tasks and related functional outcome measures. In the same group of individuals, neurodevelopmental CNVs were associated with an increased risk of depression.

I assessed 5 carriers of neurodevelopmental CNVs and 22 CNV noncarriers with psychiatric disorders from the Cardiff Cognition in Schizophrenia Study (Cardiff COGS) and National Centre for Mental Health (NCMH) cohorts. I also analysed previously collected data from 76 CNV carriers and 2,389 CNV noncarriers from NCMH. CNV carriers tended to have primary diagnoses and family histories from the neurodevelopmental spectrum, and had a greater number of both psychiatric and physical health diagnoses than CNV noncarriers. CNV carriers also displayed greater deficits in assessments of negative symptoms and tended to be more likely to report delays in walking and talking.

The findings of this PhD add to the evidence on the phenotypic spectrum of neurodevelopmental CNVs, generate hypotheses for future work investigating the role of CNVs in psychiatric disorders and provide a starting point for the translation of these findings to the clinic.

Acknowledgements

I would first like to thank my primary supervisor James Walters. His advice and mentorship played an integral role in my development and completion of this project and in securing funding. I would also like to thank George Kirov, who taught me everything I know about CNVs and Ian Jones, for his unfailing support and advice. My thanks also go to Mike Owen for his mentorship.

Thank you to the Wales Clinical Academic Track and the Wellcome Trust, and to the participants who gave up their time to take part in this research. It would not have been possible without their valuable contribution.

I would like to thank and acknowledge my colleagues at Cardiff University for their contributions to the research presented in this thesis. I would also like to acknowledge the work of the field teams of the Cardiff COGS and NCMH cohorts who recruited participants and collected data included in parts of this thesis. Thank you to Emma Chubb who assisted with data collection for the recall study in Chapter 4.

Thank you to Professor Anne Bassett and her team for hosting me at the Dalglish Family 22q Clinic, Toronto General Hospital and providing a truly inspiring learning experience.

Thank you to my family and friends for their support. I would particularly like to thank my husband Dan and daughter Gwenan for their love and support throughout my PhD.

This thesis is dedicated to Gaynor Kendall.

Contributions

The work in this thesis was made possible through the support of the Wellcome Trust in the form of a clinical research training fellowship. The ideas for the fellowship and PhD arose from my clinical interests and following discussion with James Walters. We then approached George Kirov and Ian Jones who were enthusiastic in their support of the work.

At the start of my fellowship, George Kirov taught me how to call CNVs and we set about doing this in the first wave of genotyped UK Biobank participants. At the time, this was the largest sample with both genetic and cognitive data. I took the opportunity to learn about the cognitive tasks used by UK Biobank and analysed their results for association with 53 neurodevelopmental CNVs. I later went on to repeat the analyses in the full sample for inclusion in this thesis (Chapter 2).

I had the idea to examine the role of neurodevelopmental CNVs in depression risk because I became aware that the existing literature was inconclusive and underpowered. I began this work in ~150,000 participants just as UK Biobank released genotyping data for the remaining samples. I had the option of continuing the depression analyses or pausing them to complete CNV calling with George Kirov and Matthew Bracher-Smith and repeating them later in the full sample. I chose the latter option. As a result, the analyses were much better powered, allowing me to examine associations with individual CNV loci (Chapter 3). This work took me up to my maternity leave which started in February 2018.

Throughout the initial years of the fellowship, I had developed a protocol for the phenotypic assessment of individuals from pre-existing psychiatric cohorts. In addition, I carried out cognitive assessments for 29 individuals without psychiatric disorders for inclusion in a reference sample. On my return from maternity leave in September 2018, I set about recalling CNV carrying individuals and matched CNV noncarriers for assessment. This went well, and with the help of a psychology placement student Emma Chubb, I

assessed 27 individuals (5 CNV carriers, 22 CNV noncarriers). Unfortunately, in March 2020 the COVID19 pandemic and lockdown forced us to end the study early. Recognising that this adversely affected power for my analyses, I have also analysed existing NCMH data, where possible (Chapter 4).

I carried out all analyses and writing in this thesis. Chapters 2 and 3 are based on papers, the drafts of which were commented on by my co-authors. My supervisors James Walters, George Kirov and Ian Jones provided comments on this thesis. Contributions from others are described below.

Work	Chapter	People
CNV calling	2, 3	Professor George Kirov, Matthew Bracher-Smith
Generating accessory files for CNV calling; Perl annotation script	2, 3	Dr Elliott Rees
-nopower2 command use in CNV calling	2, 3	Dr Kai Wang
Defining European genetic ancestry	2, 3	Dr Sophie Legge
CNV calling in the Cardiff COGS and NCMH samples	4	Dr Elliott Rees, Dr Leon Hubbard

Publications based on this thesis

Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J, O'Donovan MC, Owen MJ, Walters KTR, Kirov G. 2017. Cognitive performance among carriers of pathogenic copy number variants: analysis of 152,000 UK Biobank subjects. *Biological Psychiatry* 82(2), 103 - 110.

Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V, Owen MJ, O'Donovan MC, Walters JTR, Kirov G. 2019. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *British Journal of Psychiatry* 214(5), 297 - 304.

Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S, O'Donovan MC, Owen MJ, Jones I, Kirov G, Walters JTR. 2019. Association of rare copy number variants with risk of depression. *JAMA Psychiatry* 76(8), 818 - 825.

Contributions to papers relevant to this thesis

Kendall KM, John A, Sze Chim L, Rees E, Pardiñas A, Del Pozo Banos M, Owen MJ, O'Donovan MC, Kirov G, Lloyd K, Jones I, Legge SE, Walters JTR. 2020. Impact of schizophrenia genetic liability on the association between schizophrenia and physical illness: a data linkage study. *BJPsych Open* 6(6), E139. doi: 10.1192/bjo.2020.42.

Warland A, **Kendall KM**, Rees E, Kirov G, Caseras X. 2019. Schizophrenia-associated genomic copy number variants and subcortical brain volumes in UK Biobank. *Molecular Psychiatry* doi: 10.1038/s41380-019-0355-y.

Silva AI, Kirov G, **Kendall KM**, Bracher-Smith M, Wilkinson LS, Hall J, Ulfarsson MO, Walters GB, Stefansson H, Stefansson K, Linden DEJ, Caseras X. 2021. Analysis of diffusion tensor imaging data from the UK Biobank confirms dosage effect of 15q11.2 copy number variation on white

matter and shows association with cognition. *Biological Psychiatry* doi: 10.1016/j.biopsych.2021.02.969.

Brcic L, Underwood JFG, **Kendall KM**, Caseras X, Kirov G, Davies W. 2020. Medical and neurobehavioural phenotypes in carriers of X-linked ichthyosis-associated genetic deletions in the UK Biobank. *Journal of Medical Genetics* doi: 10.1136/jmedgenet-2019-106676.

Legge SE, Jones HJ, **Kendall KM**, Pardiñas A, Menzies G, Bracher-Smith M, Escott-Price V, Rees E, Davis KAS, Hotopf M, Savage JE, Posthuma D, Holmans P, Kirov G, Owen MJ, O'Donovan MC, Zammit S, Walters JTR. 2019. Association of genetic liability to psychotic experiences with neuropsychiatric disorders and traits. *JAMA Psychiatry* doi:10.1001/jamapsychiatry.2019.2508.

Crawford K, Bracher-Smith M, **Kendall KM**, Rees E, Padiñas AF, Einon M, Escott-Price V, Walters JTR, O'Donovan MC, Owen MJ, Kirov G. 2019. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *Journal of Medical Genetics* doi: 10.1136/jmedgenet-2018-105477.

Owen D, Bracher-Smith M, **Kendall KM**, Rees E, Einon M, Escott-Price V, Owen MJ, O'Donovan MC, Kirov G. 2018. Effects of pathogenic CNVs on physical traits in participants of the UK Biobank. *BMC Psychiatry* 19(1), 867, doi: 10.1186/s12864-018-5292-7.

Escott-Price V, Smith DJ, **Kendall KM**, Ward J, Kirov G, Owen MJ, Walters JTR, O'Donovan MC. 2018. Polygenic risk for schizophrenia and season of birth within the UK Biobank cohort. *Psychological Medicine* doi: 10.1017/S0033291718000454.

Gubb S, Brcic L, Underwood JFW, **Kendall KM**, Caseras X, Kirov G, Davies W. 2020. Medical and neurobehavioural phenotypes in male and female carriers of Xp22.31 duplications in the UK Biobank. *Human Molecular Genetics* doi: 10.1093/hmg/ddaa174.

Contents

Figures

Tables

Chapter 1 General introduction

This thesis is divided into five chapters – a general introduction, three results chapters and a final discussion chapter.

In this introductory chapter, I will provide a brief overview of the history of psychiatric genetics, and the role of common variation in psychiatric disorder risk. I will then focus on what is known about rare copy number variants (CNVs) and their role in neurodevelopmental spectrum disorders and psychiatric disorders. I will go on to provide an overview of the evidence relevant to cognition and depression.

In Chapter 2, I will examine neurodevelopmental CNVs for association with cognitive task results and related functional outcomes in individuals without a neurodevelopmental disorder diagnosis. In Chapter 3, I will examine the role of neurodevelopmental CNVs in risk of depression. In Chapter 4, I will describe the phenotypic spectrum of neurodevelopmental CNVs across psychiatric, cognitive and physical health domains based on my recall of individuals with and without neurodevelopmental CNVs from pre-existing psychiatric cohorts. Finally, I will provide an overview of the results presented in this thesis before discussing how they impact our knowledge of the phenotypic spectrum of neurodevelopmental CNVs and their potential for translation to the clinic.

1.1 The history of psychiatric genetics

Efforts to examine the familial nature of psychiatric disorders began in the mid twentieth century. Here, I provide a brief overview of the history of psychiatric genetics and focus on advances made in the last 10 years.

1.1.1 Family, twin and adoption studies

Before completion of the Human Genome Project, the main methods of examining the familial basis of psychiatric disorders were family studies, twin studies and adoption studies.

In family studies, the frequency of a disorder is established in relatives of individuals affected by a disorder (probands) and compared with the frequency of the disorder in i) individuals from the general population, ii) individuals from a control population, iii) relatives of healthy controls, or iv) relatives of individuals with an unrelated disorder. When carrying out a family study, care must be taken to ensure that probands are ascertained independently of each other. In an ideal situation, assessments should be made of all relatives of probands who are available, making this a time- and resource-intensive method of studying the familial nature of psychiatric disorders.(1)

Twin studies use the fact that twins share their environment but differ on the shared proportion of genes. Monozygotic twins share 100% of their DNA sequence compared with, on average, 50% in dizygotic twins. In twin studies, pairs of twins where one or both individuals are affected by a disorder are studied and the similarity between probands and their co-twins is determined. This is expressed as a concordance rate and there are two variations on this metric - i) pairwise concordance rate - the number of twin pairs with the diagnosis divided by the total number of pairs, ii) probandwise concordance rate - the number of affected twins divided by the total number of co-twins.

In adoption studies, adopted children, their biological parents and adoptive parents are studied. There are three variations on this type of study. In adoptee studies, disorder rates in adopted away offspring of affected parents are compared with those of control adoptees of unaffected parents. In adoptee's family studies, disorder rates are compared in the biological and adopted relatives of adopted individuals who have the disorder. In cross-fostering studies, disorder rates in adoptees with affected biological parents, but raised by unaffected adopting parents, are compared with those in the offspring of unaffected parents brought up by adopting parents who become affected.(1) Greater similarities in levels of disorder risk between children and their biological parents, compared with their adoptive parents, suggests a genetic basis to the disorder in question.

Disorder	Heritability Estimate (%)	Reference
Autism spectrum disorder	64 - 91	(2)
Schizophrenia	81	(3)
Bipolar affective disorder	80 - 85	(4)
Attention deficit hyperactivity disorder	77 - 88	(5)
Major depressive disorder	37	(6)

Table 1.1. Heritability estimates for psychiatric and neurodevelopmental disorders.

Family, twin and adoption studies allowed the calculation of heritability estimates for psychiatric disorders, defined as the proportion of the phenotypic variance attributable to genetic factors (Table 1.1). However, they do not allow disorder-associated genetic variants to be identified. For this, researchers turned to genetic studies.

1.1.2 Early genetic studies in psychiatry

Prior to the advent of high throughput genotyping and sequencing, insights into the genetic basis of psychiatric disorders came from cytogenetic studies and linkage studies.

Cytogenetics is the study of chromosomes, namely their number and morphology. Earlier cytogenetic studies reported increased rates of chromosomal abnormalities in both intellectual disability (ID) and autism spectrum disorder (ASD). These included Fragile X, sex chromosome aneuploidies, duplicated chromosomes (e.g. Down syndrome), chromosomal rearrangements, and some of the larger CNVs (e.g. Prader Willi syndrome).(7–10) In schizophrenia, an increased rate of deletions at 22q11.2 and sex chromosome aneuploidies were reported (11,12), along with translocations which guided early searches for schizophrenia-associated genes. Perhaps the most well supported of these was a balanced

translocation between chromosomes 1 and 11 which occurred in a large Scottish family and segregated with schizophrenia and related disorders - (1;11)(q42.1;q14.3). This translocation was found to disrupt a gene, which was subsequently named Disrupted-In-Schizophrenia 1 (*DISC1*) and became a candidate gene for the disorder.(13)

In linkage studies, markers are genotyped over large genomic regions in family pedigrees and individuals' affected status determined for a particular disorder. These data are then examined for co-segregation of a specific allele at a marker with the disorder in question. The fewer the recombination events, the less the physical distance between the marker and the disease causing variant. These studies often implicated relatively large regions of the genome in disorder risk and, in order to further refine the implicated loci, were followed up with association studies.

In association studies, markers are genotyped across a pre-determined locus or gene, more densely than in linkage studies. Similar to linkage studies, the results are then examined for association with disorder risk. The genomic regions analysed were selected on the basis of being implicated by linkage studies or, in the case of genes, due to them being previously being linked to the disorder. Examples from the schizophrenia literature include the *DISC1* gene (13) and the *DRD2* gene which encodes the dopamine D2 receptor and was implicated by antipsychotic pharmacology.(14)

These techniques were used successfully in neurodegenerative disorders to identify disease-causing variants in specific genes - for example - polyglutamine repeats in the *Huntingtin* gene in Huntington's disease (15) and variants in the *Amyloid Precursor Protein*, *Presenilin 1* and *Presenilin 2* genes in early-onset Alzheimer's disease.(16–20) However, within psychiatry, large numbers of loci were implicated, many associations failed to replicate and convincing causal variants were not robustly identified.(14) At this time, knowledge of the genetic architecture of psychiatric disorders was limited, resulting in a lack of awareness of the small effect sizes we now expect from associations. In hindsight, the historical studies were severely

underpowered to detect true associations and if true associations were found, subsequent studies were underpowered to replicate them.

1.1.3 Conclusions from historical studies in psychiatric genetics

Historical family-based and psychiatric genetic studies clearly implicated genetic factors in risk of psychiatric disorders. However, the inheritance patterns observed in family studies and the failure to find causal mutations using linkage and other techniques argued against a simple Mendelian single gene basis to any of the disorders examined. The data collected suggested that, in the case of major psychiatric and neurodevelopmental disorders, what is being inherited is not the disorders themselves but a predisposition to the development of the disorders. Another striking finding of this work was that these disorders did not 'breed true' - individuals in the same family were often affected by different disorders, one of the earliest lines of evidence for the shared aetiology of some psychiatric disorders.

1.2 Common genetic variation in psychiatric and neurodevelopmental disorders

1.2.1 Genome-wide association studies

In the early 2000s, the Human Genome Project was completed, and it was this, combined with the systematic mapping of human genetic variation and advances in high throughput genotyping, which facilitated the development of genome-wide association studies (GWAS). In GWAS, hundreds of thousands of single nucleotide polymorphisms (SNPs) are genotyped throughout the genome. In practice, the number of SNPs available for analysis can be increased to the millions by the inference of genotypes for non-genotyped SNPs using linkage disequilibrium and reference sequences, a process called imputation. SNP genotypes are then analysed individually or in combinations called haplotypes for association with a disorder. The vast number of markers analysed requires the application of a stringent p value threshold to determine significance - 5×10^{-8}, which accounts for the multiple testing burden of common variation.(21)

As results from the early GWAS came in, there was an increasing recognition that psychiatric disorders are polygenic and effect sizes for common variation are small. In order to find associated loci, very large samples would be required. GWAS for each disorder have got steadily larger and the success of this technique has been largely attributable to the move towards international collaboration in consortia.

The largest GWAS of schizophrenia to date meta-analysed data from individuals with treatment resistant schizophrenia on clozapine, and independent datasets from the Psychiatric Genomics Consortium, resulting in 40,675 cases and 64,643 controls. Pardiñas et al established 145 genetic loci associated with schizophrenia at genome-wide levels of significance and identified 33 putative causal genes. Common variation associated with the risk of schizophrenia was more likely to be present in genes intolerant of loss of function mutation.(22)

The largest GWAS of bipolar affective disorder (BPAD) analysed data from 20,352 individuals with bipolar affective disorder and 31,358 controls. In the first stage, Stahl et al identified 19 variants associated with BPAD risk at genome-wide levels of significance. They then followed up 822 variants with p values of $< 1 \times 10^{-4}$ in an additional sample of 9,412 cases and 137,760 controls. In the combined sample, a total of 30 loci achieved genome-wide levels of significance, 20 of which were novel in BPAD. Further analyses implicated genes encoding ion channels, neurotransmitter transporters and synaptic components and gene sets involved in the regulation of insulin secretion and endocannabinoid signalling.(23)

The largest GWAS of major depressive disorder (MDD) to date meta-analysed data from 246,363 cases and 561,190 controls. Howard et al established 102 independent loci associated with the risk of major depressive disorder, 87 of which replicated in an independent sample of 414,055 cases and 892,299 controls. Further analyses implicated genes and gene pathways involved in synaptic structure and neurotransmission.(24)

The largest GWAS of anxiety and stress-related disorders analysed data from 12,655 individuals with anxiety and other stress-related diagnoses and 19,225 controls. In this study, Meier et al reported 68 variants at a single locus on chromosome 1 as associated with risk of anxiety and stress-related disorders at genome-wide levels of significance. This locus overlaps with the phosphodiesterase 4B gene (*PDE4B*), a gene previously implicated in panic disorder. Further analysis of this gene in mouse models revealed altered expression of PDE4B in mice displaying anxiety-like behaviour.(25)

The largest GWAS of autism spectrum disorder (ASD) to date meta-analysed data from 18,381 cases and 27,969 controls. Grove et al identified five loci as significantly associated with the risk of ASD. Further analysis, leveraging genetic overlap with disorders such as schizophrenia, resulted in the identification of another seven loci associated with ASD risk. There was substantial polygenic heterogeneity across subtypes of ASD and processes involved in neuronal function and corticogenesis were implicated in disorder risk.(26)

The largest GWAS of attention deficit hyperactivity disorder (ADHD) meta-analysed data from 20,183 individuals with ADHD and 35,191 controls. Demontis et al identified 12 loci as associated with ADHD risk at genome-wide levels of significance. Further analyses revealed associated loci to be overrepresented in evolutionary constrained regions and in genes intolerant of loss of function mutation.(27)

GWAS have also allowed the calculation of heritability estimates for the common variation component of the genetic basis of psychiatric disorders - they are not designed to examine rare variation. SNP-based heritability estimates on a liability scale reflect the variance in liability explained by SNPs and the values for the disorders in Table 1.1 are 23% for schizophrenia, 25% for bipolar affective disorder, 21% for major depressive disorder, 28% for attention deficit hyperactivity disorder and 17% for autism spectrum disorder.(28)

1.2.2 Polygenic risk scoring

Individual SNPs increase disorder risk by a very small amount, with odds ratios of up to around 1.1 but it was recognised that, cumulatively, they may account for a much greater proportion of variation in risk. Polygenic risk scoring (PRS) is an approach developed to represent the cumulative contribution of common variants, as a group, to disorder risk or traits. PRS are calculated by adding together an individual's risk alleles weighted by their effect sizes. The result is a score, which reflects an individual's common variation liability to develop the disorder in question. This technique was pioneered by the International Schizophrenia Consortium (ISC) in 2014. The ISC compiled data across loci nominally associated with schizophrenia into quantitative scores. They used the ISC sample as a discovery sample and were able to show that the risk score derived from this sample was associated with schizophrenia in two independent samples. They also showed that risk scores for schizophrenia were associated with BPAD, demonstrating a shared genetic component between the disorders.(29)

The use of PRS in psychiatric clinics to predict future case status is not currently feasible - PRS predicts schizophrenia status weakly with an area under the curve of 0.65 and there is too great an overlap between PRS for cases and controls.(30,31) It is perhaps more likely to have initial applications to the stratification of patients with respect to clinical features such as treatment response. It is, however, currently being used to examine the genetic overlap between psychiatric phenotypes and, psychiatric and non-psychiatric phenotypes in research settings.

1.2.3 The examination of common variation across disorders

Genetic correlation is a statistical genetics approach used to estimate the shared genetic architecture between complex diseases such as those seen in psychiatry. The two main approaches used for the assessment of genetic correlation between separate phenotypes, whether psychiatric or non-psychiatric, are linkage disequilibrium score regression and polygenic risk scoring.

1.2.3.1 Linkage disequilibrium score regression

LD score regression (LDSC) takes summary statistics from GWAS and uses regression analyses to establish the relationship between a SNP's LD score and its GWAS test statistics.(32) This technique may be used to estimate heritability for the common variation component of disorder risk (SNP-based heritability) and to estimate the genetic correlation between phenotypes. A major advantage of LDSC is that it does not require individual-level genotype data, instead using GWAS summary statistics. As a result, it requires far less computing power than alternative methods such as the genomic restricted maximum likelihood (GREML) approach and can be done relatively easily and efficiently on large samples. However, it has been established that LDSC is less accurate than GREML, with accuracy decreasing further when the number of SNPs is reduced and when the degree of genetic heterogeneity between the sample and the reference sample used for LD estimation increases. The latter also results in biased SNP-heritability estimates.(32,33)

1.2.3.2 Using polygenic risk scores to examine genetic overlaps between phenotypes

Where full genotype data are available on large enough samples from GWAS, this permits the use of polygenic risk scores to examine the genetic overlap between phenotypes. This technique has been used to examine the genetic overlap between psychiatric phenotypes and, psychiatric and non-psychiatric phenotypes.

Allardyce et al examined the relationship between genetic risk for schizophrenia and the mood incongruence of psychotic symptoms in BPAD. They calculated PRS scores for schizophrenia in 4,436 individuals with BPAD, 4,976 individuals with schizophrenia and 9,012 controls. They then used regression models to estimate the differential association of schizophrenia PRS across BPAD stratified by phenotypic features and between diagnoses. PRS for schizophrenia was associated with disorders across the schizophrenia-bipolar spectrum with the strongest associations

occurring in those disorders more phenotypically similar to schizophrenia (e.g. schizoaffective disorder - bipolar subtype). In addition, individuals with BPAD I plus psychosis had higher schizophrenia PRS than individuals with BPAD and no psychosis.(34)

The Brainstorm Consortium examined the degree of overlap in the genetic basis of psychiatric, neurological, cognition-related and other phenotypes using LDSC. They utilised GWAS summary statistics from 265,218 patients and 784,643 controls and reported significant genetic correlations between psychiatric disorders, particularly ADHD, BPAD, MDD and schizophrenia. In contrast, neurological disorders had lesser genetic correlations both with other neurological disorders and with psychiatric disorders – they appeared more distinct. They reported positive associations between schizophrenia and years in education but negative associations with intelligence.(35) This paradoxical relationship was further examined by Lam et al using a pleiotropic meta-analytic technique called ASSET (association analysis based on subsets) and GWAS summary statistics for cognitive ability, educational attainment and schizophrenia. Using this method, they were able to identify a subset of genetic variants associated with increased risk of schizophrenia, lower cognitive functioning and lower educational attainment. Gene-based analysis implicated genes involved in early neurodevelopment in this subset. Another subset of genetic variants was associated in the 'unexpected' direction – increased risk of schizophrenia but higher educational attainment. Gene-based analysis implicated adulthood synaptic pruning pathways in this subset.(36) These findings are an example of how genetic studies can be used to dissect out different subtypes of a disorder, with potential relevance for clinical practice.

1.2.4 Conclusions from studies of common variation

Studies of common variation in psychiatry have substantially advanced our knowledge of the genetic architecture of psychiatric and neurodevelopmental disorders, confirming both their heritable and polygenic nature. They have also begun to illustrate how these disorders are related to each other and

implicated specific biological pathways in disorder aetiology. All this taken together provides a route for establishing phenotypes more closely related to the underlying biological processes operating aberrantly, and potentially drug targets.

1.3 Rare copy number variation in psychiatric and neurodevelopmental disorders

Rare genetic variation, occurring in the population at a rate of <1%, is established to increase the risk of several psychiatric and neurodevelopmental disorders. Here, I will discuss the role of rare copy number variants (CNVs) in disorder risk and phenotypic outcomes.

CNVs are segments of DNA 1kb or larger, which are present at variable copy number when compared to a reference genome. Smaller variants which alter copy number are often referred to as insertions or deletions (indels), and when CNVs occur in greater than 1% of the population, they are referred to as copy number polymorphisms.(37) The main types of CNV are: i) deletions – the deletion of a DNA segment, its copy number decreases; ii) duplications – the duplication of a DNA segment, its copy number increases; iii) insertions – the insertion of a DNA segment into another part of the genome and iv) translocations – the rearrangement of chromosomes.(37) Throughout this thesis, I will use the term CNV to refer to deletions and duplications, since insertions and translocations were not examined.

CNVs can be as large as an entire chromosome as in the case of trisomy 21 in Down syndrome (7), so this form of variation has been known about for several decades. CNVs were previously thought to be very rare and often related to regions of the genome rich in repeat sequences.(38) However, high throughput genotyping and genome wide approaches have revealed that, in fact, CNVs occur throughout the genome. They are present in everyone and are a source of genetic diversity which may drive genome evolution.(37,39,40) The majority of CNVs identified are rare and their

frequency is negatively correlated with their size and the number of genes they contain.(41)

1.3.1 Mechanisms of CNV formation

CNVs form during cell division. They are referred to as somatic CNVs if they occur during mitosis and germline CNVs if they occur during meiosis.(42) There are three types of mechanism known to result in the formation of CNVs i) non-allelic homologous recombination, which is responsible for most of the CNVs studied in this thesis, ii) errors in DNA repair - non-homologous end-joining and microhomology-mediated end-joining, iii) errors in DNA replication - fork stalling and template switching, and microhomology-mediated break-induced replication.

1.3.1.1 Non-allelic homologous recombination

Most CNVs established as risk factors for psychiatric disorders are recurrent, occurring at predictable genomic loci. The breakpoints of these CNVs tend to occur in regions with highly repetitive sequences such as low copy repeats (LCRs), also known as segmental duplications (Figure 1.1). LCRs are present in ~5.4% of the genome, and those with particularly high levels of homology appear to result in genomic instability at their locus by causing the misalignment of chromosomes during the crossing over which occurs during recombination.(43,44)

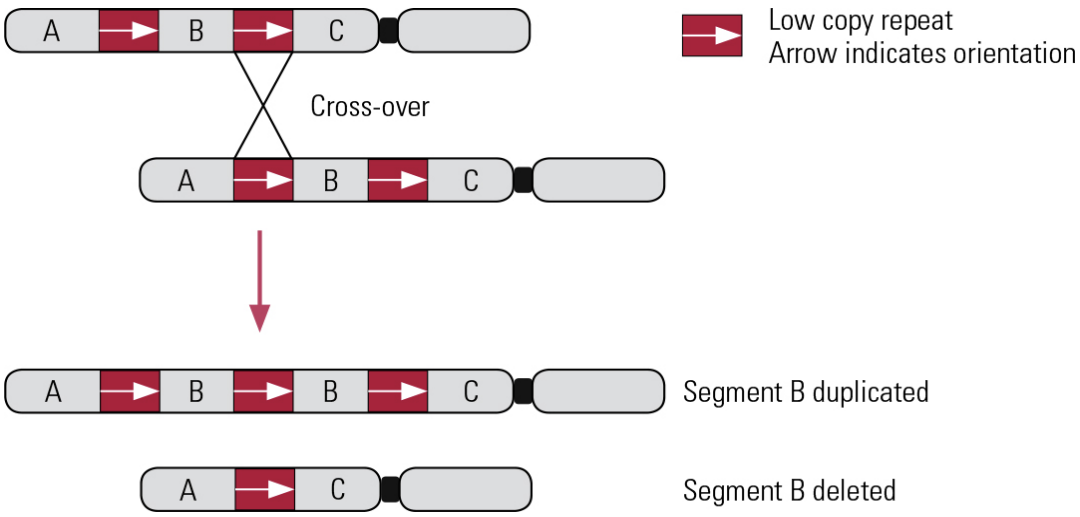

Figure 1.1 Non-allelic homologous recombination. Red boxes and arrows indicate the location and orientation of low copy repeats. Figure reproduced with permission.(42)

1.3.1.2 Errors in DNA repair

Errors in the mechanisms of DNA double stranded break repair can result in the formation of CNVs - non-homologous end-joining (NHEJ) and microhomology-mediated end-joining (MMEJ).(45) During NHEJ and MMEJ, DNA nucleotides can be lost or wrongly inserted resulting in small indels.(46) Errors in these mechanisms often result in CNVs with non-recurrent breakpoints because they are not always associated with genomic architectural features such as LCRs. They do, however, often occur at the sites of other repetitive features like long terminal repeats (LTR) and long interspersed nuclear elements (LINE).(47)

1.3.1.3 Errors in DNA replication

During DNA replication, it is thought that sometimes the DNA replication fork stalls at a particular position. The lagging strand then detaches from the original template and transfers to another replication fork nearby. It then restarts DNA synthesis. This is called fork stalling and template switching (FoSTeS) and is a mechanism by which CNVs with non-recurrent breakpoints are formed.(46,47)

A similar mechanism to FoSTeS is microhomology-mediated break-induced replication (MMBIR). In MMBIR, attempts of a replication fork to proceed through a single stranded DNA break result in fork stalling. This creates a one-ended, double-stranded DNA break that has to be processed differently from a two-ended double-stranded DNA break. Priming of DNA replication on the new fork results in microhomology at the join point (template switch).(48) Once again, the CNVs formed via this mechanism have non-recurrent breakpoints.

1.3.2 Associations between CNVs and psychiatric, and neurodevelopmental disorders

Here I will present the evidence implicating CNVs in risk of psychiatric and neurodevelopmental disorders. I will discuss the evidence in terms of associations with individual loci and associations at the level of CNV burden.

1.3.2.1 CNVs in schizophrenia

Some of the earliest evidence implicating CNVs in the risk of schizophrenia came from a report in 1992 of high rates of chronic schizophrenia in individuals with velocardiofacial syndrome (VCFS), caused by deletions at the 22q11.2 locus.(49) Systematic approaches using operationalised criteria later confirmed 22q11.2 deletions as a risk factor for the disorder.(50) This, and the observation of phenotypic overlaps between schizophrenia and CNV-associated neurodevelopmental disorders such as ASD led Kirov et al to carry out a systematic examination of CNVs in proband-parent schizophrenia trios. The result was the first implication of exonic *NRXN1* deletions in schizophrenia risk, an association which has since been replicated several times over.(51–54) Further CNV loci would later be implicated by large case-control studies.

There are now 12 individual CNV loci with robust evidence for association with risk of schizophrenia, with odds ratios of between 2 and 58 – i) deletions at 1q21.1, *NRXN1* (exonic), 3q29, 15q11.2, 15q13.3, 16p12.1 and 22q11.2, and ii) duplications at 1q21.1, the Williams Beuren syndrome locus at

7q11.23, the Angelman/Prader Willi syndrome locus at 15q11-q13, 16p13.11 and 16p11.2.(53–55) These CNVs have been reported to occur in ~2.5% of individuals with the disorder.(53)

Associations with schizophrenia risk have also been reported at the level of CNV burden. In 2008, the International Schizophrenia Consortium carried out a genome-wide case-control study of CNVs in schizophrenia in 3,391 cases and 3,181 controls. They reported burden of rare CNVs >100kb to be 1.15 times greater in individuals with schizophrenia compared to controls. They also found deletions at the 15q13.3 and 1q21.1 loci to be associated with risk of schizophrenia at genome-wide levels of significance.(56) More recently, Marshall et al carried out a much larger case-control study of CNVs in 21,094 individuals with schizophrenia and 20,227 controls. They reported an enrichment of CNV burden in cases with an odds ratio of 1.11, enriched for genes involved in synaptic function and neurobehavioural phenotypes.(54)

De novo CNVs, CNVs that occur for the first time in a proband i.e. they are not inherited, are another type of CNV implicated in risk of schizophrenia. The rationale for suspecting their involvement is that individuals with schizophrenia have low fecundity so neurodevelopmental CNVs, all of which have large effect sizes, will be quickly removed from the population by natural selection. It stands to reason that their observation in individuals with schizophrenia must mean that they occur de novo. In order to establish whether a CNV is de novo, one must analyse the DNA of the proband's parents, so trios are used for this type of study. In 2008, Xu et al carried out a genome-wide CNV study of 359 individuals with schizophrenia and their biological parents from the Afrikaner population in South Africa. They reported a statistically significant association between de novo CNVs and schizophrenia, which was accounted for by an increased rate in individuals with sporadic schizophrenia, but not familial schizophrenia.(57)

In 2012, Kirov et al carried out a genome-wide CNV study using a sample of 662 schizophrenia proband-parent trios and 2,623 controls. They reported a significant increase in the rate of de novo CNVs in cases (5.1% all cases,

5.5% cases with no family history) compared with controls (2.2%). They provided evidence for de novo CNV occurrence at individual CNV loci already established as schizophrenia risk loci (3q29, 15q11.2, 15q13.3, 16p11.2). They also implicated other neurodevelopmental CNV loci, such as deletions at the 1q21.1 TAR region and duplications at the 7q11.23 Williams Beuren region, providing early support for the neurodevelopmental spectrum. De novo CNVs were enriched for proteins involved in the post-synaptic density and this was largely explained by enrichment for members of the NMDA receptor and ARC postsynaptic signalling complexes.(58)
Finally, Malhotra et al carried out a genome-wide CNV study in 788 trios across the schizophrenia – bipolar spectrum (offspring diagnoses BPAD n = 185, schizophrenia n = 177, and controls n = 426). There was a significant enrichment in de novo CNVs in both schizophrenia (OR = 5) and BPAD (OR = 5.8), particularly for individuals with BPAD with an age of onset of < 18 years. However, this was not the case for schizophrenia with an earlier age of onset.(59)

1.3.2.2 CNVs in bipolar affective disorder

The role of CNVs in bipolar affective disorder (BPAD) is less clear than for schizophrenia and results have been conflicting. Singleton deletion CNVs >100kb have been reported to occur more frequently in BPAD (60,61), as have CNVs in general and de novo CNVs in individuals with BPAD with an earlier age of onset.(59,62) Another study reported the rate of CNVs in BPAD to be intermediate between schizophrenia and controls.(63) However, multiple studies have failed to find evidence for a significantly increased rate of CNVs in BPAD compared to controls.(60,61,64,65) One study even reported a lower rate of CNVs in BPAD compared to other non-psychiatric disorders.(66) At the level of individual CNV loci, associations with BPAD have been reported for deletions at 3q29 and duplications at 1q21.1 and 16p11.2 but it was only the latter which achieved levels of statistical significance which would survive genome wide correction.(67)

These studies may have struggled to establish a role for CNVs in BPAD because of low power – the largest examined data from 2,637 cases (67) and the majority had sample sizes of less than 1,000. In addition, BPAD is a particularly heterogeneous disorder characterised by two extremes of mood disturbance and consisting of at least two phenotypic subtypes (BPAD I and BPAD II). One of the most recent and best powered studies of CNVs in BPAD lends support to the suggestion that disorder heterogeneity may be a problem in establishing the role of CNVs in BPAD. Charney et al examined CNVs in 6,353 individuals with BPAD spectrum disorders. They failed to find evidence for an increased burden of CNVs when BPAD was treated as a single diagnostic entity but reported an increased rate of CNVs in the bipolar subtype of schizoaffective disorder when compared with BPAD I.(68) This would be in keeping with the bipolar subtype of schizoaffective disorder being phenotypically, and presumably aetiologically, close to schizophrenia, as the concept of the neurodevelopmental spectrum would suggest.

1.3.2.3 CNVs in depression

Studies examining associations between CNVs and depression have generated inconsistent results with reported associations not reaching genome-wide levels of significance or being replicated.

In 2010, the first genome-wide association study of CNVs in depression reported an association between duplications at the 5q35.1 locus and major depressive disorder.(69) This association did not replicate in subsequent studies.(70,71) Further associations with depression have been reported for deletions at 7p21.3 and 18p11.32, duplications at 15q26.3, and the combination of deletions and duplications at 16p11.2 (71), and overall burden of 100 - 200kb duplications.(70) None of these associations were at levels of significance, which would survive correction for the number of tests involved. A later genome-wide CNV study reported an association between recurrent depressive disorder and burden of deletion CNVs (72) but a re-analysis of almost exactly the same sample failed to find evidence for this association.(73) More recently, a meta-analysis of four cohorts reported an increased burden of short (<100kb) deletions in patients with major

depressive disorder.(74) Other studies examined specific phenotypic features of depression for association with CNVs – nominally significant findings have been reported for treatment resistance (70), suicide attempts (75) and treatment response.(76)

The inconsistencies between the findings of these studies likely stem from multiple factors. Studies of CNVs in depression have suffered from some of the same problems described above in relation to BPAD, namely power and heterogeneity. The largest study of CNVs in depression for some time examined 3,106 cases (72,73) until Zhang et al's recent study which examined 5,780 cases.(74) Depression is also very phenotypically heterogeneous (77), a factor likely to hinder the search for associations.(78) Each of these studies used different phenotypic definitions of depression and different research interviews – for example Glessner et al used the Composite International Diagnostic Interview to establish a DSM-IV diagnosis of major depressive disorder; Rucker et al used the Schedules for Clinical Assessment in Neuropsychiatry interview to establish an ICD-10 or DSM-IV diagnosis of two or more episodes of moderate severity depression.(69,72) We know from the common variation literature in depression that the phenotypic definition used has a considerable impact on findings.(79) The studies also used different CNV calling methods.

1.3.2.4 CNVs in attention deficit hyperactivity disorder

In attention deficit hyperactivity disorder (ADHD), an excess of large, rare CNVs has been reported at two size thresholds - >100kb and >500kb. Associations were also reported with duplications at the 15q13.3 and 16p13.11 loci and, whilst these results did not achieve genome wide levels of statistical significance, they did replicate in independent samples.(80,81) CNVs greater than 500kb in size have been reported to occur in approximately 12.2% of individuals with ADHD.(80)

1.3.2.5 CNVs in autism spectrum disorder

A substantial body of evidence has implicated rare CNVs in autism spectrum disorder (ASD) with slight differences in the reported rates according to the size of CNV examined. Girirajan reported a rate of rare CNVs >50 kb of 10% in autism. However, this appeared to have been driven by individuals with autism and intellectual disability. When those with autism but without intellectual disability were considered, there was only a modest increase in CNV burden which did not reach statistical significance.(82) De novo CNVs have been strongly implicated in risk of ASD including evidence of association at individual CNV loci – 1q21.1, 3q29, 7q11.23, 16p11.2, 15q11.2-13 and 22q11.2.(83)

1.3.3 Defining a set of neurodevelopmental CNVs

In 2014, Coe et al carried out a large systematic study of rare CNVs in neurodevelopmental disorders. Using array CGH, they compared CNVs in 29,085 children with intellectual disability (ID), ASD or developmental delay (DD) with results from 19,584 healthy adult controls. Cases in this study included 15,767 children selected for inclusion in a previous study on the basis of referral with a general diagnosis of ID and/or DD (73% ID/ASD/DD, 12% no detailed phenotypic annotation, 15% congenital anomalies) and an additional 13,318 children referred with ID and/or DD.(84) They reported an increased rate of rare CNVs in these patients, driven by deletions $\geq$ 500kb (odds ratio 5.09). In 2,086 analysed transmissions, likely deleterious CNVs were transmitted by mothers 58% of the time. They identified associations with ID/ASD/DD for autosomal genomic disorder regions. Using a genomic windowing approach focused on CNV >250kb, they were also able to identify newly significant loci that were novel or previously discussed in the literature at the level of case reports. At the time I put together this fellowship, this was the largest systematic study of CNVs in neurodevelopmental disorders. The authors identified 54 CNV loci as being associated with ID/ASD/DD at nominal levels of significance ($p < 0.05$) but knowledge of their phenotypic spectrum was limited to individuals with childhood neurodevelopmental disorders. Given that we knew CNVs act to increase disorder risk in a non-

specific manner, but it is also possible to carry one of these CNVs and be relatively unaffected, we chose to examine the phenotypic spectrum of the 54 CNVs in individuals with and without psychiatric disorders. Throughout this thesis I refer to these 54 CNVs as neurodevelopmental CNVs (Table 1.2; Appendix 1).(85)

1p36 deletion (*GABRD*)	15q24 deletion
1p36 duplication (*GABRD*)	15q24 duplication
TAR deletion	15q25 deletion
TAR duplication	16p13.11 deletion
1q21.1 deletion	16p13.11 duplication
1q21.1 duplication	16p12.1 deletion
NRXN1 deletion	16p11.2 distal deletion
2q11.2 deletion (*LMAN2L*, *ARID5A*)	16p11.2 distal duplication
2q13 del/dup	16p11.2 deletion
2q13 duplication	16p11.2 duplication
2q37 deletion (*HDAC4*)	17p13.3 deletion (*YWHAE*)
3q29 deletion	17p13.3 duplication (*YWHAE*)
Wolf-Hirschhorn deletion	17p13.3 deletion (*PAFAH1B1*)
Wolf-Hirschhorn duplication	17p13.3 duplication (*PAFAH1B1*)
Soto syndrome deletion	Smith-Magenis deletion
Williams Beuren syndrome deletion	Potocki-Lupski syndrome duplication
Williams Beuren syndrome duplication	17q11.2 deletion (*NF1*)
8p23.1 deletion	17q11.2 duplication (*NF1*)
8p23.1 duplication	17q12 deletion
9q34 duplication (*EHMT1*)	17q12 duplication
10q23 deletion (*NRG3*, *GRID1*)	17q21.31 deletion
Potocki-Shaffer deletion (*EXT2*)	22q11.2 deletion
15q11.2 deletion	22q11.2 duplication
15q11.2 duplication	22q11.2 distal deletion
PWS deletion	22q11.2 distal duplication
PWS duplication	*SHANK3* deletion
15q13.3 del	*SHANK3* duplication

Table 1.2. 54 neurodevelopmental CNVs. These CNVs were selected on the basis of association with ID/ASD/DD at $p < 0.05$ in Coe et al, 2014.(85)

1.3.3.1 The phenotypic spectrum of neurodevelopmental CNVs

Establishing the full range of phenotypes associated with individual neurodevelopmental CNVs is difficult because of their rarity. However, what is clear is that phenotypic outcomes can vary quite considerably for individuals with the same CNV, and many of these CNVs affect multiple body systems. The formation of CNV specific consortia such as the 15q11.2 Working Group has improved the level of phenotypic information available for some of the less rare neurodevelopmental CNVs. An exhaustive list of the phenotypes seen with all 54 neurodevelopmental CNVs is beyond the scope of an introductory chapter so I have focussed on describing the phenotypic features of three of the more common of these CNVs, as examples.

One of the most well-known neurodevelopmental CNVs is the 22q11.2 deletion, which is responsible for DiGeorge and velocardiofacial syndromes. The neurodevelopmental phenotypes of this CNV include developmental delay, delays in the development of expressive language and motor difficulties. Cognitive functioning in individuals with the 22q11.2 deletion falls in a normal distribution, but is shifted to the left in comparison to the general population (mean IQ 70). It is uncommon to see severe ID in the context of this CNV. In terms of psychiatric phenotypes, children with the 22q11.2 deletion have an increased risk of ADHD, ASD and anxiety.(86) Adults with this CNV have an increased risk of any psychiatric disorder (87), and anxiety is also a prominent feature. Carriers have a ~25% risk of developing schizophrenia and the CNV is found in ~1 per 100-200 people with the disorder.(86) The 22q11.2 deletion is an exemplar of the multisystem CNV – physical phenotypes presenting in childhood include immunodeficiency (~75%); congenital cardiac malformations (~75%); hypocalcaemia secondary to hypoparathyroidism (~50%); palatal abnormalities (~75%); gastrointestinal, feeding and swallowing problems (~30%); genitourinary anomalies including renal agenesis (~30%).(86) In adults, this CNV is associated with early onset Parkinson's Disease and is responsible for ~0.5% of the disorder.(88)

One of the CNVs which seems to have less severe phenotypic effects is the 15q11.2 deletion, reflected in its relatively high frequency. The 15q11.2 deletion-associated phenotypes reported in the literature vary considerably, largely due to differences in ascertainment. Cox and Butler carried out a review of the literature of around 200 carriers of the 15q11.2 deletion. Developmental delay, speech delay, difficulties in reading and writing and a verbal IQ of ≤75 occurred in 50% or more of their CNV carriers. 27% of 15q11.2 deletion carriers had ASD, 35% had ADHD and 20% had schizophrenia.(89) However, in a meta-analysis carried out by Jønch et al, this CNV was found to reduce IQ by only 4.3 points, there was no increased risk of ASD and estimates for other disorders were lower than that stated by previous studies – for example the odds ratio for schizophrenia was 1.5. The authors found no real difference in the frequency and nature of symptoms between carriers of deletions and duplications at 15q11.2. They came to the conclusion that many of the symptoms reported as being associated with the deletion are due to ascertainment bias – individuals with a more severe phenotype are more likely to be included in research samples.(90)

The 16p13.11 duplication occurs relatively frequently but evidence of its phenotypic effects is largely based on case series. This CNV has been associated with speech delay, intellectual disability and ASD.(91,92) It has also been implicated as a risk factor for cardiac malformations and thoracic aortic aneurysms and dissections.(91,93) Interestingly, this CNV was found to be enriched in males in an examination of 10,397 individuals with neurodevelopmental disorders and a subsequent search of the DECIPHER database.(94) However, the same result could not be replicated by later work (91), something not accounted for by any obvious difference in ascertainment.

1.3.3.2 Penetrance and ascertainment bias

Neurodevelopmental CNVs have incomplete penetrance, which varies depending on the phenotype under consideration. Individual CNV loci also have quite different penetrance estimates. Using some of the individual

CNVs described above as examples, the penetrance estimates for deletions at 15q11.2 and 22q11.2 are 2% and 12% respectively for schizophrenia, rising to 11% and 88% respectively for any of developmental delay, congenital malformations and ASD. These values are based on individuals with schizophrenia from two research samples, and patients with developmental delay, ASD or congenital malformations referred for genetic testing.(95,96)

The clinical ascertainment of samples for studies of neurodevelopmental CNV-associated phenotypes, by definition, cannot include CNV carriers who appear unaffected as they would not reach the threshold for referral. As a result, existing estimates of the effect sizes of these CNVs are likely inflated. Population studies are required in order to establish the true penetrance of neurodevelopmental CNVs but are much more difficult, and more expensive to conduct. iPSYCH have attempted to call large, rare CNVs from dried neonatal blood spots. They examined rates of deletions and duplications at the 22q11.2 locus in 57,377 individuals with ADHD, MDD, schizophrenia, ASD or BPAD ascertained via nationwide hospital registers and a sample of 30,000 randomly drawn individuals. They reported a population rate of 1:3,672 for the 22q11.2 deletion, and lower disease prevalence estimates compared to previous studies which used clinically ascertained samples. As the authors discuss, for childhood onset disorders they expected complete ascertainment, so the results suggest previous estimates are inflated due to ascertainment bias. This is more difficult to state for later onset disorders such as schizophrenia since the age of the sample under study did not cover the entire risk period for the disorder.(97)

It is clear that the impact of neurodevelopmental CNVs can vary considerably and it is not currently possible to predict the phenotypic outcome of an individual CNV. No neurodevelopmental CNV could be considered necessary or sufficient for the development of the phenotypes discussed so we must also look to other genetic variation (e.g. second hit CNVs, rare single nucleotide variants, common variants) and environmental risk factors for their roles in adding to or modifying risk. For example, evidence from the

fields of cancer and cardiac genetics shows that polygenic risk modifies the penetrance of monogenic variants.(98) There is evidence that individuals with schizophrenia who carry rare pathogenic CNVs also have an excess burden of common risk alleles for the disorder.(99) A recent study of individuals with the 22q11.2 deletion also found that carriers of this CNV who went on to develop schizophrenia, had a significantly higher schizophrenia PRS.(100)

1.3.4 Conclusions from studies of CNVs in psychiatric and neurodevelopmental disorders

Rare CNVs have been strongly implicated in psychiatric/neurodevelopmental disorder risk and their relative contribution to the different disorders appears to fit with the concept of the neurodevelopmental spectrum. There is considerable evidence for their role in risk of disorders classically considered to be neurodevelopmental in origin (e.g. ASD) with this decreasing as the neurodevelopmental loading of a disorder decreases. Individual loci have been implicated in specific disorders but these associations cross diagnostic boundaries. In many of the disorders, there is also a residual CNV burden, suggesting that there are many more CNV-phenotype associations to be found, with the limiting factor being one of power.

1.4 The neurodevelopmental spectrum

Childhood onset neurodevelopmental disorders, such as ID, ASD and ADHD have long been considered to lie on a neurodevelopmental spectrum.(101,102) In the 1980s, the neurodevelopmental hypothesis of schizophrenia had begun to be developed,(103,104) but it was not until the 2000s that researchers began to question whether the apparent gap between schizophrenia and ID/ASD/ADHD was a true one.(105–107) The suggestion that the concept of the neurodevelopmental spectrum be expanded to include schizophrenia was based on several lines of evidence. Schizophrenia and ID/ASD/ADHD share a number of phenotypic features. Cognitive impairment features in all of these disorders to some extent and

each is associated with developmental delay. ID, ASD, ADHD and schizophrenia have increased rates of subtle non-localised signs of neurological dysfunction, termed soft neurological signs, and increased rates of motor abnormalities. Each disorder also occurs more frequently in males than females.(108) In addition, there are elevated rates of comorbidity between the disorders and the boundaries between them are unclear.(109) Finally, the same CNVs implicated in risk of schizophrenia also increase the risk of ID/ASD/ADHD.(53,81,84,110) Similar evidence from the common variation sphere would come later, but would support the same conclusions.

The neurodevelopmental spectrum suggests the conceptualisation of schizophrenia and ID/ASD/ADHD as lying on an aetiological and phenotypic spectrum (Figure 1.2).(111) In this model, the greatest level of neurodevelopmental impairment is observed in ID. As one moves through ASD and schizophrenia, the level of neurodevelopmental impairment gradually decreases until we reach major depressive disorder, which has a relative lack of neurodevelopmental loading. The model also provides an illustration of how specific phenotypic features may vary across disorders in a similar spectrum-based manner. The greater the level of neurodevelopmental impairment the i) greater degree of cognitive impairment, ii) earlier age of onset – e.g. congenital for ID, iii) greater degree of functional impairment, and iv) greater the role for rare, damaging genetic variants, such as CNVs and rare coding variants, in disorder risk.(108)

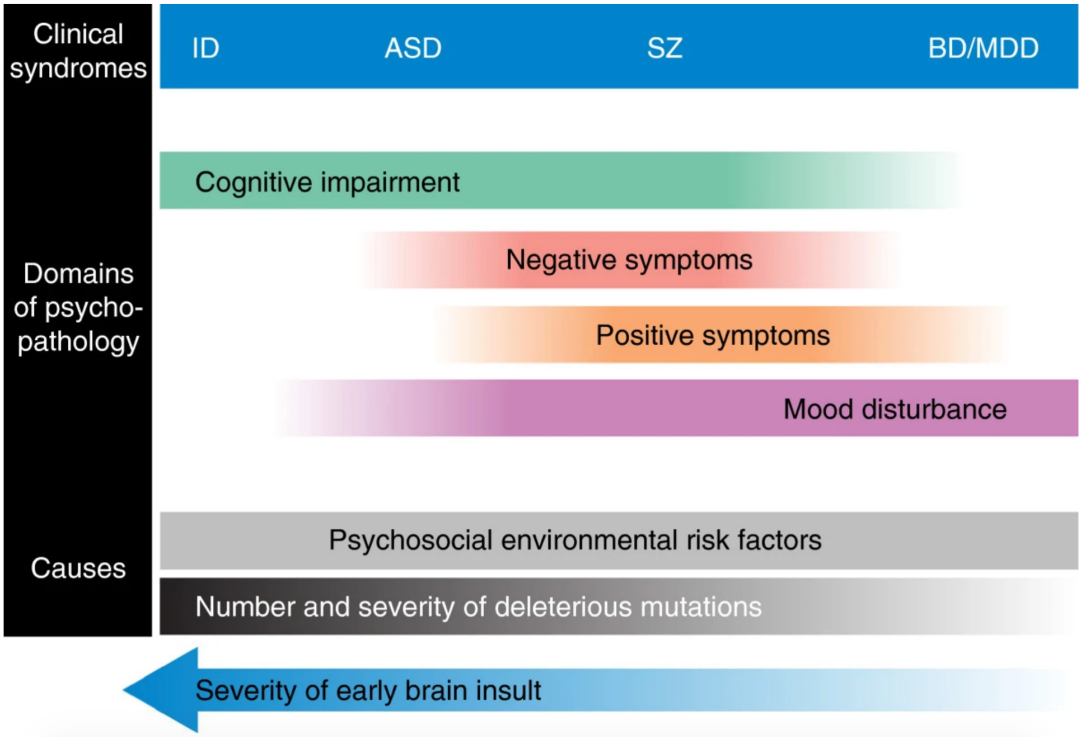

Figure 1.2. The neurodevelopmental spectrum. A simplified schematic of the potential relationship between domains of psychopathology, genetic mutations and clinical syndromes. Reproduced from Doherty and Owen, 2014 with permission.(111)

1.5 Cognition

Cognition is the process of acquiring knowledge and its understanding. Cognitive function can be broken down into domains, for example attention, psychomotor speed, executive function, memory and social cognition. These domains comprise of cognitive processes – for example the memory domain can be broken down into i) working memory – the process by which individuals hold and manipulate information in their minds, ii) episodic memory – the process by which events are associated with places and times, and iii) recognition memory – the process by which an individual recognises information. Cognitive function varies considerably throughout the general population and there is also evidence for variability in cognitive domain performance within individuals.(112) Studies of cognition and ageing have consistently demonstrated a decline in cognitive function with advancing age.(113)

If cognition is the process of acquiring knowledge and understanding, intelligence is the ability to acquire and use it. Like the psychiatric disorders/traits discussed above, intelligence is heritable. Polderman et al carried out a very large meta-analysis of traits from classical twin studies and reported intelligence to have a heritability of 0.54.(114) Some of the early genetic studies of intelligence were successful in confirming this heritable nature of intelligence but failed to find replicable associations with SNPs. Given what we now know about the extreme polygenicity of intelligence, it is clear that these earlier studies, which had sample sizes in the region of 3,000 – 5,000, were vastly underpowered to detect or replicate associations.(115–117) It has only been in the last ~2 years that samples as large as 300,000 have been possible and the result has been many more SNP associations with intelligence. Savage et al carried out a GWAS of intelligence in a sample of 269,867 individuals and reported 205 associated genomic loci.(118) Around the same time, Davies et al carried out a GWAS of intelligence in 300,486 individuals from the CHARGE and COGENT consortia and UK Biobank. They reported 148 associated genomic loci associated with general cognitive function.(119)

1.5.1 Cognitive phenotypes across psychiatric disorders

Cognitive function has been shown to be impaired in all of the neurodevelopmental spectrum disorders, although the nature and degree of impairment varies quite considerably.

1.5.1.1 Cognition in intellectual disability

Cognitive impairment is the core diagnostic feature of ID, a diagnosis of which is made in individuals with an IQ of <70. The ICD 10 classifies the severity of ID by the degree of cognitive impairment – i) mild, IQ 50 – 69; ii) moderate, IQ 35 – 49; iii) severe, IQ 20 – 34, and iv) profound, IQ < 20.(120) The greater the degree of cognitive impairment, the greater the impact on an individual's level of functioning. ID is an umbrella term for aetiologically heterogeneous disorders and risk factors include, but are not limited to, genetic factors, infections, metabolic abnormalities, toxicity and trauma.

Given this level of heterogeneity, there is no specific pattern of domain-based cognitive impairment when ID is treated as a single diagnostic entity. However, patterns begin to emerge when individuals with intellectual disability due to a single cause are examined together. For example, foetal alcohol syndrome can cause intellectual disability. Individuals with this disorder exhibit deficits in executive functioning, especially in tasks involving working memory. However, even in this more aetiologically homogeneous disorder, there is still considerable variability in cognitive performance.(121)

1.5.1.2 Cognition in autism spectrum disorder

Autism spectrum disorder consists of a spectrum of phenotypic presentations, the severity of which can be highly variable. Cognitive impairment is not in the diagnostic criteria for ASD.(120) However, ID and ASD can occur comorbidly with each other – a population based study of 30,037 children ascertained 79 children with severe ID and 99 with mild ID. ASD was the second most commonly co-occurring disorder with ID, behind ADHD.(122)

Individuals with autism have been shown to exhibit deficits in multiple areas of cognition. Pellicano et al carried out a 3 year longitudinal study of 37 children with autism and 31 neurotypical children. In group analyses, individuals with autism experienced difficulties in planning and set shifting and exhibited deficits in false-belief attribution, an index of theory of mind ability. However, individual cognitive profiles varied considerably. Over the 3 year period of the study, theory of mind and executive function changed, whereas central coherence did not (it has been suggested that individuals with autism tend to focus on individual details rather than global pictures).(123)

A systematic review and meta-analysis of cognition in ASD, this time in adults, examined 75 studies encompassing data from 3,361 individuals with ASD and 5,344 neurotypical individuals. Similarly to Pellicano et al's study in children, adults with ASD exhibited deficits in theory of mind. They also

exhibited deficits in emotion perception and processing, processing speed and verbal learning and memory.(124)

At the less severe end of the spectrum, the previously used diagnostic label of Asperger syndrome refers to a form of autism which does not have developmental delay as a core feature. Nevertheless, individuals with this form of autism display uneven cognitive profiles. Bucaille et al compared the cognitive functioning of 32 individuals with Asperger syndrome and matched typically developed controls. Full scale IQ was not significantly different between the two groups but individuals with Asperger syndrome performed more poorly on tests of working memory and executive function.(125)

1.5.1.3 Cognition in attention deficit hyperactivity disorder

Similarly to ASD, cognitive impairment is not in the diagnostic criteria for ADHD. However, they frequently co-occur. Cognitive deficits in ADHD occur across multiple domains and their severity varies between individuals. Most individuals with ADHD have deficits in one or two domains. A much smaller proportion have no cognitive deficits, or deficits in every domain.(126) There is a considerable amount of evidence describing executive functioning deficits in ADHD including impairments in working memory and planning.(127,128) In addition, individuals with ADHD have difficulties with inhibitory control and exhibit multiple deficits in reward processing, e.g. tending to prefer immediate rewards.(129) Other domains shown to be impaired in ADHD have included the processing of information, processing speed, speech and language, arousal and activation and motor control.(130)

1.5.1.4 Cognition in schizophrenia

Cognitive impairment is a core feature of schizophrenia and has been shown to be significantly associated with functional outcomes.(131) Studies examining cognition in schizophrenia have demonstrated generalised deficits across all cognitive domains.(132) The samples included in these studies consisted of individuals with schizophrenia who were middle aged and chronically unwell, laying them open to the question of whether the observed

deficits were a feature of schizophrenia or subsequent changes due to e.g. medication side effects or the effects of being chronically unwell. However, the results have been replicated many times over with individuals with schizophrenia performing, on average 1 – 2 standard deviations lower than healthy controls.(133,134) In addition, studies of antipsychotic naïve patients have suggested that cognitive impairment is not caused by medication. A meta-analysis of studies carried out in this group of patients reported impairments across multiple cognitive domains including speed of processing, attention, executive function, and working and visual memory.(135)

There is strong evidence for impaired social cognition in schizophrenia and performance in this domain has the potential for substantial effects on everyday functioning. In their systematic review and meta-analysis, Fett et al examined the relationship between cognition and functional outcomes. Social cognition was the domain most strongly associated with functioning in the community, explaining 16% of the variance.(131) The strongest association was for theory of mind, the ability to attribute mental states to oneself and to others. Theory of mind deficits have also been observed in other studies – in individuals at the onset of schizophrenia, throughout the disorder and even in those at high risk.(136)

From these lines of evidence, it is clear that there is cognitive impairment in schizophrenia. However, what is not clear is its course. Longitudinal studies have demonstrated that, for some individuals with schizophrenia, cognition deteriorates throughout the disorder. For these individuals, functional outcomes are poorer. Both cognitive impairment at first episode and a decline in cognitive function in the following 1-7 years are associated with poorer social functioning and greater functional disability.(137,138)

1.5.2 Conclusions from studies of cognition

Whilst cognitive impairment is not necessary for a diagnosis of all the neurodevelopmental disorders, it frequently occurs within them and can have

a substantial impact on functioning. The observed cognitive deficits occur across multiple domains within a given disorder and there are marked overlaps in the cognitive impairments observed between these disorders. Together, the existing evidence lends weight to these disorders lying on a neurodevelopmental spectrum and fits with the relative contribution of rare CNVs across it.

1.6 Depression

1.6.1 The depression phenotype

Depression is an affective disorder characterised by the core features of i) depressed mood, which may manifest as feelings of sadness, irritability or emptiness, and ii) anhedonia – a decreased ability to feel pleasure. Individuals with the disorder also experience symptoms from emotional, cognitive and physical symptom groups. The emotional symptoms of depression include feelings of worthlessness or hopelessness, thoughts of death and suicidality. The cognitive symptoms include difficulties in concentrating and making decisions, and subjective memory problems. Physical symptoms of the disorder include fatigue or loss of energy and alterations, in either direction, in appetite or weight, sleep, and psychomotor activity.(139)

Depression is phenotypically very heterogeneous and is perhaps better considered an umbrella term for phenotypically quite different disorders. A study of participants in the STAR*D trial reported over a thousand unique symptom profiles with the commonest being reported by only 1.8% of the sample.(77) A diagnosis of depression may be made using multiple different symptom combinations and there are also many ways of measuring the disorder's severity.(78) In addition, the terms depression, depressive disorders and major depressive disorder appear to be used inconsistently and interchangeably. This has implications for research, its translation to the clinic, and clinical practice.

1.6.2 Epidemiology

The 12-month prevalence of depression is estimated to be in the region of ~6% and the lifetime prevalence ~20%.(140) These figures vary considerably between countries. Peak onset of major depressive episodes occurs from mid-late adolescence extending through to the early 40s, and median age of onset is in the mid 20s.(141)

Established epidemiological risk factors for depression include age, sex, marital status and physical health problems. Women have been reported to have around twice the risk of developing depression compared to men.(140) Some have questioned this finding, suggesting that men present differently with more irritability and externalising symptoms such as substance misuse, which may not necessarily be recognised as depression.(142) If this is the case, the risk of depression in men may have been underestimated.

Most of the epidemiological research into depression has been done in more affluent, Western countries with less research being done in low- to middle-income countries. The evidence regarding age and marital status as risk factors for depression suggests geographical variation. In higher income countries, younger age and being separated from a partner has been reported to be associated with an increased risk of depression episodes.(141,143,144) Research carried out in low- to middle-income countries has either not found these associations, or the associations have been different – with increasing age and being divorced/widowed being risk factors.(140,145,146)

A large body of evidence has reported associations between depression and chronic physical health conditions, with examples including chronic pain syndromes, cardiovascular disease and cancer.(147–152) Depression has often been thought of as a consequence of chronic physical health conditions. However, there is some evidence that the association can occur in the opposite direction – i.e. that depression can contribute to physical health conditions. The evidence here is more limited but depression has

been reported to predict first onset of cardiovascular disease, stroke, diabetes mellitus and some forms of cancer.(153–159) In their review of depression epidemiology, Kessler and Bromet have identified potential mediators of this association reported in the literature including i) poor health behaviours such as smoking and obesity and ii) alterations in the functioning of biological systems such as the hypothalamic-pituitary-adrenal axis.(141) It appears possible that these associations can occur in both directions, and a finding that has important implications for clinical practice is that comorbid depression is associated with a worse course of the physical disorder, possibly due to an association between depression and non-adherence with treatment.(141,160–165).

1.7 Aims, objectives and hypotheses of this PhD

The aim of this PhD is to establish the phenotypic effects of neurodevelopmental CNVs on psychiatric, cognitive and physical health domains in individuals with and without psychiatric disorders. I hypothesise that individuals who carry these CNVs will have a greater burden of impairments across these domains, when compared to CNV noncarriers.

The objectives of this PhD are to:

1. Analyse cognitive and related functional outcome data in the UK Biobank for association with neurodevelopmental CNVs,
2. Examine depression phenotypes and related sub-phenotypes within UK Biobank for association with neurodevelopmental CNVs,
3. Describe the psychiatric, cognitive and physical phenotypes of carriers of neurodevelopmental CNVs with psychiatric disorders in the Cardiff Cognition in Schizophrenia Study cohort and NCMH cohort.

Chapter 2 Cognitive performance of carriers of neurodevelopmental CNVs in the UK Biobank

The research presented in this chapter expands upon work published in *Biological Psychiatry* and the *British Journal of Psychiatry*.(166,167)

2.1 Introduction

The neurodevelopmental disorders associated with large, rare copy number variants (CNVs) are all characterised by varying degrees of cognitive impairment.(53,55,56,85,168,169) Most of the literature concerning the cognitive phenotypic associations of these CNVs is based on individuals at the more severely affected end of the spectrum, and more highly penetrant CNVs such as the 22q11.2 deletion.(170,171) In contrast, relatively little is known about the cognitive phenotypes associated with less penetrant CNVs such as the 15q11.2 deletion.(172)

One thing all neurodevelopmental CNVs have in common, irrespective of their penetrance, is the lack of information on their phenotypic effects in individuals who have not developed overt neurodevelopmental disorders. The limited work done in this area prior to this thesis suggested that individuals who carry neurodevelopmental CNVs have, as a group, cognitive impairment relative to CNV noncarriers, even if they have not developed an obvious neurodevelopmental disorder. An Icelandic study of 144 carriers of 11 autism- or schizophrenia-associated CNVs reported that the cognitive performance of healthy CNV carriers was impaired, with these individuals performing intermediately between CNV noncarriers and individuals with schizophrenia.(173) An Estonian study of 56 carriers of CNVs associated with ‘known syndromes’ reported an association between CNV carrier status and lower educational attainment.(174)

The aim of the work described in this chapter is to establish the cognitive effects of neurodevelopmental CNVs in individuals without

neurodevelopmental disorders. I hypothesise that carriers of neurodevelopmental CNVs will have impaired cognitive performance relative to CNV noncarriers.

2.2 Methods

2.2.1 Sample

Between 2006 and 2010, UK Biobank recruited ~500,000 UK-based individuals from NHS registers based on the distance of their home address from 22 assessment centres in Wales, England and Scotland. 46% of participants were male and, at the time of recruitment, those who took part were 37 - 73 years of age. All participants provided informed consent to participate in UK Biobank projects and the North West Multi-Centre Ethics Committee granted ethical approval for the study (approval number 11/NW/0382).

Participants attended their local assessment centres where they provided demographic, socioeconomic and health-related information, and underwent cognitive assessments on a touchscreen device. They were then interviewed by a research nurse to collect further information, confirm/clarify answers to questions and carry out physical health assessments. Participants provided blood, urine and saliva samples. The UK Biobank work in this thesis was carried out under UK Biobank project number 14421.

2.2.2 Sample processing and genotyping

UK Biobank obtained two 10ml EDTA (ethylenediaminetetraacetic acid) tubes of blood per participant, from which DNA was extracted and purified using a modified Maxwell 16 Blood DNA Purification Kit (Promega – AS1010X). All samples were genotyped at the Affymetrix Research Services Laboratory, Santa Clara, CA. The first ~50,000 blood samples were genotyped on the UK BiLEVE genotyping array, designed for a University of Leicester project investigating genetic variation in chronic obstructive pulmonary disease (807,411 probes).(175) The remaining ~450,000 blood

samples were genotyped on the UK Biobank Axiom genotyping array (820,967 probes). We used the 95% common content between the two arrays for CNV calling.

2.2.3 CNV calling

We downloaded raw genotype .CEL files to a secure UNIX server and processed genotype data in batches of ~4,600. The UNIX commands and relevant parameters are detailed in Table 2.1, given arbitrarily for batch 1. We used the Affymetrix Power Tools apt-probeset-genotype command to generate normalised signal intensity data, genotype calls and confidence scores.(176) We then used PennCNV-Affy to generate genotype clusters using library files for the ~750,000 biallelic markers. We carried out all subsequent steps using PennCNV according to the Affymetrix CNV protocol (penncnv.openbioinformatics.org). We then used the detect cnv.pl command to call CNVs and joined adjacent CNVs using the clean cnv.pl command if they were separated by <25% of the combined length.(177)

2.2.4 Quality control filtering

For all of the UK Biobank work in this thesis, I excluded individuals if they had ≥ 30 CNVs, a genotype call rate < 96%, or a waviness factor > 0.03 or < -0.03 (a metric which accounts for signal intensity dispersion across the genome) and, I excluded individual CNVs if they were covered by <20 probes, or had a density coverage of <1 probe per 20,000 base pairs.

Command	Parameters
apt-probeset-genotype	/apt-probeset-genotype --analysis-files-path /Axiom_UKB_WCSG.r3/ --xml-file /Axiom_UKB_WCSG_96orMore_Step2_Bi-allelic.r3.apt-probeset-genotype.AxiomGT1.xml --out-dir /Batch1 --summaries --cel-files cel.list_batch1.txt
generate affy geno cluster.pl	/generate_affy_geno_cluster.pl AxiomGT1.calls.txt AxiomGT1.confidences.txt AxiomGT1.summary.txt --nopower2 -locfile mapfile.dat -sexfile sex_batch1.txt -out_batch1.genocluster
normalize affy geno cluster.pl	normalize_affy_geno_cluster.pl batch1.genocluster AxiomGT1.summary.txt -nopower2 -locfile mapfileAX.dat -out batch1_lrr_baf.txt
kcolumn.pl	kcolumn.pl batch1_lrr baf.txt split 2 -start 1 -end 1000 -tab -head 3 -name
ls	ls split1*>Batch1_signalfilelist
compile pfb.pl	compile_pfb.pl -listfile Batch1_signalfilelist.txt -output
detect cnv.pl	detect_cnv.pl -test -hmm Axiom_trained.hmm –pfb Axiom.pfb -listfile Batch1_signalfilelist -out --confidence --log –gcmodel Axiom.gcmodel
clean cnv.pl	clean_cnv.pl combineseg Batch1.rawcnv > join batch1.rawcnv -signalfile Axiom.pfb –fraction 0.25 -bp

Table 2.1. Commands and parameters used for the processing of genotype data with Affymetrix Power Tools, PennCNV-Affy and UNIX.

2.2.5 CNV annotation

We compiled a list of 93 CNVs proposed to be pathogenic for neurodevelopmental disorders, and to aid interpretation grouped sub-region CNVs where appropriate e.g. large and small 22q11.2 deletion CNVs.(85,168) We annotated the CNV calls using a Perl script. We compiled a list of CNV calling rules and inspected all CNV breakpoints to confirm they covered the required intervals. This process was particularly important where CNVs were not flanked by low copy repeats. In general, a CNV call was considered valid if the CNV covered >50% of the critical region, including key genes where these are known. In the case of single gene deletion CNVs such as Neurexin 1 (*NRXN1*), deletions could be any size but were required to intersect an exon. For duplications of single genes, the whole gene was required to be duplicated for the CNV call to be considered valid. Where a CNV covered two known, adjacent loci, it was annotated according to the more penetrant CNV. These rules are detailed in full in Appendix 1.

For the analyses in this thesis, I selected the 54 CNVs associated with intellectual disability and autism spectrum disorder at nominal levels of significance - $p < 0.05$ (Table 1.2, Appendix 1).(85) I then excluded the 15q11.2 duplication due to its relatively high frequency resulting in a final list of 53 neurodevelopmental CNVs. For analyses in this chapter, I subdivided the 53 CNVs into the 12 robustly associated with the risk of schizophrenia, and 41 CNVs termed “other neurodevelopmental CNVs”.

2.2.6 Cognitive tests

2.2.6.1 Pairs Matching Test

The Pairs Matching Test examines episodic memory and was completed by 498,737 participants on their first visit to an assessment centre (category 100030, Figure 2.1). Participants were shown symbol cards for 3 seconds (training round - 6 cards; testing round - 12 cards). The cards were turned over and participants were required to identify correct pairs in as few tries as possible. In order to exclude those who did not complete the test, I analysed data on the number of incorrect matches (field 399) for those who eventually

achieved six correct matches (field 398). The results were not normally distributed, so I applied a log+1 transformation to the data before converting them to Z scores.

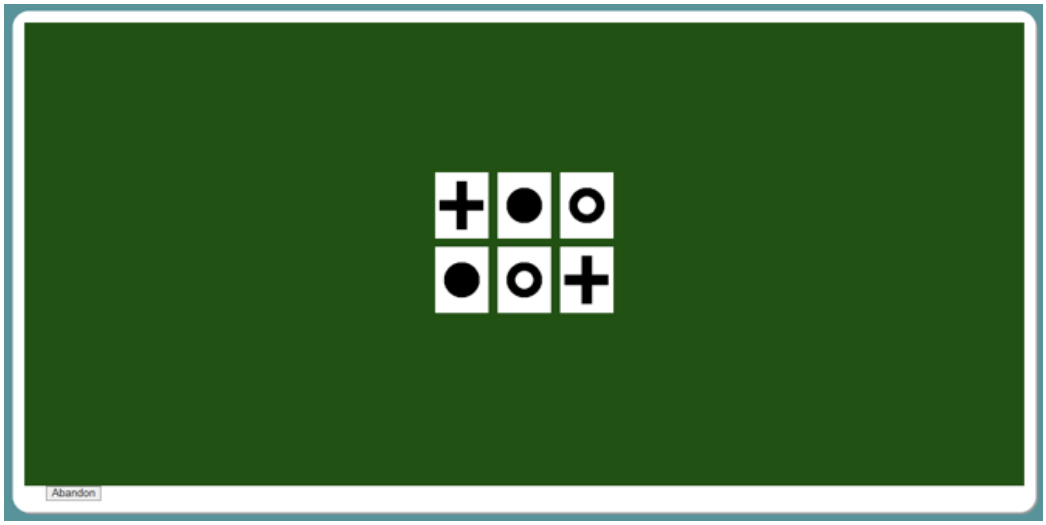

Figure 2.1. An example of the cards shown in the training round of the Pairs Matching Test.

2.2.6.2 Reaction Time Test

The Reaction Time Test examines simple processing speed and was completed by 496,850 participants on their first visit to an assessment centre (category 100032, Figure 2.2). Participants played 12 rounds of the card game 'Snap'. They were shown two cards at a time and required to press a button as quickly as possible if the cards matched. I examined data on the mean time to correctly identify matches (field 20023). The results were not normally distributed, so I recoded outlying scores <100ms and >1500ms as 100ms and 1500ms respectively. I then applied a log transformation to the data before converting them to Z scores.

Figure 2.2. An example of a round of the Reaction Time Test.

2.2.6.3 Fluid Intelligence Test

The Fluid Intelligence Test examines reasoning and problem-solving ability and was completed by 198,160 participants on their first visit to an assessment centre (category 100027). Participants completed as many questions as possible within 2 minutes. I examined data on the number of correct answers (fluid intelligence score - field 20016). The results were normally distributed, and I converted them to Z scores.

2.2.6.4 Digit Span Test

The Digit Span Test examines working memory and was completed by 72,173 participants on their first visit to an assessment centre (Numeric Memory Test, category 100029, Figure 2.3). Participants were briefly shown numbers, which got progressively longer in each round. Once the number had disappeared, they were required to enter it on a number pad. I examined data on the maximum number of digits remembered correctly (field 4282). The results were normally distributed, and I converted them to Z scores.

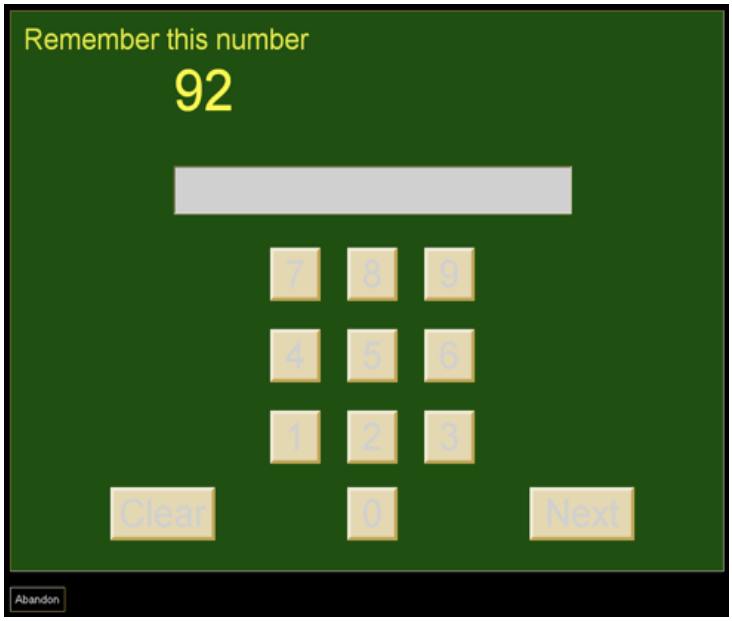

Figure 2.3. An example of a round of the Digit Span Test.

2.2.6.5 Symbol Digit Substitution Test

The Symbol Digit Substitution Test examines complex processing speed and was completed by 118,479 participants at follow-up on home computers (Category 122, Figure 2.4). Participants were required to match numbers to symbols within 2 minutes. I examined data on the number of symbol digit matches made correctly (field 20159). The results were not normally distributed, so I removed outlying scores < 3 and > 36 substitutions. I then converted the results to Z scores.

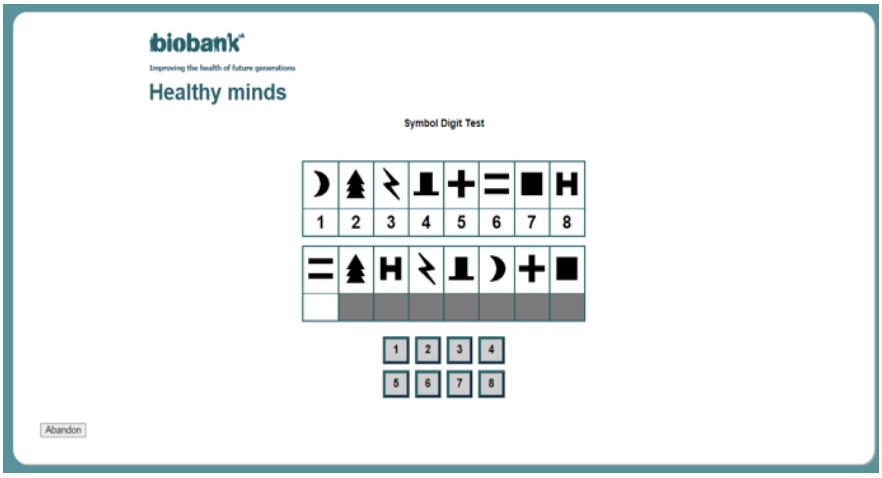

Figure 2.4. An example of a round of the Symbol Digit Substitution Test.

2.2.6.6 Trail Making Tests A and B

The Trail Making Tests A and B examine visual attention and were completed by 104,042 participants during follow-up on home computers (Category 121, Figure 2.5). Participants were required to connect scattered circles according to numbers, and alternating numbers and letters respectively. I examined data on the time taken to complete each task (Trail A – numeric path, field 20156; Trail B – alphanumeric path, field 20157). The results were not normally distributed, so I applied a log-transformation to them before converting them to Z scores.

A B

Figure 2.5. The Trail Making Tests (A – numeric; B – alphanumeric).

2.2.7 Statistical analyses

All analyses were restricted to individuals of European genetic ancestry. In order to define this, we used the covMCD function of the robustbase R package.(178,179) This uses the first five principal components to compute a minimum covariance determinant estimator of location and scatter. We then selected individuals within the 90th percentile of the minimum covariance determinant distance. This procedure is described in detail elsewhere.(180)

The cognitive tests used require proficiency in English, but UK Biobank did not collect data on participants' native language. In order to try to analyse data only from those who speak English as their native language, I restricted analyses to those who self-reported as being of white British or Irish ethnic background (field 21000).

I excluded 975 individuals with schizophrenia, autism spectrum disorder or intellectual disability from the main analyses. I defined this on the basis of self-report and hospital diagnosis codes (fields 20002, 41202, 41204). Schizophrenia is associated with impairments in cognitive performance relative to those unaffected by the disorder so I used data from 799 CNV non-carrying participants with schizophrenia for the comparison of cognitive test effect sizes.(173)

I examined the association between CNV carrier status and cognitive test results using linear regression analyses, with age, sex, genotyping array and the first 15 principal components as covariates. When a group of CNV carriers was not specifically being examined, I excluded relevant individual carriers from the analysis. For example, when I examined the association between schizophrenia CNVs and cognitive test results, I excluded carriers of the 'other neurodevelopmental CNVs'. I examined the association between schizophrenia diagnosis and cognitive test results using the same approach but omitting genotyping array and the first 15 principal components as covariates. I examined the association between CNV carrier status and educational attainment, and occupation using ordinal regression analyses, with age as a covariate and CNV carrier status, sex and the first 15 principal components as factors.

2.3 Results

Data from 400,129 participants were retained following i) exclusion of those who failed CNV quality control, ii) exclusion according to identity by descent, iii) restriction to those of genetic European ancestry and iv) restriction to those of self-reported white British and Irish ethnic background.

2.3.1 Associations between CNV carrier status and cognitive performance

Carriers of schizophrenia CNVs and other neurodevelopmental CNVs exhibited impaired performance on the seven cognitive tests when compared with CNV noncarriers. 12 out of 14 comparisons reached levels of statistical significance that survived Bonferroni correction for 14 tests (p value threshold 0.00357). These results are shown in table 2.2 and presented, for comparison, with results of the same analyses but for schizophrenia diagnosis (Figure 2.6).

I then examined the differences in cognitive performance between carriers of the 12 schizophrenia CNVs and carriers of the remaining 41 neurodevelopmental CNVs. Their performance was similar (Figure 2.6, Table 2.3). The comparison of Reaction Time Test results for the two CNV groups reached levels of statistical significance which would survive Bonferroni correction for seven tests (p value threshold 0.007) - carriers of other neurodevelopmental CNVs performed more poorly than carriers of schizophrenia CNVs.

	CNV noncarriers	Carriers of schizophrenia CNVs			Carriers of other neurodevelopmental CNVs			Individuals with schizophrenia		
	n	n	B (SEM)	p	n	B (SEM)	p	n	B (SEM)	p
Pairs Matching Test (Z)	385,350	3,077	0.13 (0.018)	1.46×10^{-13}	1,397	0.14 (0.026)	7.32×10^{-8}	682	0.35 (0.038)	1.89×10^{-20}
Reaction Time Test (Z)	391,859	3,172	0.18 (0.017)	9.06×10^{-28}	1,430	0.29 (0.025)	1.03×10^{-31}	759	0.62 (0.041)	3.86×10^{-72}
Fluid Intelligence Test (Z)	125,909	979	-0.38 (0.032)	4.78×10^{-32}	451	-0.45 (0.047)	5.28×10^{-22}	242	-0.57 (0.064)	3.69×10^{-19}
Digit Span Test (Z)	40,771	332	-0.27 (0.054)	5.59×10^{-7}	146	-0.28 (0.082)	0.001	66	-0.72 (0.122)	2.79×10^{-9}
Symbol Digit Substitution Test (Z)	93,058	596	-0.22 (0.037)	1.11×10^{-9}	257	-0.27 (0.056)	0.000001	82	-0.66 (0.099)	2.08×10^{-11}
Trail Making Test A (Z)	82,180	530	0.14 (0.041)	0.001	232	0.26 (0.062)	0.000025	67	0.68 (0.116)	4.57×10^{-9}
Trail Making Test B (Z)	82,180	530	0.31 (0.040)	3.28×10^{-15}	232	0.30 (0.06)	5.02×10^{-7}	67	0.53 (0.112)	3.0×10^{-6}

Table 2.2. Association analysis results for 1) schizophrenia CNVs, 2) other neurodevelopmental CNVs and 3) schizophrenia diagnosis with measures of cognition. B - unstandardised beta; SEM - standard error of the mean; p - uncorrected p value.

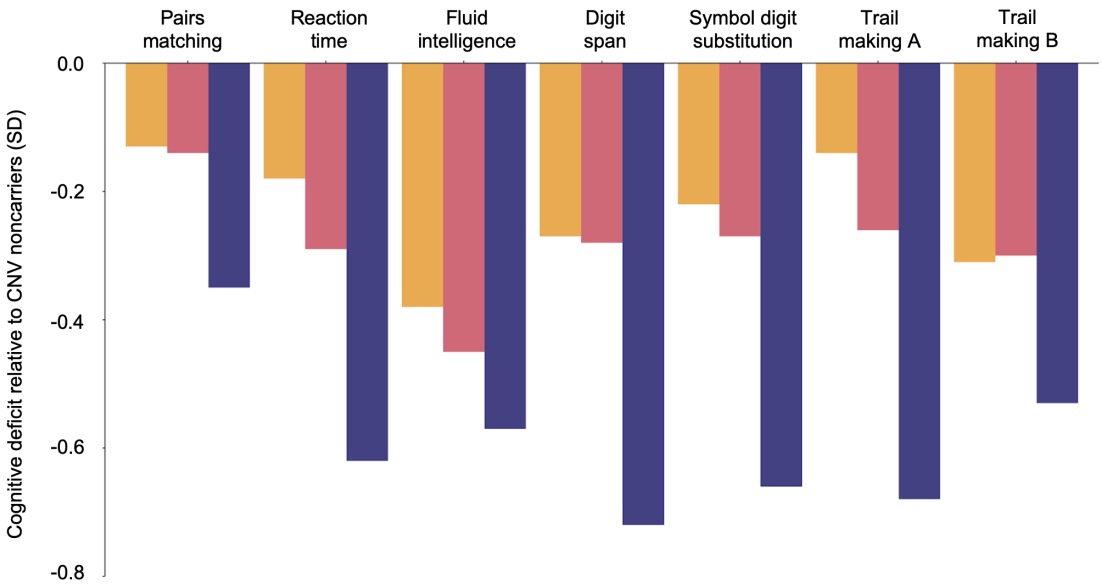

Figure 2.6. Results on seven cognitive tasks for the two groups of CNV carriers and participants with schizophrenia. Orange – schizophrenia CNVs; pink – other neurodevelopmental CNVs; purple – participants with schizophrenia. Results are expressed as Z score differences relative to CNV noncarriers. To aid interpretation, results have been orientated such that results below the line always indicate poorer performance relative to CNV noncarriers.

Test	B (SEM)	p
Pairs Matching Test (Z)	0.011 (0.031)	0.735
Reaction Time Test (Z)	0.113 (0.032)	0.00045
Fluid Intelligence Test (Z)	-0.064 (0.056)	0.247
Digit Span Test (Z)	0.0004 (0.108)	0.997
Symbol Digit Substitution Test (Z)	-0.053 (0.068)	0.437
Trail Making Test A (Z)	0.129 (0.079)	0.103
Trail Making Test B (Z)	-0.004 (0.080)	0.955

Table 2.3. A comparison of cognitive task results for the two groups of CNV carriers. B - unstandardised beta; SEM - standard error of the mean; p - uncorrected p value.

2.3.2 Associations between CNV carrier status and educational and occupational outcomes

Carriers of schizophrenia CNVs and other neurodevelopmental CNVs attained lower qualifications than CNV noncarriers (Figures 2.7 and 2.8). Ordinal regression analyses indicated lower odds for carriers of schizophrenia CNVs to finish in a higher qualifications group - OR 0.59, 95% CI 0.55 - 0.63, p 7.29 x 10^{-59}. Similar results were found for carriers of the other neurodevelopmental CNVs - OR 0.51, 95% CI 0.46 - 0.56, p 2.69 x 10^{-44}. CNV carriers also tended to have occupations that require less training or academic skills. Ordinal regression analyses indicated lower odds for carriers of schizophrenia CNVs to have a job in an occupational group that requires higher skills and longer training, as defined by the Office of National Statistics - OR 0.59, 95% CI 0.55 - 0.64, 2.30 x 10^{-64}.(181) Once again, similar results were found for carriers of other neurodevelopmental CNVs - OR 0.49, 95% CI 0.44 - 0.56, p 6.44 x 10^{-31}.

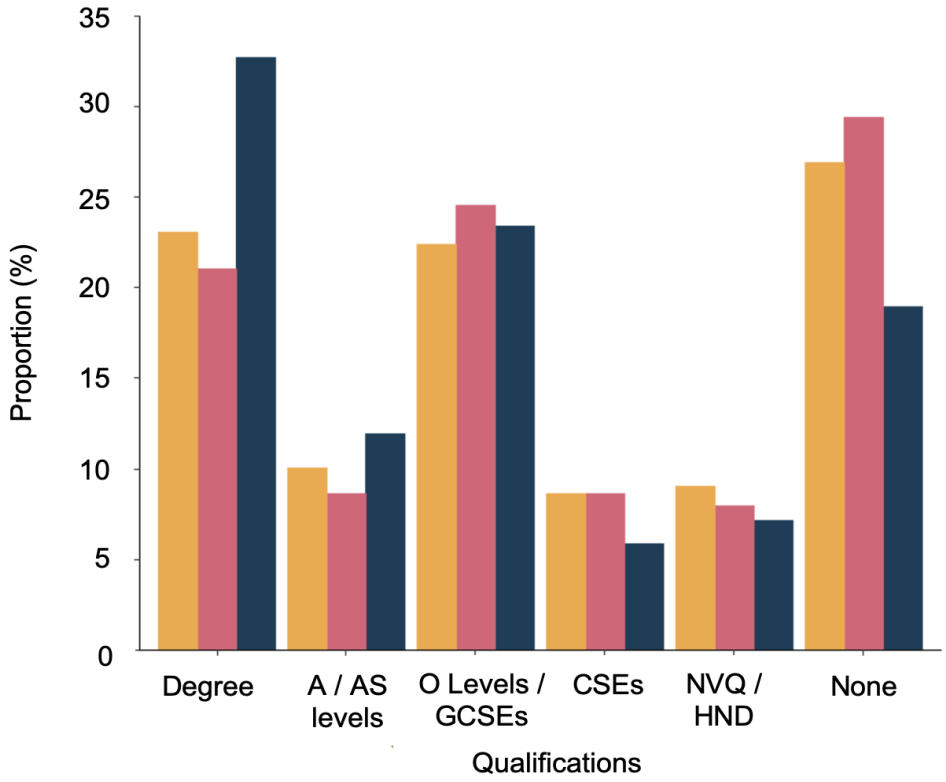

Figure 2.7. The proportion of participants in each group according to their highest qualifications.

Orange – schizophrenia CNVs; pink – other neurodevelopmental CNVs; navy – CNV noncarriers.

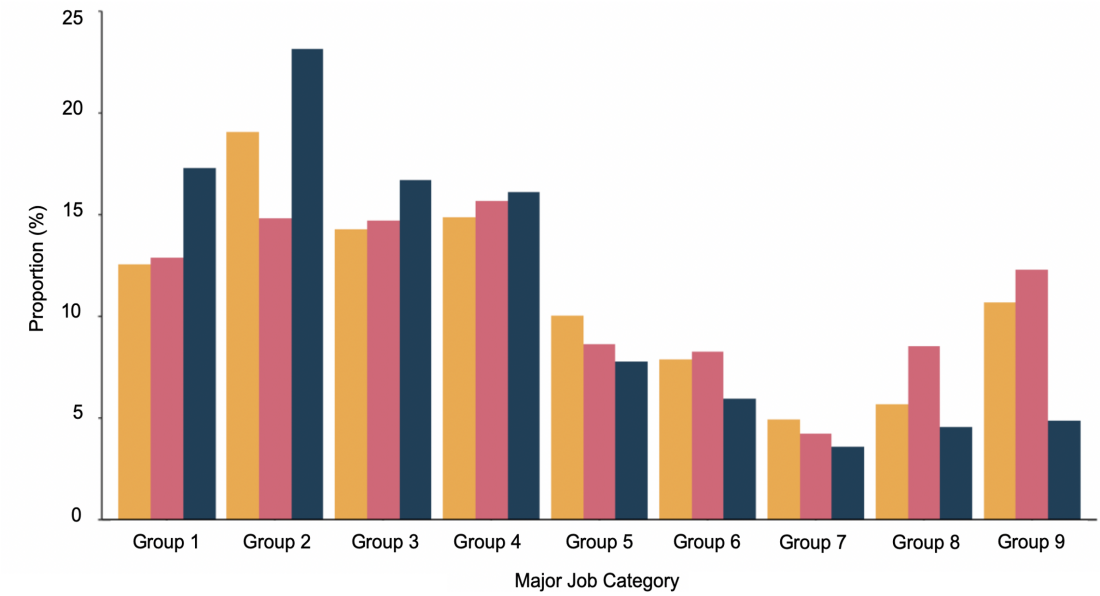

Figure 2.8. The proportion of participants in each group according to their highest occupation

Orange - schizophrenia CNVs; pink – other neurodevelopmental CNVs; navy – CNV noncarriers. Group 1 - managers and senior officials; group 2 - professional occupations; group 3 - associate professional and technical occupations; group 4 - administrative and secretarial occupations; group 5 - skilled trades occupations; group 6 - personal service occupations; group 7 - sales and customer service occupations; group 8 - process, plant and machine operatives; group 9 - elementary occupations.

2.4 Discussion

From the list of 93 CNVs broadly implicated in neurodevelopmental phenotypes, I selected the 54 with evidence of association, at $p < 0.05$, with autism spectrum disorder and intellectual disability. I excluded 15q11.2 duplications because of their relative high frequency and split the remaining 53 CNVs into 12 CNVs established as risk factors for schizophrenia and the remaining 41 other neurodevelopmental CNVs. I examined for associations between CNV carrier status, for both CNV groups, and the results of seven cognitive tasks, educational attainment and occupational attainment.

2.4.1 Carriers of schizophrenia CNVs and neurodevelopmental CNVs have reduced cognitive performance relative to CNV noncarriers

Carriers of the examined CNVs, whether defined as schizophrenia or neurodevelopmental variants, had impaired cognitive performance relative to CNV noncarriers. This was in keeping with the primary hypothesis of this work. The effect sizes observed for these associations were modest (0.13 - 0.45 SD) and tended to be smaller than that observed for associations with schizophrenia diagnosis (0.35 - 0.72 SD). The distinction between schizophrenia CNVs and other neurodevelopmental CNVs appeared to be fairly arbitrary. All schizophrenia CNVs are technically neurodevelopmental CNVs and a statistically significant difference in performance between the two CNV groups was only observed for the reaction time test, with a minimal effect size (0.113 SD).

The results of this work are consistent with that reported in the existing literature. Stefansson et al examined the cognitive performance of 144 health carriers of 11 CNVs - 1q21.1 duplication, *NRXN1* deletion, 13q31.3 duplication, 15q11.2 deletion, 16p12.1 deletion, 16p11.2 deletion, 16p11.2 duplication, 16p13.11 duplication, 17p12 deletion, 17p12 duplication and 22q11.21 duplication. All bar three of these CNVs (13q31.1 duplication, 17p12 deletion and 17p12 duplication) were present in my analyses. The pattern of results observed was the same as that reported by Stefansson et al - CNV carriers had impaired cognitive performance relative to CNV noncarriers and this was intermediate between CNV noncarriers and individuals with schizophrenia.(173)

2.4.2 Carriers of schizophrenia CNVs and neurodevelopmental CNVs are less likely to get a degree and to work in a job requiring more skills or training

Carriers of schizophrenia CNVs and neurodevelopmental CNVs were less likely than CNV noncarriers to have a degree. This finding is in keeping with a study by Männik et al, which reported lower educational attainment in carriers of CNVs associated with known syndromes.(174) Carriers of the examined CNVs were also less likely to have a job requiring greater skills or training. To my knowledge, this is the first time this association has been reported. It is consistent with the association results for the seven cognitive tests and educational attainment.

There may be several explanations for the association between CNV carrier status and the two functional outcomes - educational attainment and occupational attainment. It may be that the reduced educational and occupational outcomes observed in CNV carriers are a direct effect of the cognitive impairment associated with their CNVs. Impairments were observed across multiple cognitive domains, many of which could conceivably impair an individual's ability to engage in learning and work e.g. attention deficits could impair a child's ability to engage in schoolwork. It may also be the case that these CNVs have an effect on areas that do not explicitly come under a cognitive domain but nonetheless have an impact on educational and occupational functioning. For example, where a CNV can cause physical health problems, something applicable to all the CNVs examined, these might impair an individual's ability to attend school, university or a job. These factors, and others, need not be mutually exclusive and may work together to result in these effects on functional outcomes. In addition, there may be unmeasured factors, such as language or social communication, which may be influenced by CNV carrier status and affect an individual's ability to achieve educationally. With respect to studies in psychiatry, one limitation of the UK Biobank sample is that language ability is not well phenotyped.

2.4.3 Impairments in cognition and functional outcomes in the context of the developmental brain dysfunction model

A notable finding was the proportion of individuals who carried a CNV who appeared to function very highly - ~22% had a degree and ~12% had jobs in the highest occupational category. Whilst these individuals, as a group, had impaired cognitive performance relative to CNV noncarriers, it is clear that substantial functional impairment is not inevitable for CNV carriers. This effect may be understood in terms of the developmental brain dysfunction model. This model was developed in an attempt to describe the abnormal brain function underlying neurodevelopmental disorders in a cross-disorder manner, while taking into account the shared risk factors underlying these disorders and their frequent coexistence. Using the developmental brain dysfunction model and data on individuals with deletions at the 16p11.2 and 22q11.2 loci, Moreno-De-Luca et al, suggest that genetic variants such as CNVs reduce an individual's cognitive performance relative to their expected performance, which is based on their genetic background.

For example, parents with high levels of cognitive performance would be predicted to have a child with similarly high levels of cognitive performance. This would be reduced in the presence of a neurodevelopmental CNV but to levels that may fall within the average range and not impair functional outcomes. This is shown in section B of Figure 2.9. Contrast this with the child of parents with average levels of cognitive performance who, based on their genetic background would be predicted to also have average levels of cognitive performance. This would be reduced in the presence of a neurodevelopmental CNV but because the starting point predicted by their genetic background was lower, there is a greater chance their actual cognitive performance would drop to levels associated with an impact on their functional outcomes. This is shown in section A of Figure 2.9.(107) The CNV carriers in the UK Biobank may function so highly based on their favourable genetic background. In order to understand this better, it would be useful to know the cognitive and functional levels of their parents. It would also be possible to use a polygenic score for intelligence as a proxy - current

estimates indicate that polygenic scores for IQ explain up to 5.2% of the variance in intelligence.(118)

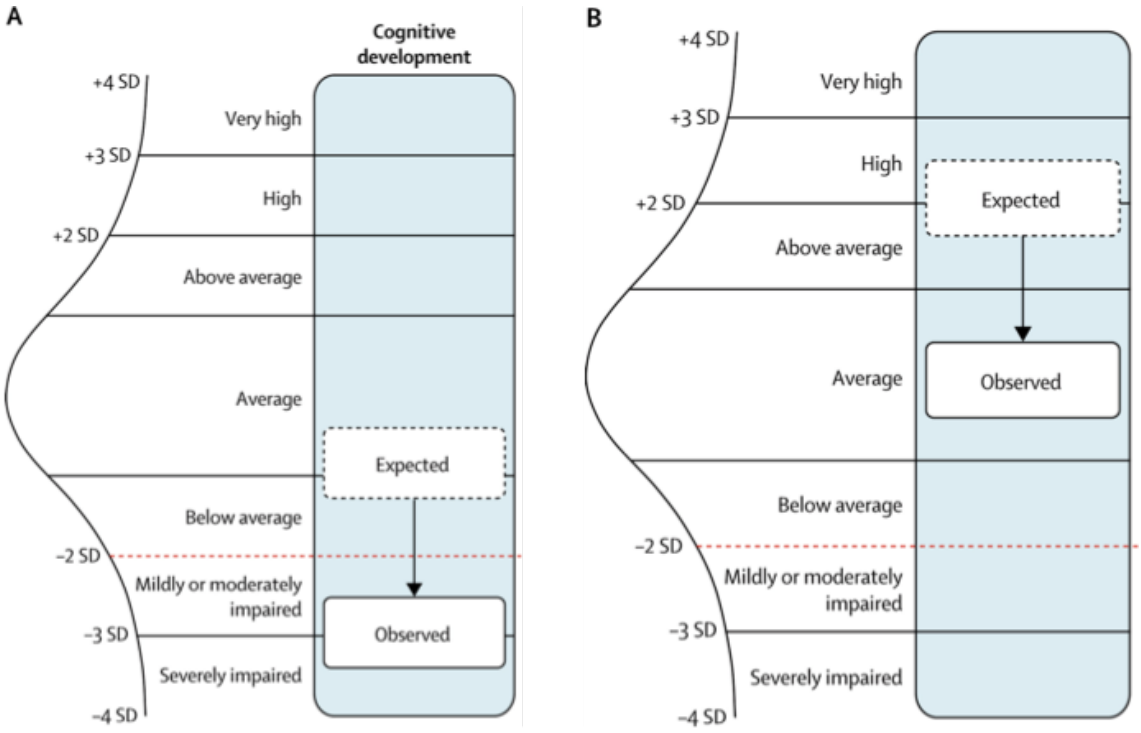

Figure 2.9. The potential effect of neurodevelopmental CNVs on cognitive performance relative to an individual's genetic background as proposed by the developmental brain dysfunction model. A and B depict examples. A – the individual's expected cognitive performance falls within the average range. The effect of the CNV in reducing cognitive performance results in them reaching the threshold for intellectual disability. B – this individual has a higher expected cognitive performance than in A. The CNV exerts an effect of similar magnitude to in A but, because of the higher starting potential, this individual's cognitive performance now falls within the average range. Adapted, with permission, from Moreno-De-Luca et al.(107)

2.4.4 Limitations

This work has several limitations. The UK Biobank is a high functioning sample, as indexed by the low number of individuals with diagnoses associated with lower levels of functioning such as schizophrenia, and high numbers of individuals with high functional outcomes such as having

degrees. Individuals would only have been recruited if they were on NHS registers and more likely to be recruited if they lived in or near a big town or city. As a result of these features of the sample, it cannot be considered to be representative of the general population in the UK. There was considerable variation in the number of individuals who completed the cognitive tasks. For tasks completed at assessment centres (PMT, RTT, FIT, DST), this was because of the people/centres chosen for the tasks and the order in which they were rolled out in the research protocol. I am not aware of any data on any strategy employed in the roll out of the cognitive tasks, but it is possible that it introduced some bias. Tasks completed at home (SDST, TMT) are potentially more likely to be subject to bias – the method of recruitment here was via e-mail. It is possible that the more highly functioning individuals were more likely to take part in this follow-up. These differing recruitment/participation strategies also limit how much the results of the different tasks can be compared. It may have been interesting, and informative, to compare phenotypic characteristics of those who did and did not complete each cognitive task.

2.4.5 Strengths

The main strength of this work is its sample size and the extent of the phenotypic information available. The rarity of neurodevelopmental CNVs means it can be difficult to achieve statistical power to detect CNV-phenotype associations. So, the UK Biobank, with its ~500,000 well-phenotyped participants represents a unique resource for the examination of the effects of rare CNVs. This study remains the largest examination of the cognitive effects of rare CNVs to date.

2.5 Conclusions

Previously, most information on the cognitive effects of neurodevelopmental CNVs came from those at the more severely affected end of the spectrum. I have shown that, whilst it is possible to carry a neurodevelopmental CNV and not develop a neurodevelopmental disorder, it would be incorrect to suggest that these individuals are not affected by their CNV. They have

cognitive impairment, and impairment in related functional outcomes relative to their CNV non-carrying counterparts. This information is a useful addition to knowledge of the phenotypic spectrum of neurodevelopmental CNVs, particularly for genetic counselling, allowing clinicians and patients alike to be fully informed about the range of possible outcomes for CNV carriers. In addition, the observation that neurodevelopmental CNVs can be associated with phenotypes of such varying severity suggests the presence of modifying factors. These may be genetic or environmental and, as they may be tractable to intervention, their further investigation is important.

Chapter 3 The role of neurodevelopmental CNVs in depression

The research in this chapter is based on work published in *JAMA Psychiatry*.(182)

3.1 Introduction

The heritability of depression is estimated at ~37%, a figure lower than that for some of the less common psychiatric disorders such as schizophrenia.(6) Similarly to these other disorders, the common variation component of genetic risk for depression consists of risk conferred by many alleles of small effect. To date, 102 independent genetic loci have been reported as associated with the risk of depression. 48 of these associations were in the direction of increasing depression risk with odds ratios of between 1.019 and 1.049.(24) Rare CNVs have the potential to increase risk of depression by a far greater amount than individual SNPs, a reflection of their substantially larger size. However, their role in depression has, to date, remained unclear.

Studies of CNVs in depression have generated inconsistent results with reported associations not reaching genome-wide levels of significance or being replicated. In 2010, the first genome-wide association study of CNVs in depression reported an association between duplications at the 5q35.1 locus and major depressive disorder (1,693 cases).(69) This association did not replicate in subsequent studies.(70,71) Further associations with depression have been reported for deletions at 7p21.3 and 18p11.32, duplications at 15q26.3, and the combination of deletions and duplications at 16p11.2 (604 cases) (71), and overall burden of 100 - 200kb duplications (1,263 cases).(70) None of these associations were at levels of significance, which would survive correction for the number of tests involved. A later genome-wide CNV study reported an association between recurrent depressive disorder and burden of deletion CNVs (72) but re-analysis of almost exactly the same sample failed to find evidence for this association (3,106

cases).(73) Some studies have examined specific phenotypic features of depression for association with CNVs – nominally significant findings have been reported for treatment resistance (452 cases) (70), suicide attempts (189 cases) (75) and treatment response (1,565 cases).(76)

The inconsistencies in these results may stem from multiple factors. Each study used different phenotypic definitions of depression and different research interviews. For example, Glessner et al used the Composite International Diagnostic Interview to establish a DSM-IV diagnosis of major depressive disorder; Rucker et al used the Schedules for Clinical Assessment in Neuropsychiatry interview to establish an ICD-10 or DSM-IV diagnosis of two or more episodes of moderate severity depression. Each study used different CNV calling methods and was underpowered to detect and replicate associations at appropriate levels of significance to allow for multiple testing.

Previous work from the MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University has shown that in schizophrenia, a substantial proportion of the CNV enrichment observed is explained by CNVs reported to be involved in neurodevelopmental disorders (intellectual disability and congenital malformations).(55,183) Depression shares genetic risk with schizophrenia (184) and frequently occurs comorbid with neurodevelopmental disorders.(185,186) Based on these data, I hypothesised that neurodevelopmental CNVs would be associated with increased rates of depression. I also tested the more general hypothesis that, outside of neurodevelopmental CNVs, there would be a residual association signal for other rare CNVs.

3.2 Methods

3.2.1 Sample

UK Biobank recruited ~500,000 individuals (46% male) aged 37 – 73 years between 2006 and 2010. Participants attended assessment centres where they provided demographic, socioeconomic and health related information

and underwent physical and cognitive assessments. All participants provided informed consent to participate in UK Biobank projects and the North West Multi-Centre Ethics Committee granted ethical approval to UK Biobank. The sample is described in detail in Chapter 2.

3.2.2 Depression phenotypes

There are multiple potential approaches to defining depression in the UK Biobank (UKBB) and a lack of consensus as to the best way to proceed.(187,188) With this in mind, I chose to begin by using a relatively liberal definition of lifetime depression, rating as cases individuals who reported a doctor had told them they have depression (self-reported depression code 1286, UKBB field 20002). I then repeated the main analyses using two alternative, more conservative definitions of depression - lifetime self-reported depression with current antidepressant prescription at the time of visit 1 and hospital discharge diagnosis of depression. *Self-reported depression with antidepressant prescription at visit 1* - I constructed a binary depression variable using (i) the self-reported depression code 1286 in UKBB field 20002 and (ii) antidepressant prescription codes in UKBB field 20003. Individuals were included as affected if they reported that a doctor had told them they have depression and they were prescribed an antidepressant medication at the time of first assessment. Individuals, who fulfilled only one of the two criteria i.e. self-reported depression or antidepressant prescription alone, were excluded from the analyses. *Hospital discharge diagnosis of depression* – individuals were included as affected if they had a hospital admission with a primary or secondary ICD-10 code for depression (UKBB fields 41202 and 41204) (Figure 3.1). For individuals assessed at Scottish assessment centres, hospital records covered general medical hospitals but not psychiatric hospitals. For Wales and England both general medical and psychiatric hospitals were covered. For individuals who attended Scottish assessment centres, I accepted a secondary ICD-10 code for depression as evidence of depression diagnosis as this could be coded during admission to a general medical hospital. However, in the absence of primary ICD-10 codes for depression, I was unable to determine the absence

of depression in controls, so I removed Scottish unaffected individuals from this variable.

In 2017, 157,397 individuals completed an online follow-up mental health questionnaire. I used data from this questionnaire to attempt to further characterise associations with depression phenotypes in individuals who stated that they had ever experienced prolonged feelings of sadness or depression (UKBB field 20446). I examined data from the following variables - (i) age at first episode of depression (UKBB field 20433), (ii) duration of worst depression (UKBB field 20438) and (iii) lifetime number of depressed episodes (UKBB field 20442). *Duration of worst depression* (UKBB field 20438) – this was coded in ranges of months e.g. less than a month, between one month and three months. Previous data has shown that the median duration of a depressive episode is 3 months.(189) Therefore, I dichotomised this variable into 0-3 months and more than 3 months. Lifetime number of depressed episodes (UKBB field 20442) – I dichotomised this variable using a median split approach (median = 1).

3.2.3 CNV calling

We carried out CNV calling using PennCNV-Affy using biallelic markers common to both genotyping platforms. This is described in detail in Chapter 2.

3.2.4 Defining CNV sets and statistical analysis

Following the approach adopted in Chapter 2, I defined a group of neurodevelopmental CNVs as those 54 CNVs for which there is at least nominally significant evidence for association with intellectual disability or autism spectrum disorder.(85) I excluded 15q11.2 duplications due to their relative high frequency resulting in a final list of 53 neurodevelopmental CNVs. In exploratory analyses, I analysed individually each of the CNVs for which there were ≥5 observations for association with self-reported depression. These results were then subject to correction for 53 tests.

I carried out CNV burden analyses using PLINK 1.07 on regions of variable copy number at three size thresholds: (i) ≥ 100KB, (ii) ≥ 500KB and (iii) ≥ 1MB.(190) I filtered CNVs for frequency at <1% using the --cnv-freq-exclude-above command and filtered out overlapping low copy repeat regions using the –cnv-exclude command. I converted PLINK outputs into CNV carrier status, which was used as the predictor in regression analyses. For these analyses, carriers of the group of 53 CNVs associated with neurodevelopmental disorders were excluded.

I carried out association analyses in R using logistic or linear regression as appropriate with age, sex, genotyping array and the first 15 principal components as covariates. Analyses were restricted to individuals of European genetic ancestry using a variable generated by Dr Sophie Legge (detailed in Chapter 2). I excluded individuals who had ASD, ID, ADHD, schizophrenia or bipolar affective disorder diagnosed by a doctor (UKBB fields 20002 and 20544) or coded during a hospital admission (UKBB fields 41202/41204).

3.2.5 Further investigation of the neurodevelopmental CNV-depression association

In order to better understand the association between neurodevelopmental CNVs and depression I investigated whether the association was explained by variables known to be associated with depression (191–194) and postulated to be associated with CNVs – (i) educational attainment, (ii) physical health, (iii) social deprivation, (iv) smoking, or (v) alcohol consumption.

Educational attainment - Prior to data analysis, I dichotomised and recoded data from the academic qualifications field into college/university degree or all other qualifications, an approach previously used by Davies et al for this data field (UKBB field 6138).(195) *Physical health* - I used affected status for one of the medical phenotypes associated with these CNVs.(196) *Social deprivation* - I used Townsend Deprivation Index, a metric which

incorporates, for an individual's home area, percentage unemployment, percentage of households without a car, percentage of households who do not own their home and household overcrowding (UKBB field 189). *Smoking* - I examined this using smoking status (UKBB field 20116). *Alcohol consumption* - I examined this using alcohol intake frequency (UKBB field 1558). I carried out analyses using structural equation modelling in the lavaan package in R.(197)

3.2.6 Sex-specific analyses

An excess of rare CNVs ≥ 500KB has previously been reported in females.(198) Recently, Martin et al examined for association between rare CNVs and neurodevelopmental problems, anxiety and depression in 12,982 children (5.3% neurodevelopmental problems; 3% anxiety or depression). They reported a higher rate of large CNVs in female children with anxiety or depression when compared to male children.(199) Since I found an excess of female CNV carriers with depression, I added an interaction term to the regression model, consisting of the product of neurodevelopmental CNVs and sex.

3.3 Results

Following exclusions (detailed in Figure 3.1), 23,979 individuals (5.89%) had self-reported depression and 383,095 individuals reported no lifetime depression.

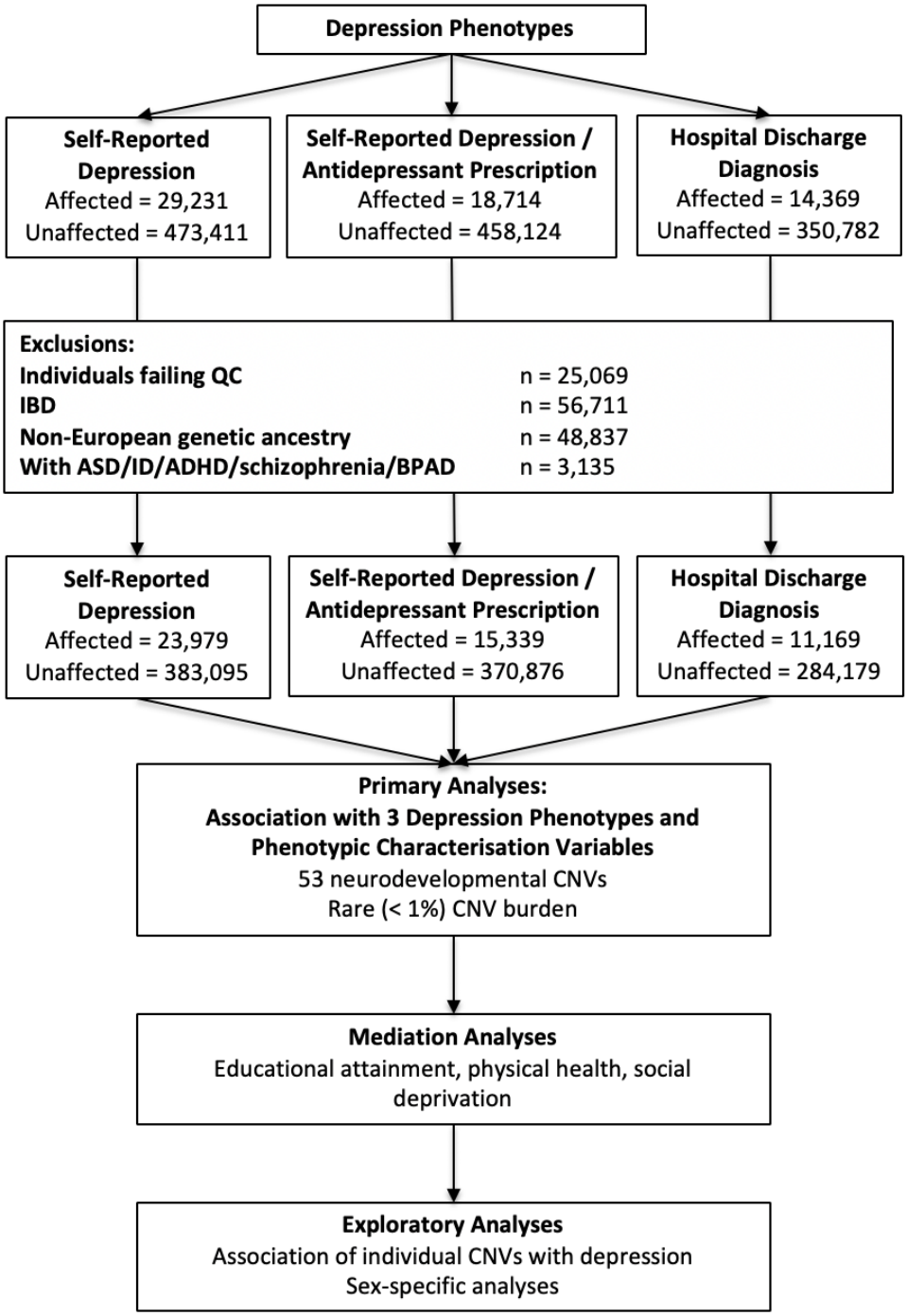

Figure 3.1. A flowchart of the study design.

3.3.1 The association between neurodevelopmental CNVs and depression

In the primary analysis, the group of 53 neurodevelopmental CNVs was associated with depression - OR 1.34, 95% CI 1.19 - 1.49, uncorrected p value 1.38×10^{-7} (Table 3.1). Of those individuals defined as affected according to the self-reported depression variable, 1.51% (n = 363) carried at least one of the 53 neurodevelopmental CNVs compared with 1.14% (n = 4,368) of those defined as unaffected. Analyses using the alternative, more conservative, depression variables gave consistent results and the effect sizes observed increased with the more conservative definitions used (Table 6). When individuals likely to have a more severe depressive illness, defined as a hospital discharge diagnosis of depression, were removed there remained an association between neurodevelopmental CNVs and self-reported depression - OR 1.35, 95% CI 1.21 - 1.51, uncorrected p 5.48×10^{-8}.

Following exclusion of carriers of the 53 neurodevelopmental CNVs, there was a weak association between CNVs ≥ 500KB and depression. This did not survive correction for the number of tests performed. There was no evidence for an association between CNVs ≥ 100KB, ≥ 1MB and depression, however it was defined (Table 6). I carried out exploratory analyses of individual neurodevelopmental CNV loci for association with risk of self-reported depression. Eight CNVs were nominally associated with self-reported depression and three of these associations survived Bonferroni correction for 53 tests - 1q21.1 duplication, Prader Willi syndrome duplication, 16p11.2 duplication (p value threshold 0.00094, Table 3.2).

	Neurodevelopmental CNVs		CNVs ≥ 100kb		CNVs ≥ 500kb		CNVs ≥ 1mb	
	OR 95% CI	p	OR 95% CI	p	OR 95% CI	p	OR 95% CI	p
Self-reported depression 23,979 affected 383,095 unaffected	1.34 1.19 - 1.49	1.38×10^{-7}	1.01 0.98 - 1.03	0.58	1.05 1.005 - 1.10	0.029	1.01 0.94 - 1.08	0.80
Self-reported depression with antidepressant prescription on visit 1 15,339 affected 370,876 unaffected	1.42 1.25 - 1.62	1.18×10^{-7}	1.02 0.98 - 1.05	0.53	1.08 1.02 - 1.14	0.01	1.02 0.94 - 1.12	0.59
Hospital discharge diagnosis of depression 11,169 affected 284,179 unaffected	1.51 1.30 - 1.75	2.95×10^{-8}	1.04 1.00 - 1.08	0.04	1.08 1.01 - 1.15	0.03	1.04 0.94 - 1.15	0.41

Table 3.1. Association analysis results for neurodevelopmental CNVs, and measures of CNV burden, and three depression phenotypes. OR - odds ratio; 95% CI - 95% confidence interval; p - uncorrected p value. For analyses of burden, carriers of the 53 neurodevelopmental CNVs were excluded.

CNV	n	OR	95% CI	p
TAR deletion	70	0.95	0.29 - 2.29	0.92
TAR duplication	405	1.17	0.77 - 1.69	0.44
1q21.1 deletion	104	1.11	0.46 - 2.22	0.79
1q21.1 duplication	173	2.17	1.34 - 3.36	**9.08×10^{-4}**
NRXN1 deletion	160	2.01	1.18 - 3.19	0.0057
2q11.2 deletion	31	2.34	0.69 - 6.02	0.11
2q13 deletion	51	0.98	0.24 - 2.67	0.97
2q13 duplication	68	1.29	0.49 - 2.90	0.59
3q29 deletion	8	11.22	2.27 - 46.52	0.001
8p23.1 duplication	6	9.64	1.32 - 50.21	0.009
15q11.2 deletion	157	1.11	0.90 - 1.35	0.30
PW duplication	17	8.14	2.77 - 21.69	**4.61×10^{-5}**
15q13.3 deletion	43	0.77	0.13 - 2.52	0.72
15q24 duplication	8	1.89	0.10 - 10.74	0.55
16p13.11 deletion	126	2.21	1.25 - 3.63	0.003
16p13.11 duplication	791	0.87	0.63 - 1.78	0.39
16p12.1 deletion	236	1.47	0.90 - 2.27	0.09
16p11.2 distal deletion	57	2.23	0.92 - 4.63	0.05
16p11.2 distal duplication	131	1.57	0.82 - 2.73	0.14
16p11.2 deletion	110	1.21	0.54 - 2.34	0.60
16p11.2 duplication	124	2.65	1.53 - 4.31	**2.04×10^{-4}**
Potocki Lupski duplication	5	4.31	0.22 - 29.82	0.19
17q11.2 deletion	9	2.16	0.12 - 11.86	0.47
17q12 duplication	95	1.55	0.69 - 3.02	0.23
22q11.2 deletion	10	1.69	0.09 - 9.17	0.62
22q11.2 duplication	267	1.72	1.12 - 2.53	0.009

Table 3.2. Association analysis results for individual neurodevelopmental CNVs and self-reported depression. n – number of CNV carriers (affected and unaffected combined), OR – odds ratio, 95% CI – 95% confidence interval, p – uncorrected p value. Results, which survived Bonferroni correction for 53 tests are shown in bold.

3.3.2 The association between neurodevelopmental CNVs and depression severity

I used data from 157,397 individuals who completed an online follow-up mental health questionnaire to examine the association between CNVs and three markers of depression severity - age at onset, number of episodes of depression, duration of worst depressive episode. On this questionnaire, the phenotype of ever experiencing prolonged feelings of sadness and depression was associated with neurodevelopmental CNV carrier status - OR 1.20, 95% 1.07 - 1.36, p 0.002. I restricted these analyses to the 68,684 individuals who reported experiencing these feelings. There was no evidence for an association between neurodevelopmental CNVs and any of the three markers of depression severity that survived correction for the number of tests performed (Table 3.3).

CNV Class	Effect Size (95% CI)	p
20446 Ever had prolonged feelings of sadness or depression 68,684 affected, 57,243 unaffected		
Neurodevelopmental CNVs	1.20 (1.07 - 1.36)	0.002
CNVs ≥ 100kb <1%	0.99 (0.98 - 1.02)	0.96
CNVs ≥ 500kb <1%	0.99 (0.95 - 1.03)	0.65
CNVs ≥ 1MB <1%	0.96 (0.90 - 1.02)	0.25
20433 Age at first episode of depression, n = 65,106		
Neurodevelopmental CNVs	-0.07 (-0.15 - -0.002)	0.06
CNVs ≥ 100kb <1%	-0.001 (-0.016 – 0.014)	0.86
CNVs ≥ 500kb <1%	0.002 (-0.027 – 0.030)	0.91
CNVs ≥ 1MB <1%	-0.005 (-0.047 – 0.037)	0.82
20438 Duration of worst depression (0-3 months vs > 3 months), n = 69,971		
Neurodevelopmental CNVs	0.98 (0.84 – 1.14)	0.74
CNVs ≥ 100kb <1%	0.99 (0.97 – 1.03)	0.78
CNVs ≥ 500kb <1%	0.99 (0.94 – 1.05)	0.82
CNVs ≥ 1MB <1%	1.07 (0.98 – 1.17)	0.10
20442 Lifetime number of depressed episodes (1 vs 2+), n = 57,482		
Neurodevelopmental CNVs	1.04 (0.87 – 1.24)	0.68
CNVs ≥ 100kb <1%	1.00 (0.97 – 1.04)	0.86
CNVs ≥ 500kb <1%	1.02 (0.96 – 1.09)	0.51
CNVs ≥ 1MB <1%	1.13 (1.03 – 1.25)	0.009

Table 3.3. Analyses of depression sub-phenotypes for individuals who stated they had ever experienced prolonged feelings of sadness or depression. Effect size – odds ratio (except for age at first episode of depression – standardised Beta), 95% CI – 95% confidence interval, p – uncorrected p value.

3.3.3 Further investigation of the neurodevelopmental CNV-depression association

I investigated whether the association between neurodevelopmental CNVs and depression could be explained by measures of educational attainment (qualifications), physical health (presence/absence of an associated medical phenotype), social deprivation (Townsend Deprivation Index), smoking (smoking status) and alcohol consumption (alcohol intake frequency). I chose these variables because of their known associations with depression,(191–194) their postulated or proven associations with CNVs (166,167) and their availability in a large proportion of the UK Biobank sample. The association between neurodevelopmental CNVs and depression was partially explained by each variable examined - educational attainment 1.2%, physical health 2.9%, social deprivation 8.1%, smoking 4.8% and alcohol consumption 16.6% (Figure 3.2, Table 3.4). When all of these measures were incorporated into the regression analysis, there remained a strong independent association between neurodevelopmental CNVs and depression - OR 1.26, 95% CI 1.11 – 1.43, p 2.87 x 10^{-4}.

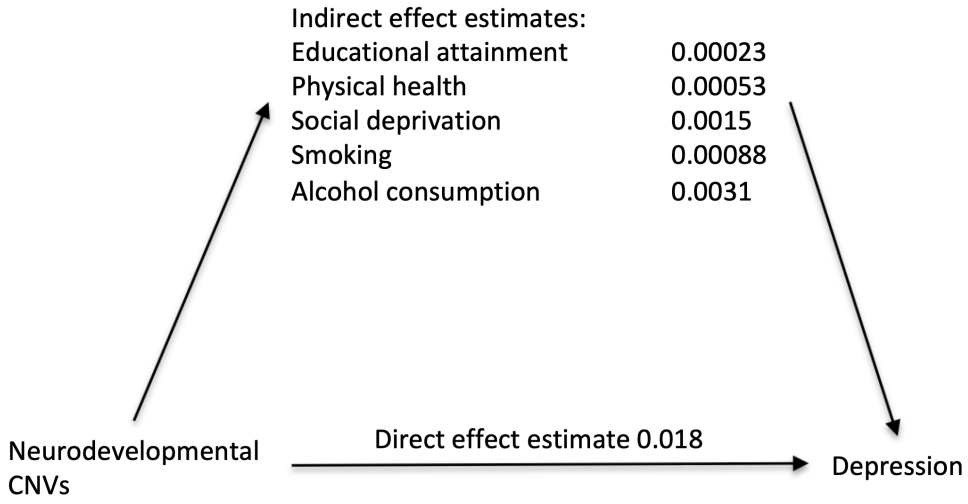

Figure 3.2. Further investigation of the association between neurodevelopmental CNVs and depression via potential explanatory variables. Numbers shown are the estimates for direct and indirect effects calculated using lavaan.(197)

Indirect Effects	Proportion Explained	Estimate	Standard Error	p
Explanatory Variable				
Educational attainment	1.2%	0.00023	7.78×10^{-5}	0.0032
Physical health	2.9%	0.00053	1.85×10^{-4}	0.0045
Social deprivation	8.1%	0.0015	1.59×10^{-4}	< 0.0001
Smoking	4.8%	0.00088	1.58×10^{-4}	2.9×10^{-8}
Alcohol consumption	16.6%	0.0031	2.39×10^{-4}	< 0.0001
Sum of Indirect Effects	21.5%	0.0039	3.85×10^{-4}	< 0.0001
Direct Effect				
Neurodevelopmental CNVs		0.018	0.0038	2.5×10^{-6}

Table 3.4. Direct and indirect effect results of analyses examining whether the association between neurodevelopmental CNVs and depression could be explained by other variables. The proportion explained was estimated by indirect effect/total effect. SE – standard error, p – uncorrected p value.

3.3.4 Sex specific analyses

Following the emergence of evidence for an increased rate of large CNVs in female children with anxiety or depression,(199) I undertook exploratory analyses examining the differences in depression rates between male and female CNV carriers. 10.24% of female carriers of neurodevelopmental CNVs had self-reported depression compared with 5.02% of male carriers of neurodevelopmental CNVs (chi square 45.45, p 1.22×10^{-11}). The increased rate in female neurodevelopmental CNV carriers was over and above the baseline-increased rate of self-reported depression in females - interaction term OR 0.66, 95% CI 0.53 – 0.83, uncorrected p = 0.0002. However, the effect was weaker for the secondary depression definitions (Table 3.5).

Self-Reported Depression 23,979 affected 383,095 unaffected		Self-Reported Depression and Antidepressant Prescription on Visit 1 15,339 affected, 370,876 unaffected		Hospital Discharge Diagnosis of Depression 11,169 affected 284,179 unaffected	
OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p
All (females and males)					
1.34 (1.19 – 1.49)	1.38×10^{-7}	1.42 (1.25 – 1.62)	1.18×10^{-7}	1.51 (1.30 – 1.75)	2.95×10^{-8}
Females					
1.46 (1.28 – 1.67)	2.23×10^{-8}	1.51 (1.29 – 1.78)	3.75×10^{-7}	1.67 (1.39 – 2.01)	2.66×10^{-8}
Males					
1.14 (0.95 – 1.38)	0.17	1.27 (1.01 – 1.59)	0.041	1.28 (0.99 – 1.63)	0.052
Interaction term (product of neurodevelopmental CNVs and sex)					
0.66 (0.53 – 0.83)	0.0002	0.67 (0.51 – 0.87)	0.003	0.67 (0.49 – 0.91)	0.009

Table 3.5. Association analyses of neurodevelopmental with the three depression phenotypes according to sex. OR – odds ratio, 95% CI – 95% confidence interval, p – uncorrected p value.

3.4 Discussion

I carried out association analyses for 53 neurodevelopmental CNVs and depression, and also examined for residual burden of large, rare CNVs in 407,074 individuals with European genetic ancestry. To my knowledge, this is the largest depression CNV study to date. I used as my primary depression definition self-reported depression - self-report of ever receiving a medical diagnosis of depression. In order to ensure that my findings were not just present in this definition of depression, I also selected two more conservative depression phenotypes for analysis - lifetime self-reported depression with antidepressant prescription at the time of visit 1 and an ICD-10 hospital discharge diagnosis of depression. The results support my main hypothesis - neurodevelopmental CNVs were associated with an increased risk of depression according to all three phenotypic definitions. These associations could not be explained by associations with neurodevelopmental or neuropsychiatric diagnoses since individuals with these disorders were excluded from the analyses. There was no evidence to support my second hypothesis - the weak association between CNVs ≥ 500KB and depression would not survive correction for the number of tests performed and there was no evidence for an association between CNVs ≥ 100KB, ≥ 1MB and depression. These results are not consistent with a meta-analysis carried out around the same time which reported an association between CNVs <100kb and major depressive disorder with these CNVs being over-represented in intergenic regions. However, that study was smaller (5,780 cases and 6,626 controls).(74)

I found associations between three neurodevelopmental CNV loci and self-reported depression which would survive Bonferroni correction for 53 tests - 1q21.1 duplication, Prader-Willi syndrome duplication and 16p11.2 duplication. On both an individual and a group level, the risk of depression in CNV carriers was lower than that identified for previous studies of schizophrenia. Qualitatively, however, the results followed a similar pattern - for both disorders the highest risk was conferred by deletion CNVs at the 3q29 locus (OR depression 11.22, OR schizophrenia 57.65, and the lowest

risk was conferred by deletion CNVs at the 16p12.1 locus (OR depression 1.47, OR schizophrenia 3.3).(53,55)

The Prader Willi locus is an imprinted region. Prader Willi syndrome (PWS) results from a loss of expression of paternally expressed genes at this locus – either due to the deletion of paternally inherited genes or the inheritance of two copies of the maternally marked genes (maternal uniparental disomy – mUPD). Individuals with PWS have a high rate of psychiatric disorders. A study of 17 adolescents and 21 adults with the diagnosis found that the commonest psychiatric disorder was anxiety, but rates of depression were also high, frequently occurring comorbidly with anxiety.(200) There is some evidence that depression may be more common in individuals with the paternal deletion form of the disorder (201) but the genetic abnormality associated with depression in this study was the duplication. The mUPD form of PWS is associated with psychosis with a large effect size – lifetime prevalence of psychosis in these individuals has been estimated at 60-100%.(202) 6 of the 17 carriers of the PW duplication self-reported a depression diagnosis and I examined the data to try to establish more about the nature of their depressive disorders e.g. evidence of psychosis. Unfortunately, only one of these individuals completed the mental health questionnaire but I was able to establish that two had received hospital diagnoses of depression, suggesting a more severe form of the illness.

3.4.1 The association between neurodevelopmental CNVs and depression is partially explained by multiple other variables

Structural equation modelling analyses provided evidence that the association between neurodevelopmental CNVs and depression could be partially explained by measures of educational attainment, social deprivation, physical health, smoking status and alcohol consumption. To my knowledge, this is the first time an association has been reported between large, rare CNVs and measures of social deprivation. This may be an important mechanism by which CNV carrier status could increase the risk of

depression, but longitudinal data is required to establish the direction of association between depression and social deprivation.

3.4.2 Limitations

This work has a number of limitations. My primary phenotypic definition of depression required self-report. This approach is subject to information bias.(203) However, the consistent results based on a phenotypic definition using clinicians' hospital discharge diagnoses would suggest that the self-report aspect of the primary definition is unlikely to have substantially affected the findings. There was also a relatively low rate of depression when compared to population estimates.(204) This may reflect the high functioning nature of the UK Biobank sample combined with an imprecise definition of depression. However, these factors seem unlikely to have generated spurious CNV associations, instead having the effect of diluting associations with CNV carrier status.

3.5 Conclusions

This piece of work was the first, to my knowledge, to robustly demonstrate the association of neurodevelopmental CNVs with risk of depression. This extends our knowledge of the phenotypic spectrum of these CNVs and reinforces that carriers of neurodevelopmental CNVs who may not have developed a neurodevelopmental disorder, cannot be assumed to be unimpaired. Along with cognitive and physical health phenotypes, there are wider implications for depression and social deprivation which must be considered in assessing CNVs at the population level.

Chapter 4 Clinical and cognitive effects of risk factors for mental illness

4.1 Introduction

In Chapters 2 and 3, I showed that individuals from a population sample who carry a neurodevelopmental CNV have, relative to CNV noncarriers, i) impairments in cognitive function and related functional outcomes, and ii) an increased risk of depression. Existing psychiatric genetic samples, by virtue of their size, have been successfully used to establish associations between CNVs and case status for neurodevelopmental disorders.(53,55,80,82) However, they lack the depth of phenotypic information required to establish their effects on individuals with psychiatric disorders and to dissect out phenotypes which may better reflect underlying biology. The aim of this study was to carry out in-depth phenotypic assessment of individuals with psychiatric disorders and CNVs, and to compare their phenotypes with those of their CNV non-carrying counterparts.

In this study, I chose to focus on negative symptoms, the assessment of which presents a particular challenge in both research and clinical practice. Data on negative symptoms are not typically available in large scale recruitment with more minimal phenotyping. Negative symptoms are characterised by deficits, are difficult to quantify, and are often not volunteered by patients during assessment. As a result, and perhaps because they do not present as dramatically as the positive symptoms of psychosis, they are under-researched. Nonetheless, negative symptoms are an important cause of disability, with substantial effects on functional outcomes, and existing treatments are of limited effectiveness.(205)

Negative symptoms are a core feature of schizophrenia but cross-disorder research has demonstrated their presence in a range of other psychiatric and neurological disorders including bipolar affective disorder, major depressive disorder, autism spectrum disorder and neurocognitive disorders.(206)

Factor analytic studies of negative symptoms have revealed a 5 factor structure – i) anhedonia – a deficit in the ability to experience pleasure; ii) avolition – a deficit in motivation to engage in goal-directed behaviour; iii) asociality – a deficit in social behaviour; iv) blunted affect – deficits in facial and emotional expression; v) alogia – a deficit in verbal output.(207)

A substantial body of evidence from psychiatric genetics and neuroscience research supports the existence of psychiatric disorders on a continuum, and argues against existing diagnostic labels representing aetiologically and biologically distinct disorders.(108) In an attempt to conceptualise psychiatric and related symptoms in a way that might better reflect underlying biology, the National Institute of Mental Health (NIMH) developed the Research Domain Criteria (RDoC) project. In RDoC, negative symptoms are conceptualised as the Positive Valence system, which is primarily responsible for responses to positive motivational situations, with additional input from the Cognitive systems.

The Positive Valence system consists of three constructs, which in turn each comprise of three sub-constructs (Table 4.1). *Reward responsiveness* – this construct consists of i) reward anticipation - processes involved in the ability to anticipate and/or represent future incentive, ii) initial response to reward - processes evoked by the initial presentation of a positive reinforcer, iii) reward satiation - processes associated with the change in the incentive value of a reinforcer over time. *Reward learning* – this construct consists of i) probabilistic and reinforcement learning, ii) reward prediction error - processes associated with a difference between anticipated and obtained rewards, iii) habit - behaviours elicited that can go to completion without constant conscious oversight. *Reward valuation* – this construct consists of i) reward probability - processes involved in the computation of the value of a reinforcer, ii) delay - processes involved in the computation of the value of a reinforcer before its expected delivery, iii) effort - processes involved in the computation of the value of a reinforcer and the perceived costs of the effort to obtain it.

The Cognitive systems construct mapped by RDoC to negative symptoms is cognitive control and is described as *"a system that modulates the operation of other cognitive and emotional systems, in the service of goal-directed behaviour, when prepotent modes of responding are not adequate to meet the demands of the current context".* The cognitive control construct consists of four sub-constructs - i) goal selection, ii) updating, representation and maintenance, iii) response selection, and iv) inhibition / suppression.(208)

Assessment tools mapped by RDoC to Positive Valence and Cognitive system sub-constructs are proposed to reflect the processes underlying the separate elements of negative symptoms. I used these mappings to select phenotypic assessments for this study (Table 11). Results of this work could, through further research in areas such as neuroimaging, be used to elucidate the biological circuits involved in these symptoms facilitating the development of targeted treatments.

The CNVs examined in this study were selected because of their established associations with neurodevelopmental disorders such as autism spectrum disorder (ASD) and intellectual disability (ID).(85) These CNVs vary considerably in their size, and the number and function of genes they affect but their rarity necessitates their analysis as a group. Given their associations with cognitive impairment and related functional outcomes, I hypothesise that neurodevelopmental CNVs will exert a generalised effect across all measures of negative symptoms. I also hypothesise that CNV carriers will have a higher frequency of disorders from the neurodevelopmental spectrum, and a higher rate of physical health problems.

Construct	Sub-construct	Measure/Scale
Positive Valence system		
Reward responsiveness	Reward anticipation	CAINS TEPS - anticipatory subscale
	Initial response to reward	TEPS - consummatory subscale
	Reward satiation	
Reward learning	Probabilistic and reinforcement learning	CANTAB Cambridge Gambling Task - risk adjustment
	Reward prediction error	TEPS - anticipatory subscale
	Habit	
Reward valuation	Reward probability	
	Delay	Monetary Choice Questionnaire
	Effort	BAS - drive subscale
Cognitive systems		
Cognitive control	Goal selection	CAINS, CANTAB One Touch Stockings of Cambridge
	Updating, representation and maintenance	CAINS, CANTAB One Touch Stockings of Cambridge
	Response selection	CANTAB Stop Signal Test
	Inhibition / Suppression	CANTAB Stop Signal Test

Table 4.1. The RDoC Positive Valence and Cognitive Control systems with mapped assessment tools. CAINS - Clinical Assessment Interview for Negative Symptoms; TEPS - Temporal Experience of Pleasure Scale; CANTAB - Cambridge Neuropsychological Test Automated Battery; BAS - Behavioural Activation Scale.

4.2 Methods

4.2.1 Samples

CNV carriers and CNV noncarriers with psychiatric disorders were recruited from existing studies within Cardiff University, which had carried out CNV calling prior to this study. I gained ethical approval for this study from the School of Medicine Ethics Committee, Cardiff University - SMREC 16/38.

Cardiff Cognition in Schizophrenia Study (Cardiff COGS)

Individuals aged 17 – 84 years, 41% female, with schizophrenia and other psychotic disorders, schizoaffective disorder and bipolar affective disorder were recruited from community, inpatient and voluntary sector mental health services in the UK. Participants underwent a Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview and the data obtained were combined with casenote review to arrive at a best-estimate lifetime diagnosis according to DSM-IV criteria.(209,210) CNVs were called in 965 participants using PennCNV by Dr Elliott Rees as previously described.(177)

National Centre for Mental Health

Individuals aged 12 – 103 years, 57% female, were recruited to the National Centre for Mental Health (NCMH) cohort from community, inpatient and voluntary sector mental health services in the UK, and also from the general public via specific campaigns. Individuals with any psychiatric disorder were eligible to be included in the cohort, along with some recruited on the basis of being a relative of someone with a psychiatric disorder or a control. Participants self-reported psychiatric disorders with which they had ever been diagnosed by a healthcare professional. CNVs were called in 2,465 participants using PennCNV by Dr Leon Hubbard.(177)

Reference sample

Early in this study, I recruited 29 individuals with no history of mental health problems from the community, via adverts in e.g. leisure centres, and on Gumtree. I screened these individuals for the presence of psychiatric

disorders using the Mini-International Neuropsychiatric Interview.(211) All 29 individuals completed the CANTAB One Touch Stockings (OTS), Emotion Recognition (ERT), Reaction Time (RTT), Spatial Working Memory (SWM), Paired Associates Learning (PAL) and Verbal Recognition Memory (VRM) tasks and these data were used in the calculation of Z scores.

4.2.2 CNV quality control and annotation

We adopted a quality control approach consistent with the UK Biobank work presented in this thesis - we excluded individuals if they had ≥ 30 CNVs, a genotype call rate <96%, or a waviness factor > 0.03 or < -0.03 and, we excluded individual CNVs if they were covered by <20 probes, or had a density coverage of <1 probe per 20,000 base pairs. Carriers of 54 CNVs associated with autism spectrum disorder and intellectual disability at nominal levels of significance ($p < 0.05$) were selected for recall.(85)

4.2.3 Recall Process

In the Cardiff COGS sample (n = 965), 24 participants had at least one of the 54 neurodevelopmental CNVs, and 76 did so in the NCMH cohort (n = 2,465). For each CNV carrier, I double checked to ensure that the CNV call was correct and that it passed the QC parameters described above. I matched each CNV carrier to two CNV noncarriers according to age and sex. I then ensured that each individual had consented to being re-contacted for research purposes and, if so, obtained from each parent study, individuals' contact information. From October 2019, I was joined by psychology placement student Emma Chubb and, throughout the study, we both remained blind to CNV carrier status. We sent individuals letters inviting them to take part in the study. They were then required to contact us to discuss their participation and to receive an information sheet. Once they had received this, there was at least a 48-hour delay before they could be booked in for their first assessment. Anyone contacted who did not wish to take part was removed from the recall list. If there was no response within 2 weeks, we attempted to contact them by telephone.

4.2.4 Statistical analyses

I carried out statistical analyses in SPSS.(212) For the comparison of group characteristics such as age and sex, I used chi square and t tests as appropriate. I compared average numbers of diagnoses and medications using Poisson regression, and scores on negative symptom assessments between CNV carriers and CNV noncarriers using t tests. I examined for associations between CNV carrier status and diagnosis, family history and developmental delay using logistic regression with age and sex as covariates.

4.2.5 Phenotypic assessments

4.2.5.1 Clinical phenotypes

Participants were asked their i) primary psychiatric diagnosis in their opinion, ii) primary psychiatric diagnosis according to their clinical team, and iii) presented with a list of psychiatric and physical health diagnoses and asked with which they had ever been diagnosed by a health professional. They were also asked which medications they were currently prescribed and, where possible, this was corroborated via their prescriptions. In order to establish a family history of psychiatric and physical health disorders, participants were presented with a list of diagnoses and asked with which a biological relative had ever been diagnosed. Numbers of first- and second-degree relatives with specific psychiatric and physical health diagnoses were recorded (Appendix 2).

We initially began attempting to assess participants' development using an adapted section A of the Development and Motor History Form.(213) However, it quickly became apparent that none of the participants were able to answer these questions and only one or two had parents available to help with the information. In the absence of appropriate brief validated assessment instruments, I made the decision to switch to asking participants a few brief questions related to developmental windows which reflect developmental histories taken in clinical services – i) birth weight, ii) whether they were born at term, early or late, iii) birth complications, iv) postnatal

complications, v) whether they walked at approximately the right time or late (definition provided), and vi) whether they talked at approximately the right time or late (definition provided). Participants were given the opportunity to elaborate when responding positively to these questions.

4.2.5.2 Assessments of the RDoc Positive Valence system

Clinical Assessment Interview for Negative Symptoms (CAINS) – this is a 13-item clinical rating scale which assesses five sub-domains of negative symptoms – i) asociality, ii) avolition, iii) anhedonia, iv) affective blunting and, v) alogia. Each item is rated from 0 – 4 (0 = no impairment, 1 = mild deficit, 2 = moderate deficit, 3 = moderately severe deficit, 4 = severe deficit). Factor analytic studies of the CAINS have demonstrated two factors reflecting impairments in i) experience – impaired motivation and enjoyment of activities and, ii) expression – impaired communication, both verbal and non-verbal.(214)

Temporal Experience of Pleasure Scale (TEPS) – this is an 18-item, questionnaire-based scale designed to measure experiences of pleasure. Participants are presented with a series of statements, for example “I look forward to a lot of things in my life” or “the smell of freshly cut grass is enjoyable to me”, and are required to indicate whether this statement is very false, moderately false, slightly false, slightly true, moderately true or very true for them. The TEPS consists of two subscales – a 10 item anticipatory pleasure scale and, an 8 item consummatory pleasure scale.(215)

Monetary Choice Questionnaire (MCQ) – this is a 27-item questionnaire which measures delayed reward discounting. Each statement offers two amounts of money, the second with a time delay e.g. “would you prefer $54 today, or $55 in 117 days?”. Participants are required to indicate whether they would prefer the smaller reward today or the larger reward in the specified number of days. The discounting rate was calculated for each participant using the geometric mean of the degree of discounting between the two questions which reflected when the participant switched between

choosing the delayed reward and the immediate reward. The higher the discounting rate, the higher the level of impulsivity.(216)

Behavioural Activation / Inhibition Scale (BAS / BIS) – this is a 24-item questionnaire. Participants are presented with a series of statements, for example “a person’s family is the most important thing in life” and are required to indicate whether this statement is very true, somewhat true, somewhat false or very false for them. The scale comprises of i) a behavioural activation component – 4 items assessing drive, 4 items assessing fun seeking and 5 items assessing reward responsiveness; ii) a behavioural inhibition component – 7 items; iii) 4 filler questions.(217)

Cambridge Gambling Task (CGT) – this task examines decision making and risk-taking behaviour outside a learning context and was done on an iPad Air 2. In this task, participants are presented with a row of 10 boxes across the top of the screen, some red, some blue, with the ratio of the two colours varying between stages (Figure 4.1).

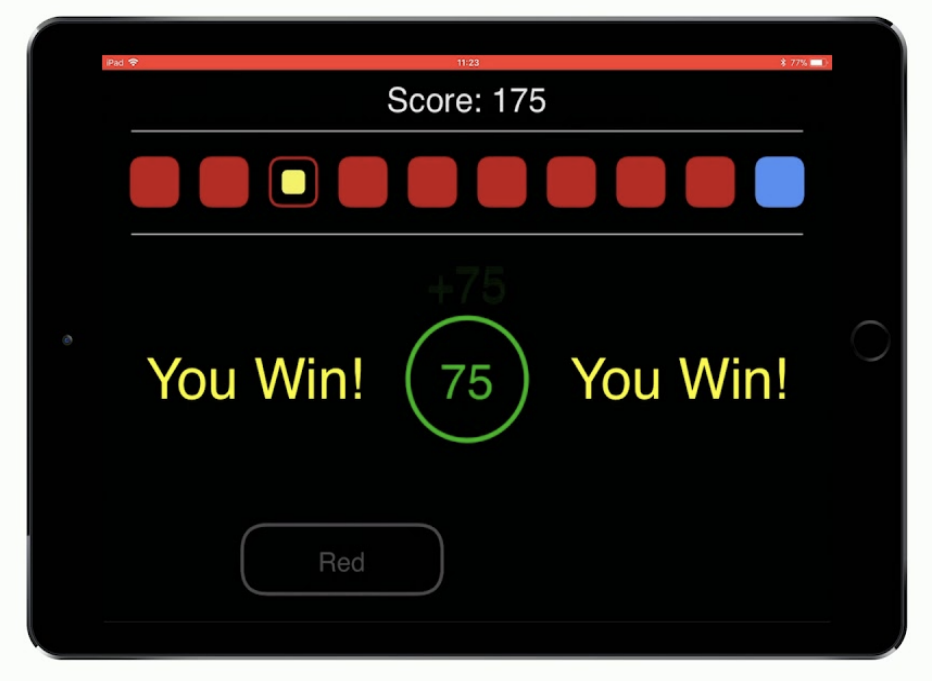

Figure 4.1. The CANTAB Cambridge Gambling Task.

In each round, a yellow token is in one of the boxes. Participants must first select whether they think the token will be in a red or blue box and then decide on a proportion of their points to bet on their decision. The current bet value is displayed in a circle in the centre of the screen. These points are then added or taken away from the total score depending on whether their decision is correct. I analysed data for the risk adjustment score merged. This variable measures the participant's ability to modify their choices on the number of points bet based on the probability of different outcomes. Normative data were not available from CANTAB and this task was not included in the original battery administered to my reference sample, so I used raw scores for analysis.

4.2.5.3 Assessments of the RDoC Cognitive Control systems

One Touch Stockings of Cambridge Task (OTS) – this task, based on the Tower of Hanoi task, examines the spatial planning and working memory subdomains of executive function and was done on an iPad Air 2. In this task, participants are presented with two displays containing three coloured balls each. The displays are presented such that they can be perceived as stacks of coloured balls held in stockings/socks suspended from a beam (Figure 4.2). There is a row of numbered boxes along the bottom of the screen. Participants must work out in their head how many moves it would require for the bottom display to be made to look identical to the top display. They must then select the appropriately numbered box. I analysed data for the number of problems solved on the participant's first choice. I used data from my reference sample to calculate Z scores.

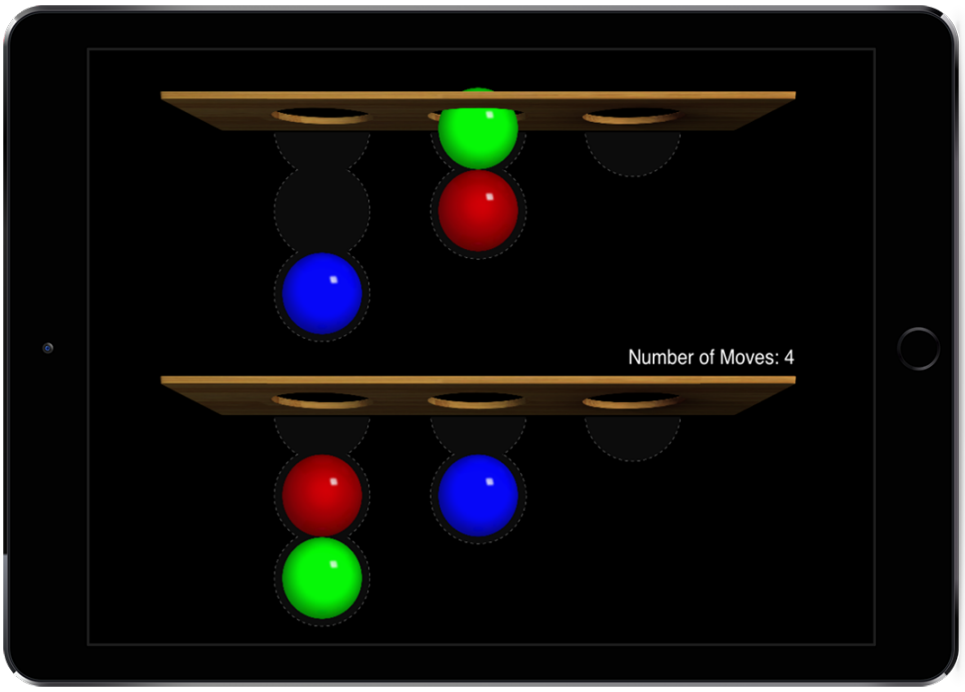

Figure 4.2. The CANTAB One Touch Stockings of Cambridge Task.

Stop Signal Task (SST) – this task examines response inhibition, which is a measure of impulse control, and was done on an iPad Air 2. In this task, participants are presented with a circle and two rectangular buttons on the screen (Figure 4.3). An arrow flashes on the screen and participants are required to press the button of the direction in which the arrow points. If an audio tone is played, they must withhold their response. The task adapts to the performance of the participant, narrowing in on a 50% success rate for inhibition. I analysed data on the stop signal reaction time – this is a covert measurement calculated from the length of time between the go stimulus and the stop stimulus at which the subject is able to successfully inhibit their response 50% of the time. Normative data were unavailable from CANTAB and this task was not included in the original battery administered to my reference sample, so I used raw scores for analysis.

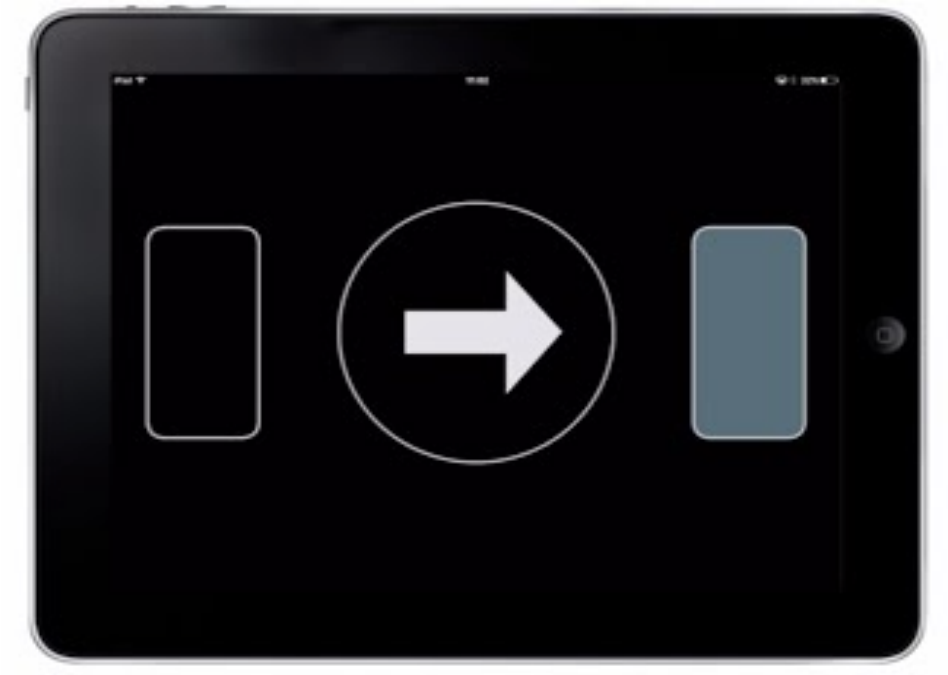

Figure 4.3. The CANTAB Stop Signal Task.

4.2.5.4 Cognitive tasks

CANTAB cognitive tasks were administered on an iPad Air 2 from Cambridge Cognition, from which results were automatically synced with a cloud-based platform. For each task, participants completed a training round to learn how the task worked. They then moved on to the assessed rounds.

Emotion Recognition Task (ERT) – this task examines the ability to identify six basic emotions in facial expressions along a continuum of expression magnitude. In this task, participants are shown computer-morphed images derived from the facial features of real individuals, each showing a specific emotion (Figure 4.4). Each face is displayed for 200 milliseconds and then immediately covered up to prevent residual processing of the image. Participants are required to select from six options which emotion the face displayed - sadness, happiness, fear, anger, disgust or surprise. I analysed data for the number of correct responses. I used data from my reference sample to calculate Z scores.

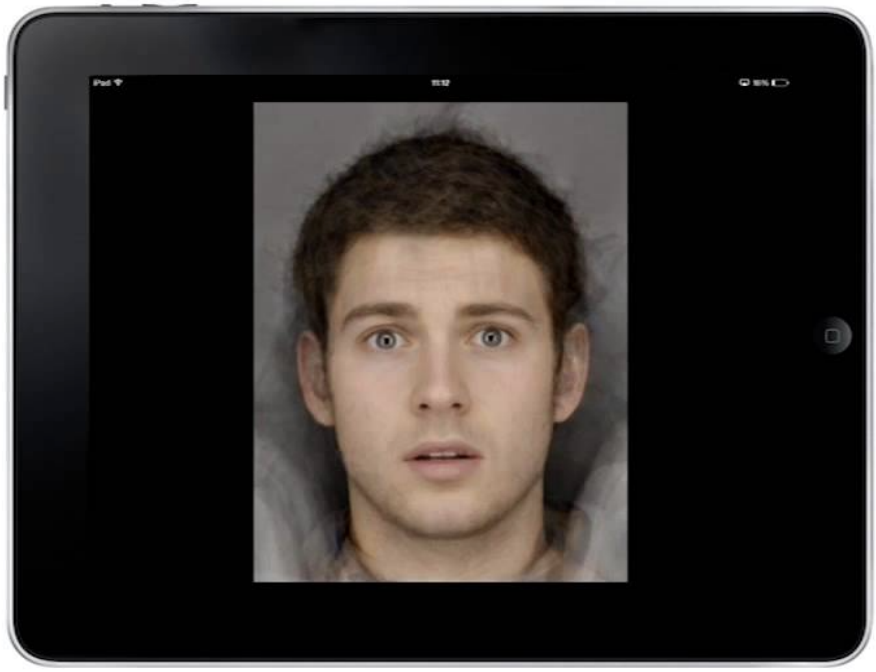

Figure 4.4. The CANTAB Emotion Recognition Task.

Reaction Time Task (RTT) – this task examines motor and mental response speeds. Participants are shown a screen with either one circle or five circles (simple or five-choice model rounds respectively). They are required to select and hold down a button at the bottom of the screen until a yellow dot appears in one of the circles. They are then required to release the button and select the circle in which the yellow dot appeared as quickly and accurately as possible (Figure 4.5). I analysed data for median reaction time on the five-choice task using data from my reference sample to calculate Z scores.

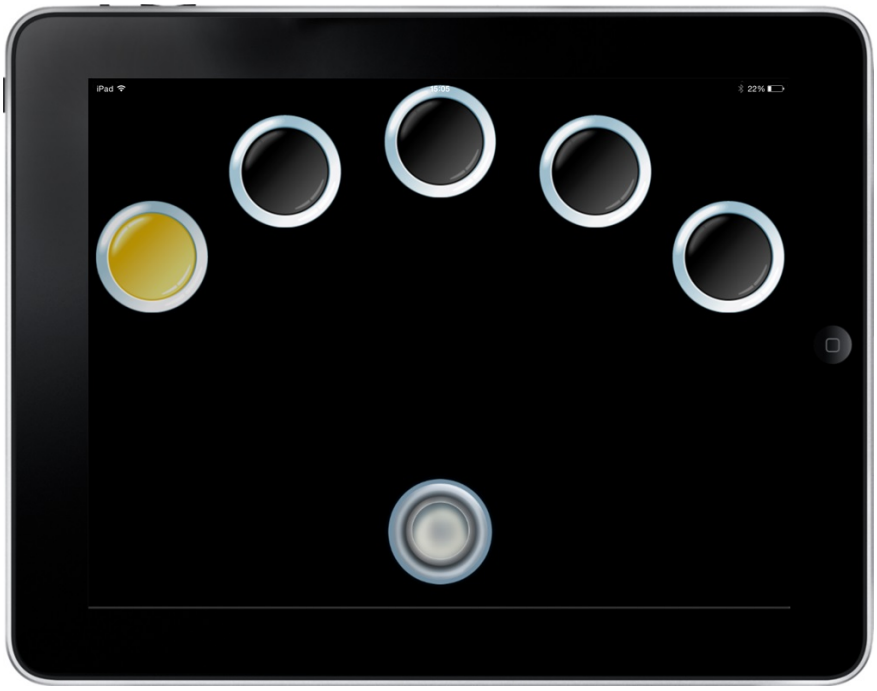

Figure 4.5. The CANTAB Reaction Time Task.

Spatial Working Memory Task (SWM) – this task examines executive function, providing measures of strategy and working memory errors. Participants are shown coloured boxes. The aim is to select the boxes and, using a process of elimination, locate yellow tokens and use them to fill up an empty column on the right-hand side of the screen (Figure 4.6). The number of boxes gradually increases. I analysed data on the number of errors made across all trials, using data from my reference sample to calculate Z scores.

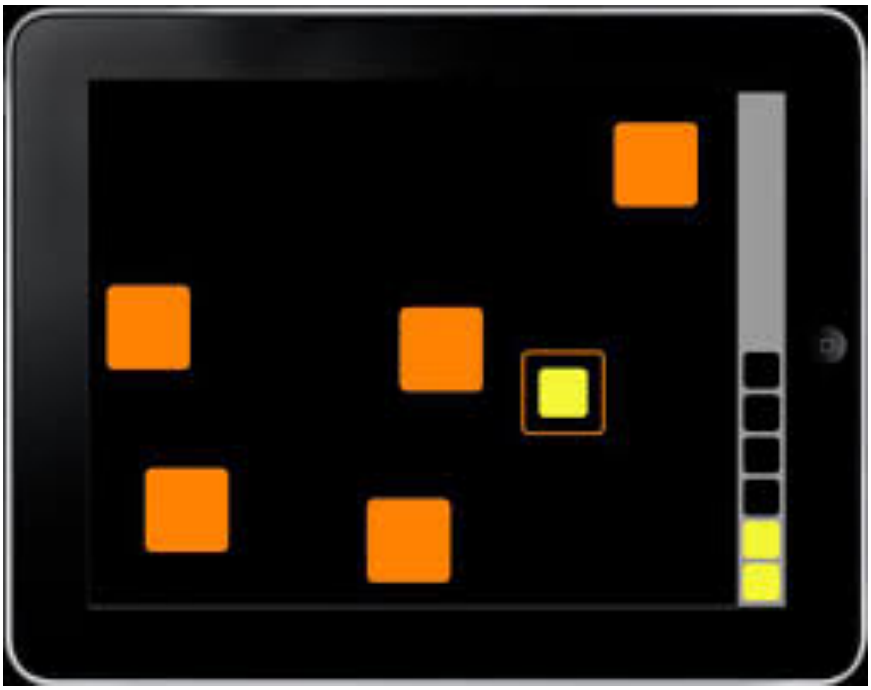

Figure 4.6. The CANTAB Spatial Working Memory Task.

Paired Associates Learning Task (PAL) – this task examines visual memory and new learning. Boxes containing various patterns are displayed on the screen. The boxes are opened at random and their patterns shown in the middle of the screen, one at a time (Figure 4.7). Participants are required to select the box in which each pattern was originally located. When they make an error, the boxes are opened again in their original sequence to remind participants of the locations of the patterns. I analysed data for total number of errors across all trials using data from my reference sample to calculate Z scores.

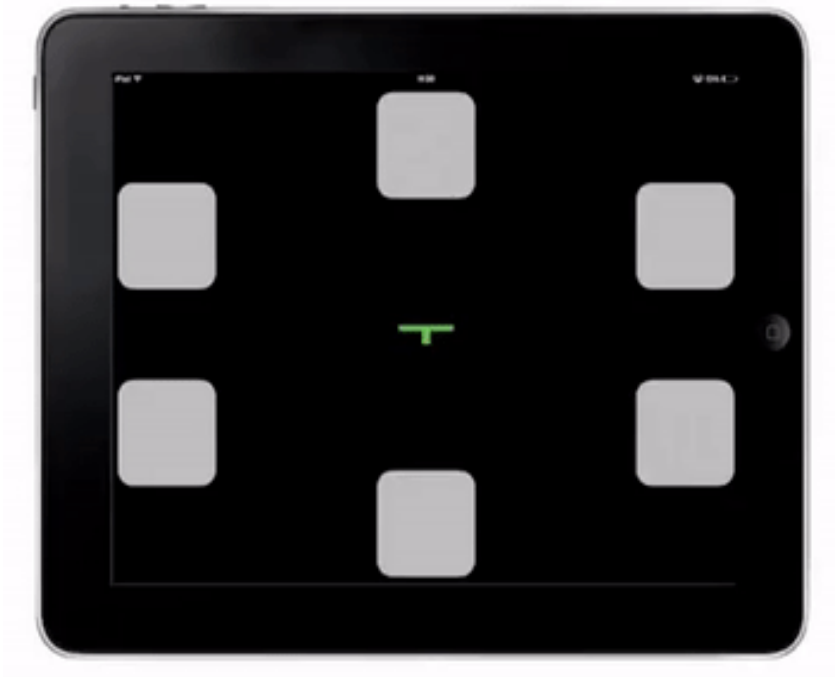

Figure 4.7. The CANTAB Paired Associates Learning Task.

Verbal Recognition Memory task (VRM) – this task examines verbal memory and new learning. Participants are shown a sequence of words on the screen one by one. They are then required to recall the words, while a rater marks which ones they remembered. In the second phase of the task, participants are presented with pairs of words with each pair consisting of one word from the original list and a distractor word (Figure 4.8). They are required to choose which word they saw previously in a forced choice paradigm. Following a delay, there is an additional recognition phase. I analysed data for free recall distinct stimuli – this is the total number of distinct words correctly recalled in the immediate free recall stage. I used data from my reference sample to calculate Z scores.

Figure 4.8. The CANTAB Verbal Recognition Memory Task.

4.3 Results

There were 24 individuals who carried one of the 54 neurodevelopmental CNVs in the Cardiff COGS sample (2.5% of 965) and 76 in the NCMH cohort (3.1% of 2,465). Taking into account a small degree of overlap between the samples and following the exclusion of individuals who i) had died, ii) were too unwell to take part, iii) failed risk assessment, iv) lived outside of Wales, 87 CNV carriers were eligible to take part in the study. We aimed to assess 20 CNV carriers and 40 CNV noncarriers and, by late March 2020, 5 CNV carriers and 16 CNV noncarriers had completed assessments at both recall sessions. A further 6 CNV noncarriers had completed assessments at the initial recall session only (Figure 4.9, Table 4.2). The study was suspended in early March 2020 due to the COVID19 pandemic because of lack of access to facilities, which we were not able to access again until after the end of my PhD fellowship in July 2020.

	CNV carriers n = 5	CNV noncarriers n = 22	p
Average age Age range	60.8 years 51 - 69 years	52.1 years 20 - 78 years	0.173
Sex	2 female 3 male	13 female 9 male	0.628
CNV genotype 15q11.2 deletion 16p11.2 distal deletion 16p13.11 duplication 22q11.2 deletion	 1 1 2 1		

Table 4.2. Characteristics of participants who took part in recall assessments. Statistical comparison was carried out using a t test for average age and a chi square test for sex.

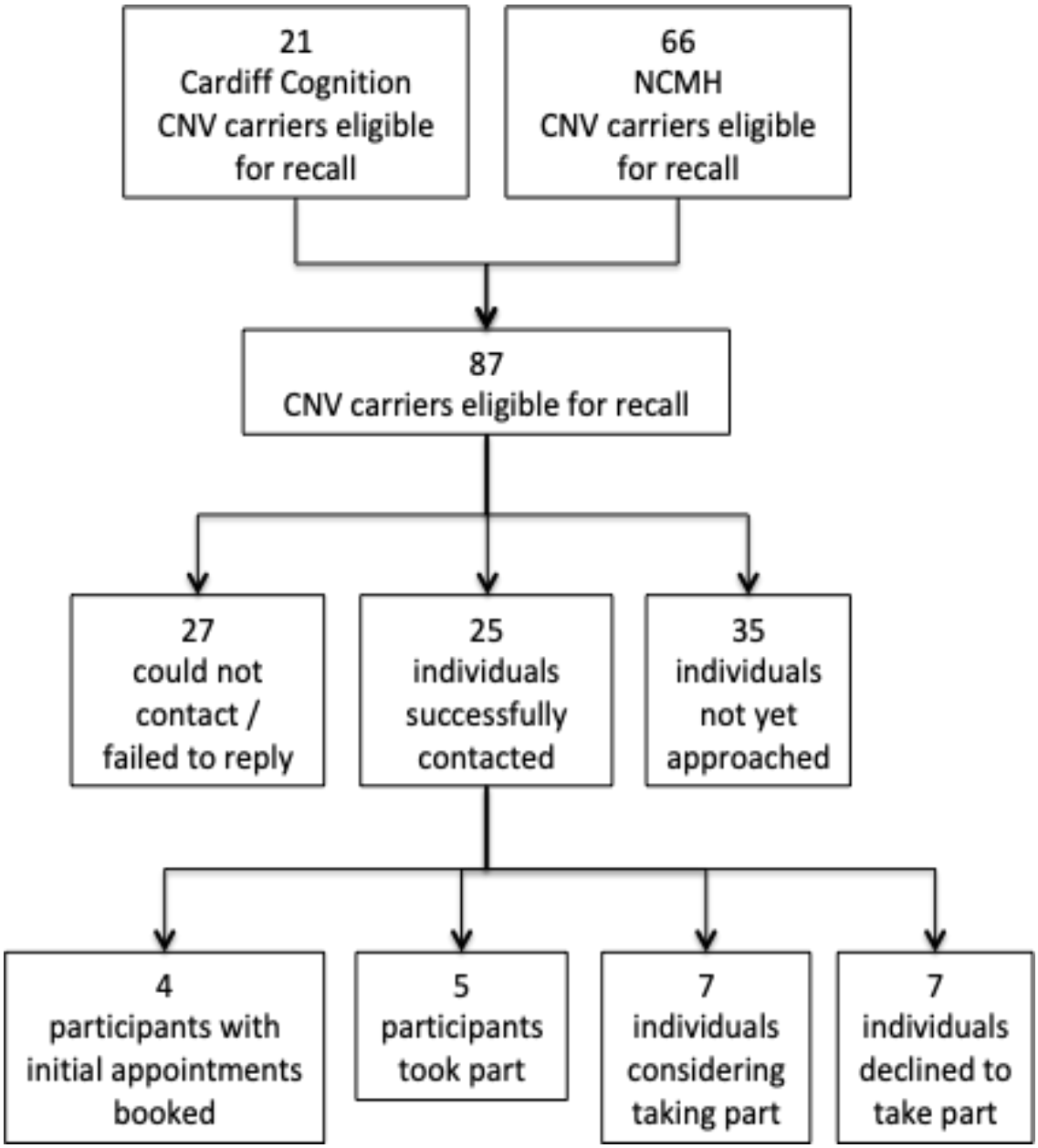

Figure 4.9. Progress in the recruitment of CNV carriers from the Cardiff COGS and NCMH cohorts as of late March 2020.

4.3.1 Clinical phenotypes

Only one CNV carrier, a participant with a diagnosis of velocardiofacial syndrome, was aware of their CNV carrier status. The remaining four CNV carriers were not aware that they carry a CNV. Displaying single counts in conjunction with phenotypic data in this group risks participants being able to identify themselves and establish their status as a CNV carrier. I have attempted to display data in such a way as to avoid this happening.

Diagnosis – we recorded self-reported psychiatric and physical health diagnoses at interview and were able to confirm these in health records for two CNV carriers and six CNV noncarriers. The primary psychiatric diagnoses of the five CNV carriers were autism spectrum disorder (ASD), schizophrenia, schizoaffective disorder and bipolar affective disorder (BPAD). For CNV noncarriers, primary psychiatric diagnoses were attention deficit hyperactivity disorder (ADHD, n = 1), depression (n = 12) and schizophrenia or other psychoses (n = 9). The main difference between the two groups was for a primary diagnosis of depression. All 5 CNV carriers had a diagnosis of depression but this was not primary for any of the five participants, whereas 54% (n = 12) of CNV noncarriers had a primary diagnosis of depression.

I examined existing data on self-reported diagnosis for the 76 CNV carriers and 2,389 CNV non-carrying individuals from the NCMH cohort. The characteristics of these groups are shown in Table 4.3. I compared the average number of psychiatric diagnoses between the two groups - CNV carriers reported a greater average number of psychiatric diagnoses than CNV noncarriers (3.5 vs 2.9). A Poisson regression, which included age and sex, showed that CNV carriers had 1.2 times the number of psychiatric diagnoses as CNV noncarriers (95% CI 1.06 – 1.36, p = 0.004). The range of psychiatric diagnoses in CNV carriers is shown in Figure 4.10.

	CNV carriers n = 76	CNV noncarriers n = 2,389	p
Average age Age range	54.6 15 – 97 years	54.6 12 – 103 years	0.274
Sex (% female)	56.7% (n = 43)	57.2% (n = 1366)	0.907

Table 4.3. Characteristics of participants from the NCMH cohort in whom CNVs were previously called and phenotypic data available. Statistical comparison was carried out using a t test for average age and a chi square test for sex.

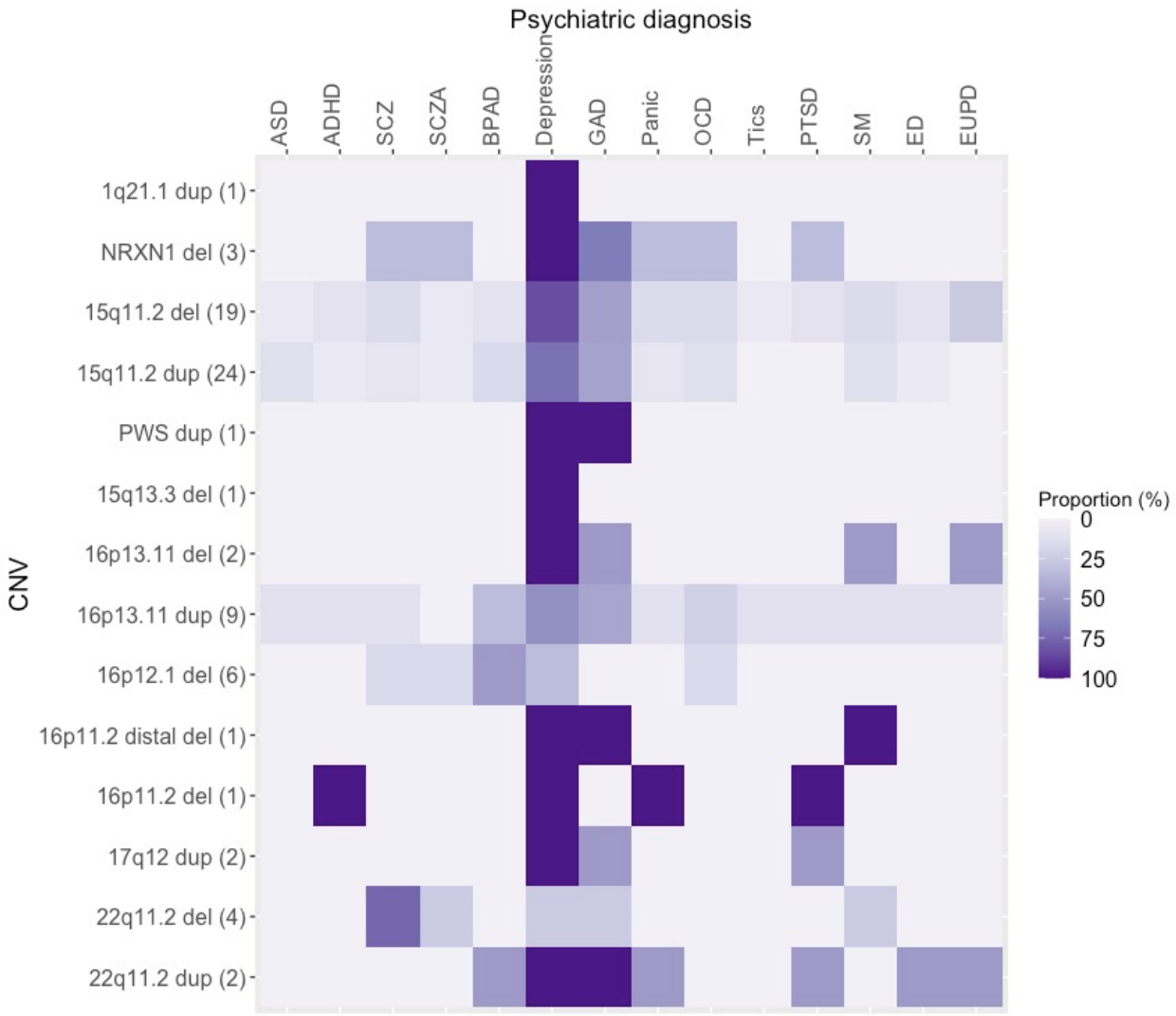

Figure 4.10. Psychiatric diagnoses in carriers of neurodevelopmental CNVs.

Proportion refers to the percentage of carriers of a specific CNV with a diagnosis. Number in parentheses after each CNV indicates the number of carriers.
del – deletion, dup – duplication, SCZ – schizophrenia, SCZA – schizoaffective disorder, GAD – generalised anxiety disorder, panic – panic disorder, tics – tic disorders, SM – substance misuse, ED – eating disorders, EUPD – emotionally unstable personality disorder.

All participants, except six CNV noncarriers had at least one physical health diagnosis but logistic regression analyses, including age and sex as covariates, revealed no statistically significant associations between CNV carrier status and physical health diagnosis (Table 4.4). In Table 4.4, there are two instances of a single count in the CNV carrier column and zero count in the CNV noncarrier column – this is where the CNV carrier is already aware of their carrier status. I compared the average number of physical health diagnoses between the two groups - CNV carriers reported an average of 5.0 physical health diagnoses compared to 2.9 in CNV noncarriers. A Poisson regression revealed that CNV carriers had 1.65 times the number of physical health diagnoses as CNV noncarriers (95% CI 1.007 – 2.716, p = 0.047).

Medications – we collected data on prescribed medications and compared the average numbers between groups using Poisson regression. There was not a significant difference in the average number of medications prescribed to CNV carriers and CNV noncarriers, regardless of whether this was for a psychiatric indication (1.0 vs 1.6, OR 0.57, 95% CI 0.21 – 1.51, p = 0.257) or physical health indication (2.0 vs 3.1, OR 0.52, 95% CI 0.26 – 1.04, p = 0.063).

Family history – we collected data on history of psychiatric or physical health disorders in participants' first- and second-degree relatives. 80% of CNV carriers (n = 4) reported a history of intellectual disability, ASD, ADHD or schizophrenia in at least one first- or second-degree relative compared to 27% (n = 6) of CNV noncarriers. However, there was not a statistically significant association between CNV carrier status and family history of neurodevelopmental disorder in logistic regression analyses with age and sex as covariates (OR 11.32, 95% CI 0.85 – 150.72, p = 0.066). 80% (n = 4) of CNV carriers reported a history of cancer in first- or second-degree relatives compared to 40% (n = 8) of CNV noncarriers but once again, there was not a statistically significant association between CNV carrier status and family history of cancer in logistic regression analyses (OR 10.01, 95% CI 0.71 – 140.72, p = 0.087).

	CNV carriers n = 5, % (n)	CNV noncarriers n = 22, % (n)	OR (95% CI)	p
Cerebrovascular accident	0	4.5 (1)	NA	0.999
Cataracts	20 (1)	0	NA	0.989
Asthma / chronic obstructive pulmonary disease	0	18.2 (4)	NA	0.999
Obstructive sleep apnoea	20 (1)	9.1 (2)	2.38 (0.14 – 41.79)	0.552
Structural heart problems or cardiomyopathy	40 (2)	4.5 (1)	15.89 (0.68 – 368.99)	0.085
Ischaemic heart disease	20 (1)	13.6 (3)	0.52 (0.03 – 11.07)	0.673
Hypertension	20 (1)	22.7 (5)	0.61 (0.05 – 7.76)	0.703
Hyperlipidaemia	40 (2)	45.5 (10)	0.54 (0.06 – 4.53)	0.570
Gastroesophageal reflux disease / gastric or duodenal ulcers	20 (1)	31.8 (7)	0.22 (0.01 – 4.23)	0.319
Type 2 diabetes mellitus	60 (3)	9.1 (2)	13.76 (0.98 – 192.79)	0.052
Autoimmune disorder	20 (1)	18.2 (4)	1.59 (0.10 – 24.49)	0.741
Osteoarthritis	20 (1)	18.2 (4)	0.98 (0.07 – 13.09)	0.989
Fibromyalgia	20 (1)	9.1 (2)	2.81 (0.09 – 81.91)	0.548
Cancer	20 (1)	0	NA	0.989

Table 4.4. Association analyses for physical health diagnoses and CNV carrier status. Logistic regression analyses included age and sex as covariates.

Development - 13 participants were able to state their birthweight (range 0.91kg to 4.25kg). Out of the 18 participants who were able to state when they were born, eight were born at term, eight were born early and two were born late. 10 participants stated that there were birth complications. 19 participants (3 CNV carriers, 16 CNV noncarriers) were able to comment on when they started walking. 33% (n = 1) of CNV carriers and 12.5% (n = 2) of CNV noncarriers reported walking late but there was not a statistically significant association between CNV carrier status and delays walking in logistic regression analyses (OR 9.78, 95% CI 0.27 – 360.71, p = 0.215). 20 participants (4 CNV carriers and 16 CNV noncarriers) were able to comment on when they started talking. 75% (n = 3) of CNV carriers and 18.8% (n = 3) of CNV noncarriers reported talking late. Logistic regression analyses examining for an association between CNV carrier status and talking late yielded a non-significant result (OR 17.62, 95% CI 0.90 – 344.54, p = 0.059).

4.3.2 Assessments of the RDoc Positive Valence system

We assessed negative symptoms using questionnaires and tasks mapped to the RDoC Positive Valence system. I compared mean scores across these measures between CNV carriers and CNV noncarriers using t-tests. There were no statistically significant differences in mean scores between the two groups, but some patterns were observed (Table 4.5).

Clinical Assessment Interview for Negative Symptoms (CAINS) – CNV carriers had higher mean scores across both the motivation and pleasure, and expression domains, reflecting greater deficits (Table 4.5). The former domain consists of social, work and school, and recreation sub-domains and the difference observed here was largely accounted for by CNV carriers having higher mean scores in the social sub-domain (mean score 6.8 vs 3.8, p – 0.116). *Temporal Experience of Pleasure Scale (TEPS*) – CNV carriers had lower mean scores on both the anticipatory and consummatory subscales, reflecting greater deficits. This difference was more marked for the anticipatory subscale, reflecting an apparent deficit in anticipating reward. *Monetary Choice Questionnaire (MCQ)* – CNV carriers tended to have a

higher mean discounting rate than CNV noncarriers, indicating higher levels of impulsivity. *Behavioural Activation Scale (BAS) drive subscale* – CNV carriers tended to have lower mean scores than CNV noncarriers, indicating lower levels of drive. *Cambridge Gambling Task (CGT)* – CNV carriers tended to have lower mean scores on the risk adjustment measure than CNV noncarriers. This reflects difficulties in adjusting the number of points bet based on the ratio of blue-red boxes.

Assessment	CNV carriers		CNV noncarriers		p
	n	Mean score (SD)	n	Mean score (SD)	
CAINS - motivation and pleasure	5	13.8 (10.6)	18	9.6 (9.5)	0.402
CAINS - expression	5	3.8 (4.4)	18	2.2 (3.4)	0.377
TEPS - anticipatory	5	30.2 (8.5)	21	37.4 (6.8)	0.054
TEPS - consummatory	5	34 (5.8)	21	36.7 (6.8)	0.423
MCQ - mean discounting rate	5	0.039 (0.042)	21	0.027 (0.057)	0.661
BAS - drive	5	9.2 (3.1)	20	11.2 (2.8)	0.178
CGT - risk adjustment score	5	-0.16 (0.32)	16	0.67 (1.14)	0.132

Table 4.5. A comparison of mean scores on measures of negative symptoms between CNV carriers and CNV noncarriers. Mean scores were compared using t-tests.

4.3.3 Assessments of the RDoC Cognitive Control systems

I examined the RDoc Cognitive Control systems using the CANTAB One Touch Stockings of Cambridge (OTS – goal selection and updating, representation and maintenance sub-constructs) and Stop Signal tasks (SST – response selection, and inhibition / suppression sub-constructs). In the OTS, CNV carriers tended to be less likely than CNV noncarriers to solve problems on their first choice. In the SST, the stop signal reaction time is calculated from the length of time between the go stimulus and the stop stimulus at which the subject is able to successfully inhibit their response 50% of the time. CNV carriers had an average score lower than CNV noncarriers (Table 4.6).

Assessment	CNV carriers		CNV noncarriers		p
	n	Mean score (SD)	n	Mean score (SD)	
CANTAB One Touch Stockings of Cambridge					
Problems solved on first choice (Z)	5	-2.25 (1.59)	15	-1.18 (1.32)	0.153
CANTAB Stop Signal Task					
Stop signal reaction time (ms)	5	266.9 (52.4)	15	312.5 (118.6)	0.421

Table 4.6. A comparison of mean scores on the problems solved on first choice and stop signal reaction time measures of the One Touch Stockings of Cambridge (OTS) and Stop Signal Tasks (SST) respectively. Data for the OTS are Z scores calculated using normative data from my reference sample. Normative data were not available for the SST so raw scores are compared. ms – milliseconds.

4.3.4 Cognitive tasks and related functional outcomes

21 participants attempted the cognitive tasks - 5 CNV carriers, 16 CNV noncarriers. Five participants terminated the testing battery early - three cited physical problems such as arm pain, one had poor eyesight and one just wanted to stop. A further six participants did not have a second assessment, where they would have had the opportunity to complete the CANTAB tasks, due to the study finishing early.

The tasks were completed by 3-4 CNV carriers and 13-15 CNV noncarriers. A comparison of mean results between the two groups using t-tests did not yield any statistically significant results but some patterns were observed. CNV carriers tended to perform more poorly than CNV noncarriers across most tasks - they got fewer correct answers on the ERT; had slower reaction times on the RTT; made a greater number of errors on the SWM and could recall fewer words on the VRM (Table 4.7).

I analysed education and occupation data, where available, for the wider NCMH sample (71 CNV carriers, 2239 - 2272 CNV noncarriers). CNV carriers were significantly less likely than CNV noncarriers to have achieved a degree, in logistic regression analyses which included age and sex as covariates (OR 0.40, 95% CI 0.19 – 0.82, p = 0.012) (Table 4.8). They also tended to be less likely to work in a job requiring longer training and more likely to be unemployed or to have never worked (Table 4.9).

Assessment	CNV carriers		CNV noncarriers		p
	n	Mean score (SD)	n	Mean score (SD)	
CANTAB Emotion Recognition Task					
Number of correct responses (Z)	4	-2.03 (1.36)	15	-1.26 (0.86)	0.173
CANTAB Reaction Time Task					
Five choice reaction time (Z)	4	0.77 (0.52)	13	0.35 (1.15)	0.501
CANTAB Spatial Working Memory Task					
Number of errors across all trials (Z)	3	0.62 (0.43)	13	0.47 (0.73)	0.741
CANTAB Paired Associates Learning Task					
Total number of errors (Z)	3	0.75 (0.83)	13	0.87 (1.32)	0.892
CANTAB Verbal Recognition Memory Task					
Free recall distinct stimuli (Z)	4	-1.39 (0.54)	15	-0.78 (0.56)	0.067

Table 4.7. A comparison of mean scores on the Emotion Recognition (ERT), Reaction Time (RTT), Spatial Working Memory (SWM), Paired Associates Learning (PAL) and Verbal Recognition Memory (VRM) tasks.

	CNV carriers n = 71 % (n)	CNV noncarriers n = 2,239 % (n)
Degree level or above	12.7 (9)	25.9 (580)
A levels or equivalent	29.6 (21)	27.3 (611)
GCSEs or equivalent	32.4 (23)	23.7 (530)
11+	5.6 (4)	1.6 (35)
No qualifications	19.7 (14)	21.6 (483)

Table 4.8. Highest qualifications in CNV carriers and CNV noncarriers in the wider NCMH sample.

	CNV carriers n = 71 % (n)	CNV noncarriers n = 2,272 % (n)
Professional	2.8 (2)	7.9 (181)
Legislator / senior official / manager	1.4 (1)	1.6 (36)
Health / educational professionals / business and public service associate professionals	2.8 (2)	5.9 (136)
Skilled agricultural workers	0	0.3 (6)
Administrative workers	1.4 (1)	3.9 (90)
Leisure, travel related occupations	0	0.04 (1)
Shop workers	11.3 (8)	9.8 (222)
Craft and related trade workers	2.8 (2)	1.4 (31)
Elementary occupations		1.5 (33)
Full time student	4.2 (3)	4.7 (106)
Home maker	2.8 (2)	1.7 (38)
Retired	16.9 (12)	24.7 (561)
Voluntary work	4.2 (3)	2.3 (52)
Unemployed / never worked	49.3 (35)	34.3 (779)

Table 4.9. Occupations in CNV carriers and CNV noncarriers from the wider NCMH sample.

4.3.5 Knowledge of CNV carrier status

During the process of providing information on the study and gaining consent, I informed participants that we selected individuals to approach based on their genetic results. I told them that we approached individuals with and without genetic variants known to be associated with psychiatric disorders, that I was blind to genetic variant status and that we planned on comparing findings between the two groups. None of the participants showed concern at this and none withdrew their participation. Several expressed a desire to know their results and I took the opportunity to explore their reasons, which fell under four broad categories – i) interest, ii) explanation of aetiology, iii) options for intervention and iv) family interests. *Interest* – several participants said they would like to know their genetic variant status because it is interesting, with one participant saying it would "be cool" to find out you have a genetic variant. *Explanation of aetiology* – some participants, who tended to have been more psychiatrically unwell than others in the sample, expressed a desire to know their genetic variant status to provide an explanation about why they had become so unwell. Such individuals had a tendency to ruminate on potential reasons they were unwell, engaged in self-blame, and expressed the idea that knowing it was "something genetic" would mean that it was not their fault. *Options for intervention* – several participants expressed the hope that psychiatric genetics research might be able to find treatment options that would be more effective and knowing their genetic variant status would allow them to try "better treatments". *Family interests* – some participants with family histories of psychiatric and neurodevelopmental disorders stated that finding out their genetic variant status might help affected family members understand the aetiology of their disorder, potentially leading to them being able to have genetic testing. Other participants expressed neutral opinions regarding finding out results. All of the participants expressed an interest in the study's findings and said they would like to receive a lay summary.

4.4 Discussion

In this project, I carried out psychiatric, cognitive and physical health/developmental assessments for 5 carriers of neurodevelopmental CNVs and 22 CNV noncarriers (16 assessments complete, 6 partially complete). Unfortunately, the COVID19 pandemic curtailed the study resulting in it being underpowered. However, trends emerged from the data collected which may be used to generate hypotheses for future work. *Psychiatry domain* - CNV carriers tended to have primary diagnoses from the neurodevelopmental spectrum, a finding also reflected in their family histories. This may reflect CNV carrying status in their family members but is something we cannot confirm in this sample. CNV carriers also seemed to attract a larger number of diagnostic labels than CNV noncarriers. In assessments of negative symptoms, CNV carriers tended to have lower levels of activity, drive and motivation, with the latter of these deficits being particularly pronounced for social activities. They had higher levels of impulsivity, and greater deficits in both anticipating rewards and adjusting their responses to rewards. *Cognition domain* - CNV carriers tended to perform more poorly than CNV noncarriers across most tasks and were less likely to perform highly in terms of qualifications and occupations. *Physical health/development* – CNV carriers had a greater number of physical health problems than CNV noncarriers. They were the only individuals to report a history of cataracts, congenital heart disease or cancer and tended to be more likely to report a history of obstructive sleep apnoea. They were also more likely to report delays in achieving the developmental milestones of walking and talking.

The trends identified suggest that individuals with neurodevelopmental CNVs have, perhaps unsurprisingly, a higher burden of neurodevelopmental phenotypes than their CNV non-carrying counterparts. They also begin to suggest the phenotypes which may be used in clinical practice to identify individuals who should be offered CNV testing. The trends for CNV carriers to have higher rates of developmental delay and neurodevelopmental disorders are in keeping with the results of a study which examined

phenotypic predictors of carrier status for a narrower range of schizophrenia-associated CNVs in individuals with schizophrenia or schizoaffective disorder.(218) Unfortunately, congenital disorders and other physical health problems were not included in that study. I plan on examining psychiatric and physical health variables in a much larger sample linked to electronic health records to establish more predictors of CNV carrier status that may be useful in clinical practice.

The trends observed for negative symptoms are potentially interesting. To my knowledge, no studies have examined negative symptoms in CNV carriers. Existing literature has reported associations between polygenic risk scores for schizophrenia and negative symptoms both in the general population (219), and individuals with psychosis (220), and a greater severity of negative symptoms in those with psychosis.(221) Parallels can also be drawn with findings in ASD. ASD has, as a core feature, deficits in the ability to initiate and sustain social interactions / communication.(120) These symptoms are not usually referred to as negative symptoms in ASD but there are clear similarities with multiple elements of the negative symptoms described in psychosis. RDoC proposed a reconceptualisation of ASD in a dimensional framework, and provided examples of how phenotypes such as failure to initiate conversation, lack of social smile, gaze avoidance and failure to develop functional language could be considered negative symptoms.(222) Associations between large, rare CNVs and case status for ASD are already established,(82,223) but CNV carrier status has also been reported to predict outcomes of social skills training in individuals with ASD.(224) These links between CNVs and elements of what could be considered negative symptoms are interesting, and provide added impetus to the idea that these links should be examined further in additional diagnoses. Collecting genetic data, and data on negative symptoms and related phenotypes in a large cross-disorder sample would allow us to establish 1) the nature and extent of negative symptoms, conceptualised dimensionally, across psychiatric and neurodevelopmental presentations and, 2) the nature of any association between genetic variants such as CNVs and individual elements of negative symptoms. For example, using mediation analysis via

structural equation modelling, we could establish whether the deficits observed in adjusting responses to reward are due to the direct effect of CNVs disrupting the physiological processes underlying these behaviours, or perhaps mediated by something else such as general cognitive impairment.

This project has a number of limitations. Informed by a priori power calculations, my initial target was to assess 20 CNV carriers and 40 CNV noncarriers and we were on course to achieve these numbers. At the time the study was stopped due to COVID19, I had reached over 50% of the target for CNV noncarriers and 25% of the target for CNV carriers. As a result, the project lacks the power to be able to statistically compare the two groups. This was clearly a problem that could not have been foreseen and I have tried to identify trends which it may be interesting to follow up in the future. I believe the discrepancy between the proportion of CNV carriers and noncarriers ascertained was because it was more difficult to recruit the CNV carriers. I revisited my recall records post-unblinding and found that CNV carriers were more likely to be uncontactable and the only individuals to state outright that they did not wish to participate. This I believe is in itself interesting, and something I need to take into account in my future work. As a result of the low number of CNV carriers assessed, some of the findings for physical health diagnosis rely on only one individual having the phenotype in question, limiting the amount to which these results can be interpreted. The CNV carriers assessed were not a homogeneous group - they had four different CNVs. This would limit the generalisability of results, but the trends identified may still be interesting to examine in larger samples.

The assessment of developmental milestones using self-report is not ideal. In this study, people frequently did not know when they achieved specific milestones. They seemed more likely to have this knowledge if their development was somehow abnormal, potentially skewing the data. One solution might be to seek this information from their parents. However, this approach is limited by a number of factors. For the individuals included in this study, parents were often not available. In addition, there is evidence that most people lack knowledge of the normal development of young

children.(225) Obtaining these data from linkage with electronic health records in psychiatry would be subject to the same problems, since they originate from self-report. An alternative, likely more reliable approach would be to obtain the information from childhood developmental records kept by health visitors.

CNVs may exert their effects by i) affecting specific physiological processes in the brain resulting in specific phenotypic findings/deficits, ii) having a generalised effect on brain networks by, for example, reducing resilience to environmental insults. The results of this study cannot be used to draw any conclusions about which is correct, and this is for multiple potential reasons. It is possible that the phenotypic measures used were not suitable to detect specific effects, but I do not believe this to be the case - many of these measures were chosen based on a dimensional approach. I believe that this was appropriate and the best option available at the time for maximising the ability to find specific effects. An alternative explanation is that the study was underpowered. That was certainly the case. The CNVs analysed here were selected because of their known associations and analysed as a group. This was necessary, due to their rarity, but not ideal. This group of CNVs is very heterogeneous – they vary in size and gene content, some result in loss of DNA and some in a gain, and they occur throughout the genome. I believe that it would actually be surprising to find specific phenotypic effects from such a group of variants. What is required is the collection of dimensional phenotypic data on a very large sample that allows the analysis of individual CNV loci for association with specific phenotypic effects, a monumental but important task.

Chapter 5 Discussion

The aim of this PhD was to establish the phenotypic spectrum of neuropsychiatric copy number variants (CNVs). I selected 54 CNVs based on their established associations with autism spectrum disorder (ASD), intellectual disability (ID) and developmental delay – termed neurodevelopmental CNVs. I chose to focus on the analysis of three phenotypic domains – psychiatry, cognition and physical health/development.

In Chapter 2, I examined data from ~82,000 to ~391,000 individuals in UK Biobank without a neurodevelopmental diagnosis for association between neurodevelopmental CNVs and i) the results of seven cognitive tasks, ii) educational attainment, and iii) occupational attainment. I showed that CNV carriers performed more poorly than CNV noncarriers by 0.13 – 0.45 standard deviations across all cognitive tasks. I established that CNV carriers were significantly less likely to finish in a higher qualifications group or to have a job requiring greater skills or longer training. Results for carriers of 12 CNVs associated with risk of schizophrenia and carriers of the other 41 neurodevelopmental CNVs were broadly equivalent.

In Chapter 3, I examined data from ~407,000 individuals in UK Biobank without a neurodevelopmental disorder or bipolar affective disorder (BPAD) diagnosis for association between neurodevelopmental CNVs and depression. I showed that CNV carriers have an increased risk of depression, whether this is defined on the basis of self-report, antidepressant prescription combined with self-report, or hospital diagnosis. I found three CNVs, duplications at 1q21.1, Prader Willi and 16p11.2 loci to be individually associated with self-reported depression at levels of statistical significance that survive correction for the number of tests. The association between neurodevelopmental CNVs and self-reported depression was partially explained by educational attainment, physical health, social deprivation, smoking and alcohol consumption.

In Chapter 4, I recalled carriers of neurodevelopmental CNVs and matched CNV noncarriers from existing psychiatric cohorts. This study was curtailed by the COVID19 pandemic, but I was able to conduct assessments for 5 CNV carriers and 22 CNV noncarriers. I was also able to analyse existing data from the wider NCMH cohort of 76 CNV carriers and 2,389 CNV noncarriers. CNV carriers reported, more frequently, primary diagnoses and family histories of neurodevelopmental disorders, and a greater number of psychiatric and physical health diagnoses. They also tended to be more likely to report delays in attaining the developmental milestones of walking and talking. In assessments of negative symptoms, the results for CNV carriers suggest i) greater deficits in motivation and pleasure, particularly for social activities, ii) greater deficits in anticipating rewards, iii) higher levels of impulsivity, iv) lower levels of drive, and v) difficulties adjusting responses to changes in the probability of reward. I had originally hypothesised that neurodevelopmental CNVs would exert a generalised effect across all measures of negative symptoms and that CNV carriers would have a higher frequency of neurodevelopmental disorders and a higher rate of physical health problems. However, due to lack of statistical power I was unable to reject the null hypothesis for any of these.

The results of analyses of cognition and related functional outcomes I report in Chapter 2 are consistent with those reported by two existing studies. Stefansson et al from deCODE Genetics, Iceland examined the impact of schizophrenia/autism associated CNVs on cognitive tasks and reported the same pattern of results – CNV carriers performed more poorly than CNV noncarriers but not as poorly as individuals with schizophrenia.(173) Männik et al examined the impact of CNVs associated with known syndromes on unselected populations from Estonia, Italy, USA and the UK. They reported a significant association between CNV carrier status and lower educational attainment.(174)

As discussed in Chapter 3, the literature on CNVs in depression has historically been less clear. However, a study published around the time of the work from Chapter 3 implicated an increased burden of short deletions in

depression, lending weight to the suggestion that rare CNVs play at least some role in risk of the disorder.(74) Depression is phenotypically very heterogeneous.(77) It is possible that neurodevelopmental CNVs exert their impact on overall depression risk by predisposing individuals to a neurodevelopmental subtype of the disorder. A study by Rice et al reported an illness trajectory, in 9% of adolescents with depression, characterised by onset in early adolescence, associations with childhood ADHD and neurodevelopmental traits, and associations with polygenic risk scores for major depressive disorder, schizophrenia and ADHD.(226) The authors did not examine CNVs in that study and my analysis for association between CNV carrier status and age of onset of depression in UK Biobank did not reach statistical significance. However, this variable may be subject to recall bias in CNV carriers. Participants who completed this questionnaire were 45-82 years old at the time of assessment and were required to recall the age of onset for a disorder which has its average onset in the mid 20s.(139,227) As I, and other authors, have shown, CNV carriers have impaired memory.(156,157,163) If this extends to autobiographical memory, it may mean that the recall of disorder related features, such as age of onset, is impaired in CNV carriers. It would be interesting to investigate whether neurodevelopmental CNVs, and other rare CNVs outside this group, play a role in a subtype of depression with an earlier age of onset, and psychiatric and physical comorbidities suggesting a greater degree of neurodevelopmental loading.

The analyses from my recall study in Chapter 4 are underpowered and it would be inappropriate to draw firm conclusions from the results. However, I feel they are informative for the purpose of generating specific hypotheses for further study. A lot of the results from this chapter are also new – to my knowledge no one has described the effects of neurodevelopmental CNVs on individuals with psychiatric disorders beyond associations with case status. One small study of sib pairs (total 20 cases) reported an association between having a greater number of CNV-disrupted genes and more severe negative symptoms in schizophrenia.(228) The findings regarding delayed milestones make sense when we consider that these CNVs predispose

individuals to disorders characterised by varying degrees of altered neurodevelopment – i.e. ASD, ID and ADHD.

The findings of this PhD fit with the concept of the neurodevelopmental spectrum and the existing disease threshold model. As we have seen in UK Biobank, it is possible to carry a neurodevelopmental CNV and to not develop a neurodevelopmental disorder. These individuals fall on the less severe end of the phenotypic spectrum, although it would be incorrect to suggest that they are phenotypically unaffected. Equally, individuals with the very same CNV can be found at the opposite end of the spectrum having developed schizophrenia, ASD or even severe intellectual disability. This brings about the question of, if not solely neurodevelopmental CNVs, what exactly determines the range of outcomes in those with CNVs? How can there be such a degree of difference in how two individuals with the same CNV are affected? Existing literature, and the neurodevelopmental spectrum and disease threshold models suggest that the difference is likely to come from i) common genetic variation, ii) second hit CNVs and other rare variants, iii) environmental factors and iv) stochastic factors. *Common variation* - individuals with pathogenic CNVs and schizophrenia have been shown have a higher polygenic risk score (PRS) for schizophrenia compared to controls (99), with another study reporting an interactive effect of PRS and CNVs in the disorder.(229) *Second hit CNVs and other rare variants* – evidence from work on 16p12.1 deletions has suggested that this CNV predisposes to neuropsychiatric phenotypes but also exacerbates these phenotypes in the presence of other large, rare CNVs (230); another study reported an association between deleterious variants and neurodevelopmental phenotypes in 16p12.1 deletion carriers.(231) *Environmental factors* - Mazina et al reported interactive effects between CNV carrier status and maternal infection during pregnancy on autism symptoms.(232) How these factors work together is unclear but the development of models which include individuals' CNVs, other rare variants, common variation and environmental factors will be key to understanding how these phenotypes occur and may help us to understand the underlying biology in greater detail.

This PhD project has several limitations. The UK Biobank sample is not representative of the general population – there is evidence of "healthy volunteer" selection bias.(233,234) Diagnoses used for exclusions and for the depression analyses in Chapter 3 relied on self-report, which is subject to information bias.(203) The wording used by UK Biobank should have mitigated against this – "in the touch screen you selected that you have been told by a doctor that you have other serious illnesses or disabilities, could you now tell me what they are?". In addition, our analyses suggested an under reporting of psychiatric diagnoses in this variable, which would have the effect of diluting associations rather than inflating them. I also repeated my depression analyses using alternative, more conservative variables with the same results. The recall study was underpowered due to the curtailment of the study because of the COVID19 pandemic. Clearly, I could do little about this. I attempted to increase power by analysing existing data where possible and these analyses will be used to generate hypotheses for future work.

My work has filled in some of the gaps of the neurodevelopmental CNV phenotypic spectrum and begun to explore how this information may be translated to the clinic. However, there remains a dearth of evidence on the effect of rare CNVs on individuals with psychiatric disorders. Future work should aim to further establish the range of phenotypes associated with these CNVs and also aim to address what other factors determine phenotypic outcomes. Key to translation to the clinic will also be the examination of the long-term impact of these CNVs and work around genetic counselling. The identification of biological mechanisms underlying psychiatric disorders and of new drug targets would be facilitated by the examination of CNVs i) in very large samples – this would allow the study of CNVs at individual loci rather than in heterogeneous groups, and ii) using a dimensional approach to phenotypic classification. However, this is clearly a far from straightforward task.

My future plans are to develop a project aimed at translating CNV and other rare variant results to the clinic. *Establishing the effects of rare variants over*

time – I plan on examining longitudinal data from large samples linked to electronic health records to i) establish which phenotypic variables predict rare variant carrier status and, ii) establish how rare variants affect individuals with psychiatric disorders over time. *Translation to the clinic* – I plan on developing a research clinic for individuals with rare variants and psychiatric disorders where I will collect detailed longitudinal phenotypic data which may be used to establish how best to improve the health of these individuals.

Appendix 1 – Neurodevelopmental CNVs

The original list of 93 CNVs identified by Coe et al, their genomic coordinates and calling criteria. N – number of CNV carriers in the UK Biobank sample.(196) CNVs associated with ASD/ID/DD at p values of < 0.05 are indicated in the column 'ND54' by the number 1. del – deletion, dup – duplication, chr – chromosome. At the *EHMT1* and *SHANK3* loci, small CNVs were found to be common in samples with poor QC criteria. Due to the likelihood of small CNVs at these telomeric loci being false positives, CNVs at these loci were required to intersect the loci by at least 1Mbp.

CNV	Critical/Unique Sequence Region (hg19)	CNV Calling Rules	N	ND 54
1p36 del (*GABRD*)	chr1:0-2,500,000	Size >50% of critical region, affecting *GABRD*	0	1
1p36 dup (*GABRD*)	chr1:0-2,500,000	Size >50% of critical region, affecting *GABRD*	0	1
TAR del	chr1:145,394,955-145,807,817	Size >50% of critical region	75	1
TAR dup	chr1:145,394,955-145,807,817	Size >50% of critical region	436	1
1q21.1 del	chr1:146,527,987-147,394,444	Size >50% of critical region	113	1
1q21.1 dup	chr1:146,527,987-147,394,444	Size >50% of critical region	177	1
NRXN1 del	chr2:50,145,643-51,259,674	Exonic deletions	163	1
2q11.2 del (*LMAN2L, ARID5A*)	chr2:96,742,409-97,677,516	Size >50% of critical region, affecting both *LMAN2L & ARID5A*	31	1
2q11.2 dup (*LMAN2L,ARID5A*)	chr2:96,742,409-97,677,516	Size >50% of critical region, affecting both *LMAN2L & ARID5A*	29	
2q13 del (*NPHP1*)	chr2:110,862,716-110,983,948	Size >50% of critical region, affecting *NPHP1*	2448	
2q13 dup (*NPHP1*)	chr2:110,862,716-110,983,948	Size >50% of critical region, affecting *NPHP1*	1976	

2q13 del	chr2:111,394,040-112,012,649	Size >50% of critical region	53	1
2q13 dup	chr2:111,394,040-112,012,649	Size >50% of critical region	71	1
2q21.1 del	chr2:131,481,308-131,930,677	Size >50% of critical region	41	
2q21.1 dup	chr2:131,481,308-131,930,677	Size >50% of critical region	59	
2q37 del (*HDAC4*)	chr2:239,716,679-243,199,373	Size >50% of critical region, affecting *HDAC4*	0	1
2q37 dup (*HDAC4*)	chr2:239,716,679-243,199,373	Size >50% of critical region, affecting *HDAC4*	0	
3q29 del	chr3:195,720,167-197,354,826	Size >50% of critical region	9	1
3q29 dup	chr3:195,720,167-197,354,826	Size >50% of critical region	5	
Wolf-Hirschhorn del	chr4:1,552,030-2,091,303	Size >50% of critical region	0	1
Wolf-Hirschhorn dup	chr4:1,552,030-2,091,303	Size >50% of critical region	0	1
Sotos syndrome del	chr5:175,720,924-177,052,594	Size >50% of critical region	0	1
5q35 dup	chr5:175,720,924-177,052,594	Size >50% of critical region	0	
6q16 del (*SIM1*)	chr6:100,836,750-100,911,811	Exonic deletions	0	
6q16 dup (*SIM1*)	chr6:100,836,750-100,911,811	Whole gene duplications	0	
Williams-Beuren syndrome (WBS) del	chr7:72,744,915-74,142,892	Size >50% of critical region	0	1
WBS dup	chr7:72,744,915-74,142,892	Size >50% of critical region	14	1
7q11.23 distal del (1.2-Mb)	chr7:75,138,294-76,064,412	Size >50% of critical region	0	
7q11.23 distal dup (1.2-Mb)	chr7:75,138,294-76,064,412	Size >50% of critical region	24	
8p23.1 del	chr8:8,098,990-11,872,558	At least 1Mbp of critical region	0	1
8p23.1 dup	chr8:8,098,990-11,872,558	At least 1Mbp of critical region	6	1
9q34 del (*EHMT1*)	chr9:140,513,444-140,730,578	At least 1Mbp, including *EHMT1*	0	
9q34 dup (*EHMT1*)	chr9:140,513,444-140,730,578	At least 1Mbp, including *EHMT1*	0	1
10q11.21q11.23 del	chr10:49,390,199-51,058,796	Size >50% of critical region	57	
10q11.21q11.23 dup	chr10:49,390,199-51,058,796	Size >50% of critical region	43	

10q23 del (*NRG3, GRID1*)	chr10:82,045,472-88,931,651	At least 1Mbp, including *NRG3* & *GRID1*	0	1
10q23 dup (*NRG3, GRID1*)	chr10:82,045,472-88,931,651	At least 1Mbp, including *NRG3* & *GRID1*	7	
Potocki-Shaffer syndrome del (*EXT2*)	chr11:43,940,000-46,020,000	Size >50% of critical region, including *EXT2*	0	1
11p11.2 dup (*EXT2*)	chr11:43,940,000-46,020,000	Size >50% of critical region, including *EXT2*	0	
13q12 del (*CRYL1*)	chr13:20,977,806-21,100,012	Exonic deletions	379	
13q12 dup (*CRYL1*)	chr13:20,977,806-21,100,012	Whole gene duplications	10	
13q12.12 del	chr13:23,555,358-24,884,622	Size >50% of critical region	85	
13q12.12 dup	chr13:23,555,358-24,884,622	Size >50% of critical region	236	
15q11.2 del BP1-BP2	chr15:22,805,313-23,094,530	Size >50% of critical region	1664	1
15q11.2 dup BP1-BP2	chr15:22,805,313-23,094,530	Size >50% of critical region	2041	1
Prader-Willi syndrome/Angelman syndrome (PWS/AS) del	chr15:22,805,313-28,390,339	Full critical region, ~4Mbp	0	1
PWS/AS dup	chr15:22,805,313-28,390,339	Full critical region, ~4Mbp	19	1
15q11q13 del BP3-BP4 (*APBA2, TJP1*)	chr15:29,161,368-30,375,967	Size >50% of critical region	16	
15q11q13 dup BP3-BP4 (*APBA2, TJP1*)	chr15:29,161,368-30,375,967	Size >50% of critical region	53	
15q11q13 del BP3-BP5	chr15:29,161,368-32462776	Size >50% of critical region	0	
15q11q13 dup BP3-BP5	chr15:29,161,368-32462776	Size >50% of critical region	9	
15q13.3 del BP4-BP5	chr15:31,080,645-32,462,776	Size >50% of critical region	42	1
15q13.3 dup BP4-BP5	chr15:31,080,645-32,462,776	Size >50% of critical region	240	
15q13.3 del (*CHRNA7*)	chr15:32,017,070-32,453,068	Size >50% of critical region, affecting *CHRNA7*	10	
15q13.3 dup (*CHRNA7*)	chr15:32,017,070-32,453,068	Size >50% of critical region, affecting *CHRNA7*	3031	
15q24 del	chr15:72,900,171-78,151,253	At least 1Mbp between the A-E intervals	0	1
15q24 dup	chr15:72,900,171-78,151,253	At least 1Mbp between the A-E intervals	9	1

15q25 del	chr15:83,219,735-85,722,039	At least 1Mbp between the A-D intervals	0	1
15q25 dup	chr15:83,219,735-85,722,039	At least 1Mbp between the A-D intervals	0	
Rubinstein-Taybi del (*CREBBP*)	chr16:3,775,056-3,930,121	Exonic deletions	0	
Rubinstein-Taybi dup (*CREBBP*)	chr16:3,775,056-3,930,121	Whole gene duplications	0	
16p13.11 del	chr16:15,511,655-16,293,689	Size >50% of critical region	131	1
16p13.11 dup	chr16:15,511,655-16,293,689	Size >50% of critical region	828	1
16p12.2-p11.2 del (7.1-8.7Mb)	chr16:21,596,415-28,347,808	Size >50% of critical region	0	
16p12.2-p11.2 dup (7.1-8.7Mb)	chr16:21,596,415-28,347,808	Size >50% of critical region	0	
16p12.1 del (520kb)	chr16:21,950,135-22,431,889	Size >50% of critical region	246	1
16p12.1 dup (520kb)	chr16:21,950,135-22,431,889	Size >50% of critical region	202	
16p11.2 distal del (220kb)	chr16:28,823,196-29,046,783	Size >50% of critical region	58	1
16p11.2 distal dup (220kb)	chr16:28,823,196-29,046,783	Size >50% of critical region	137	1
16p11.2 del (593kb)	chr16:29,650,840-30,200,773	Size >50% of critical region	110	1
16p11.2 dup (593kb)	chr16:29,650,840-30,200,773	Size >50% of critical region	138	1
17p13.3 del (*YWHAE*)	chr17:1,247,834-1,303,556	Exonic deletions	0	1
17p13.3 dup (*YWHAE*)	chr17:1,247,834-1,303,556	Whole gene duplications	0	1
17p13.3 del (*PAFAH1B1*)	chr17:2,496,923-2,588,909	Exonic deletions	0	1
17p13.3 dup (*PAFAH1B1*)	chr17:2,496,923-2,588,909	Whole gene duplications	0	1
Hereditary Neuropathy with Pressure Palsies del (HNPP)	chr17:14,141,387-15,426,961	Size >50% of critical region, affecting *PMP22*	237	
Charcot-Marie-Tooth disease type 1A dup (CMT1A)	chr17:14,141,387-15,426,961	Size >50% of critical region, affecting *PMP22*	124	
Smith-Magenis syndrome del	chr17:16,812,771-20,211,017	Size >50% of critical region	0	1
Potocki-Lupski syndrome dup	chr17:16,812,771-20,211,017	Size >50% of critical region	5	1
17q11.2 del (*NF1*)	chr17:29,107,491-30,265,075	Size >50% of critical region, affecting *NF1*	19	1

17q11.2 dup (*NF1*)	chr17:29,107,491-30,265,075	Size >50% of critical region, affecting *NF1*	0	1
Renal cysts and diabetes syndrome del (RCAD)	chr17:34,815,904-36,217,432	Size >50% of critical region	9	1
17q12 dup	chr17:34,815,904-36,217,432	Size >50% of critical region	101	1
17q21.31 del	chr17:43,705,356-44,164,691	Size >50% of critical region	0	1
17q21.31 dup	chr17:43,705,356-44,164,691	Size >50% of critical region	0	
17q23.1q23.2 del	chr17:58,302,389-60,289,141	Size >50% of critical region	0	
17q23.1q23.2 dup	chr17:58,302,389-60,289,141	Size >50% of critical region	0	
22q11.2 del	chr22:19,037,332-21,466,726	Size >50% of critical region	10	1
22q11.2 dup	chr22:19,037,332-21,466,726	Size >50% of critical region	280	1
22q11.2 distal del	chr22:21,920,127-23,653,646	Size >50% of critical region	5	1
22q11.2 distal dup	chr22:21,920,127-23,653,646	Size >50% of critical region	26	1
SHANK3 del	chr22:51,113,070-51,171,640	At least 1Mbp, including *SHANK3*	0	1
SHANK3 dup	chr22:51,113,070-51,171,640	At least 1Mbp, including *SHANK3*	0	1

Appendix 2 – Clinical and cognitive effects of risk factors for mental illness assessment booklet

Community Visit

Study Introduction

This study aims to help us understand more about how genetic changes influence measures known to be different between people with mental illness (such as psychosis) and people without mental illnesses.

Look through study information sheet – any questions?
Complete consent form.

Arrangements for HEB visit

Book in a HEB assessment session with the participant (check a room is available in the clinic and book it). Give the participant an appointment card. Check transportation arrangements – do we need to arrange a taxi? Go over procedure for reimbursing travel expenses and sending out reimbursement for time.

Education and Work

Pre-fill prior to assessment and check data is still correct with participant.

Highest qualification recorded last time (LT) _____

Still correct? Yes No

If no, correct response _____

Highest occupation recorded LT _____

Still correct? Yes No

If no, correct response _____

Mother's highest qualification recorded LT _____

Still correct? Yes No

If no, correct response _____

Mother's highest occupation recorded LT _____

Still correct? Yes No

If no, correct response _____

Father's highest qualification recorded LT _____

Still correct? Yes No

If no, correct response _____

Father's highest occupation recorded LT _____

Still correct? Yes No

If no, correct response _____

Highest Qualification

No qualifications': No academic or professional qualifications.

'1-4 GCSEs or equivalent':1-4 O Levels/CSE/GCSEs (any grades), Entry Level, Foundation Diploma, NVQ level 1, Foundation GNVQ, Basic/Essential Skills.

'5+ GCSEs or equivalent': 5+ O Level (Passes)/CSEs (Grade 1)/ GCSEs (Grades A*-C), School Certificate, 1 A Level/ 2-3 AS Levels/VCEs, Intermediate/Higher Diploma, Welsh Baccalaureate, Intermediate Diploma, NVQ level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First/General, Diploma, RSA Diploma.

'Apprenticeship': Apprenticeship.

'2+ A Levels or equivalent' (Level 3 qualifications): 2+ A Levels/VCEs, 4+ AS Levels, Higher School Certificate, Progression/Advanced Diploma, Welsh Baccalaureate Advanced Diploma, NVQ Level 3; Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma.

'Degree level or above' (Level 4 qualifications and above): Degree (for example BA, BSc), Higher Degree (for example MA, PhD, PGCE), NVQ Level 4-5, HNC, HND, RSA Higher, Diploma, BTEC Higher level, Foundation degree (NI), Professional qualifications (for example teaching, nursing, accountancy).

'Other qualifications': Vocational/Work-related Qualifications, Foreign Qualifications/ Qualifications gained outside the UK (NI) (Not stated/level unknown)

Highest Lifetime Occupation

Corporate managers and directors

Science, research, engineering and technology professionals

Health professionals

Teaching and educational professionals

Business, media and public service professionals

Other managers and proprietors

Science, engineering and technology associate professionals

Health and social care associate professional

Protective service occupations

Culture, media and sports occupations

Business and public service associate professionals

Skilled agricultural and related trades

Skilled metal, electrical and electronic trades

Skilled construction and building trades

Textiles, printing and other skilled trades

Administrative occupations

Secretarial and related occupations

Caring personal service occupations

Leisure, travel and related personal service occupations

Sales occupations

Customer service occupations

Process, plant and machine operatives

Transport and mobile machine drivers and operatives

Elementary trades and related occupations

Elementary administration and service occupations

Participant's Medical History

Primary diagnosis (participant's opinion) _____

Primary diagnosis _____

(opinion of clinical team as reported by participant)

Psychiatric Diagnoses

Has a health professional ever told you that you have any of the following diagnoses? (tick all that apply):

ADHD		Anorexia nervosa	
Autism		Bulimia nervosa	
Asperger's or other ASD		Obsessive compulsive disorder (OCD)	
Dyslexia		Agoraphobia	
Dyspraxia		Panic disorder	
Conduct disorder		Phobias	
Oppositional Defiant Disorder		Anxiety	
Tic disorder		Borderline personality disorder	
Tourette's disorder		Other personality disorder	
Intellectual disability		Post-traumatic stress disorder	
Depression		Alzheimer's disease	
Bipolar disorder / manic depression		Other dementia	
Mania / hypomania		Alcohol abuse / misuse	
Schizoaffective disorder		Other substance abuse / misuse _____	
Psychosis		Genetic syndrome (e.g. VCFS) _____	
Schizophrenia		Self harm or suicide attempts	
Postnatal psychosis		Other _____	
Postnatal depression			

Physical Health Diagnoses

Have you ever been told by a health professional that you have? (tick all that apply)

Heart disease		Hernia	
Heart failure		Overactive thyroid (hyperthyroid)	
Structural heart problems in childhood (hole in heart, VSD, ASD)		Underactive thyroid (hypothyroid)	
High lipids / cholesterol		Osteoarthritis	
High blood pressure		Osteoporosis	
Type 1 diabetes mellitus		Rheumatoid arthritis	
Type 2 diabetes mellitus		Gout	
Asthma		Back problems	
Chronic obstructive pulmonary disease (COPD)		Anaemia	
Migraine		Breast cancer	
Head injury with loss of consciousness		Cancer (other)	
Epilepsy/seizures		Immune system disorder	
Dementia (Alzheimer's or other type)		HIV	
Meningitis / encephalitis		Kidney problems	
Multiple sclerosis		Liver problems	
Parkinson's disease		Autoimmune disease (other)	
Stroke/brain haemorrhage		Cleft lip/palate	
Aneurysm		Other (please specify)	
Cataracts			
Gastric reflux			
Gastric or duodenal ulcers			
Coeliac disease or other malabsorption disorder			
Inflammatory bowel disease			
Irritable bowel syndrome			

Prescription Medication

Do you take prescription drugs routinely? Yes No

If yes, please list with doses:

Participant's Family History

Family History of Psychiatric Disorders

Record family history in biological relatives. First degree (parents, children, siblings). Second degree (grandchildren, grandparents, half-siblings, aunts/uncles). Record number of relatives.

	1st	2nd		1st	2nd
ADHD			Anorexia nervosa		
Autism			Bulimia nervosa		
Asperger's or other ASD			Obsessive compulsive disorder (OCD)		
Dyslexia			Agoraphobia		
Dyspraxia			Panic disorder		
Conduct disorder			Phobias		
Oppositional Defiant Disorder			Anxiety		
Tic disorder			Borderline personality disorder		
Tourette's disorder			Other personality disorder		
Intellectual disability			Post-traumatic stress disorder		
Depression			Alzheimer's disease		
Bipolar disorder / manic depression			Other dementia		
Mania / hypomania			Alcohol abuse / misuse		
Schizoaffective disorder			Other substance abuse / misuse _____ _____ _____ _____		
Psychosis			Genetic syndrome (e.g. VCFS) _____ _____ _____ _____		
Schizophrenia			Self harm or suicide attempts		
Postnatal psychosis			Other _____ _____ _____ _____		
Postnatal depression					

Family History of Physical Health Diagnoses

Biological relatives - 1st degree (parents, children, siblings), 2nd degree (grandchildren, grandparents, half-siblings, aunts/uncles). If at least one, but participant is not sure if more than one, record one.

	1st	2nd		1st	2nd
Heart disease			Inflammatory bowel disease		
Heart failure			Irritable bowel syndrome		
Structural heart problems in childhood (hole in heart, VSD, ASD)			Hernia		
High lipids / cholesterol			Overactive thyroid (hyperthyroid)		
High blood pressure			Underactive thyroid (hypothyroid)		
Type 1 diabetes mellitus			Osteoarthritis		
Type 2 diabetes mellitus			Osteoporosis		
Asthma			Rheumatoid arthritis		
Chronic obstructive pulmonary disease (COPD)			Gout		
Migraine			Back problems		
Head injury with loss of consciousness			Anaemia		
Epilepsy/seizures			Breast cancer		
Dementia			Cancer (other)		
Meningitis / encephalitis			Immune system disorder		
Multiple sclerosis			HIV		
Parkinson's disease			Kidney problems		
Stroke/brain haemorrhage			Liver problems		
Aneurysm			Autoimmune disease (other)		
Cataracts			Cleft lip/palate		
Gastric reflux			Other (please specify)		
Gastric or duodenal ulcers					
Coeliac disease or other malabsorption disorder					

Developmental History

Birth weight _____ lbs _____oz

Born: At term Early Late

By how much? _____

Birth complications Yes No

Please explain: _____

Postnatal complications Yes No

Please explain: _____

Milestones:

Walking About the right time Delayed

Talking About the right time Delayed

Hadyn Ellis Building Visit

CANTAB Cognitive Battery

Break – 25 minutes

Clinical Assessment Interview for Negative Symptoms (CAINS)

Introduction
In this interview, I'll be asking you some questions about things you have been doing over the past week. In the first section, I'm going to ask you some questions about your family, romantic partners, and friends, including how motivated you have been to spend time with them and how you felt when you were around them.

Social (Motivation and Pleasure)

Item 1: Motivation for Close Family/Spouse/Partner Relationships
The following questions are about your family. This can include relatives like parents, brothers or sisters and other relatives, as well as your spouse or live-in partner. Have you been in contact with or visited with any family members in the past week (in person, phone, e-mail)? Any contact with a spouse or partner?

If contact:
Who have you been in contact with? Anybody else? What things have you done with your family? What things have you done with your spouse/partner? How much time did you spend together?

Behaviour:
What have you done to see or contact your [family/spouse/partner] in the past week? When you were with your [family/spouse/partner] who decided what you would do? Who started the conversation? Did you start it? Did you [family/spouse/partner]? Were you involved in the conversation?
Did you ever find that you quickly wanted to end your interactions with your [family/spouse/partner]? Did you want them to last longer?

Motivation and interest in closeness:
Have you been motivated to be around or in touch with your [family/spouse/partner] in the past week? (Why is that?)
What did you talk about? Can you talk about good and bad times with your [family/spouse/partner]?
How close do you feel to your [family/spouse/partner]? What does being close mean for you? Were there times in the past week when you just didn't want to be around or in touch with your [family/spouse/partner]?
How important is being part of a family to you?
What about that is important to you? Have you felt this way throughout the past week?

If no family contact:
(This section applies when not part of a close family or if available relatives could be contacted but person has chosen not to interact. If the person is not currently in a relationship with a live-in spouse/partner, interest in romantic relationships is assessed in Item 2)
Has your family tried to contact you or visit you in the last week?
Has anything kept you or held you back from being in contact with your family?
Do you wish you were closer to your family? OR Do you wish you were part of a close family? Did you miss interacting with your family in the past week?
Is having a relationship with your family important to you? What about having a relationship is important to you?
Have you preferred to spend your time alone rather than with your family?

Item 1: Motivation for Close Family/Spouse/Partner Relationships _____

(Romantic relationships can be rated in either Item 1 or Item 2 but NOT both. A spouse/ partner relationship in which the couple is living together should be assessed in Item 1. A dating/romantic relationship in which the couple is not living together should be assessed in Item 2.)

0 No impairment	VERY INTERESTED in and highly values close family bonds as one of the most important parts of life. Strongly desires and is highly motivated to be in contact with family. Regularly initiates and persists in interactions with family and actively engages in these interactions; good and bad times are openly discussed. Well within normal limits
1 Mild deficit	GENERALLY INTERESTED in and values close family bonds though response suggests some minor or questionable reduction. Generally desires and is motivated to maintain contact with family. Has a close relationship with family member(s) in which good and bad times can be discussed. Mild deficit in initiating and persisting in regular interactions with family – generally actively engaged when interactions occur.
2 Moderate deficit	SOMEWHAT INTERESTED in family relationships and considers them somewhat important. May occasionally miss close connections with family but is only somewhat motivated to seek out interaction with family. Notable deficit in initiating and persistently engaging in interactions; discussion of good and bad times is limited. Interactions with family members may occur but are largely superficial and participation is best characterized as “going through the motions”; interactions are more likely initiated by family with mostly passive involvement of the person.
3 Moderately severe deficit	LITTLE INTEREST in family relationships (could “take it or leave it”) and does not describe family bonds as important. Describes hardly any motivation and minimal effort to have close family relationships. Rarely has discussion of good and bad times with family members. Contact and engagement with family is superficial and passive with almost all initiation and efforts to engage coming from others.
4 Severe deficit	NO INTEREST in family relationships and does not consider them at all important. Prefers to be alone and is not at all motivated to be with family. If person does see family, it is done so grudgingly, passively and with no interest.

Item 2: Motivation for Close Friendships and Romantic Relationships

Let’s talk about friends (and dating or romantic relationships) now. By friends, I mean people who you know and spend time with, anyone you consider a friend, or people you can rely on and count on. Have you had any contact with friends in the last week (in person, phone, email)? Have you been in contact with a romantic partner or dating in the last week (if relevant)?

If contact:
In the past week, what have you done with your [friends/partner/dates]?
Tell me about what you did [or what you talked about] during that [visit, activity, conversation]?
How much time did you spend together with [friends/partners/dates]?

Behaviour:
What steps did you take to see or contact your [friends/partner/dates] in the past week?
When you were with your [friends/partner/dates], who decided what you would do?
When you spoke with your [friends/partner/dates], who started the conversation? Did you?
Did you ever find that you quickly wanted to end your interaction with your [friends/partner/dates]? Did you want them to last longer?

Motivation and interest in closeness:
Have you been motivated to be around your friends (partner/dates) in the past week? Why is that?
Can you talk about both good times and bad times?
Were there times in the past week when you just didn't feel like being around your friends (partner/dates)?
How important is having friendships (partner/dates) to you? What about that is important to you?
How close do you feel to your friends (partner/dates)? What does being close mean for you?

If no friends/romantic contact:
Are you interested in having friends or dating?
Is having friendships [or being in a romantic relationship] important to you? If Yes, what about [specify friendships/romantic partner] is important?
Did you miss these types of relationships in the past week?
Would you like to have friends [or a romantic partner] with whom you could talk about good and bad times?

(If any indication of interest) Have you taken any steps to meet someone who might be a friend (or romantic partner)?

Has anything kept you or held you back from being in contact with your friends?

Would you prefer to have friendships [or a romantic relationship] or would you prefer to be alone?

Item 2: Motivation for Close Friendships and Romantic Relationships _____

0 No impairment	VERY INTERESTED in and highly values friend/romantic relationships as one of the most important parts of life. Strongly desires and is very motivated to engage in friendships. Regularly initiates and persists in interactions with friends/partner and actively engages in these interactions; good and bad times are openly discussed. Well within normal limits.
1 Mild deficit	GENERALLY INTERESTED in and values friend/romantic relationships though response suggests some minor or questionable reduction. Generally desires and is motivated to engage in friendships. Has friendships/relationship in which good and bad times can be discussed though this may be less consistent. Mild deficit in initiating or persistently engaging during interactions with friends/partner. If no friends/relationship, misses friend/romantic relationships, is motivated to have friends/relationship, and makes efforts to seek out friends/relationship.
2 Moderate deficit	SOMEWHAT INTERESTED in friend/romantic relationships and considers them somewhat important. May occasionally miss close connections with friends/partner and is somewhat motivated to have friends/partner. Notable deficit in initiating and persistently engaging in interactions; discussion of good and bad times is limited. Interactions with friends/romantic partner may occur but are largely superficial and participation is best characterized as “going through the motions”; interactions are initiated by others with mostly passive involvement of the person. If no friend/romantic relationships, is only somewhat motivated to have friends/partner and rarely if ever seeks out friends/partner.
3 Moderately severe deficit	LITTLE INTEREST in friend/romantic relationships (could “take it or leave it”) and does not describe friends/partner as important. Describes hardly any motivation to have friendships, and would just as soon be alone. Contact and engagement with others is superficial and passive with

	almost all initiation and efforts to engage coming from others.
4 Severe deficit	NO INTEREST in friend/romantic relationships and does not consider them at all important. Prefers to be alone and is not at all motivated to have friends/partner.

Item 3: Frequency of Pleasurable Social Activities – *Past Week*

Now, I want to talk to you about how you felt during the times you spent with or were in contact with others during the past week. You can include times with any of the people we have talked about so far or anyone else. Did you have any enjoyable interactions with other people, such as:

Family (PAUSE)

Romantic or dating partners (PAUSE)

Friends (PAUSE)

Any other enjoyable social interactions or time spent with people? (PAUSE)

Ask about people brought up in other sections that were described as enjoyable interactions (if needed).

If yes:

What about that was enjoyable?

How many days did you enjoy/get pleasure from these interactions [time spent with xx person(s)] (for each)?

[If many (i.e., 5 or 6) days mentioned or if not clear which days of week interactions were enjoyed] Were there any days that you did not have enjoyable interactions with other people?

Sun	Mon	Tues	Wed	Thurs	Fri	Sat

[NOTE: Ratings are based on NUMBER OF DAYS IN THE WEEK that pleasurable activity with other people is experienced. When there are reports of several different activities occurring, clarify if these happened on same or different days.]

Item 3: Frequency of Pleasurable Social Activities – Past Week _____

0 No impairment	Pleasure experienced daily.
1 Mild deficit	Pleasure experienced 5-6 days.
2 Moderate deficit	Pleasure experienced 3-4 days.
3 Moderately severe deficit	Pleasure experienced 1-2 days.
4 Severe deficit	No pleasure reported.

Item 4: Frequency of Expected Pleasurable Social Activities – *Next Week*

Now I would like you to think ahead to next week (next 7 days), thinking about whom you will spend time with. You can include people you have already talked about or anyone else.

What do you think you will enjoy doing in the next week with other people?

For each answer provided:

What about it do you expect to enjoy?

How often do you think you will enjoy this in the next week?

Follow up:

Are there other experiences with people you think you will enjoy in the next week?

[NOTE: Ratings are based on total number of expected pleasurable activities, regardless of days on which they are expected to occur].

Item 4: Frequency of Expected Pleasurable Social Activities – Next Week ___

0 No impairment	Expecting MANY (7 or more) pleasurable experiences.
1 Mild deficit	Expecting enjoyment from SEVERAL (5-6) pleasurable experiences.
2 Moderate deficit	Expecting enjoyment from a FEW (3-4) pleasurable experiences.
3 Moderately severe deficit	Expecting a COUPLE (1-2) pleasurable experiences.
4 Severe deficit	Expecting NO pleasurable experiences.

Work and School (Motivation and Pleasure)

Item 5: Motivation for Work and School Activities

Now I am going to ask you some questions about work and school, including how motivated you have been for work or school activities and how you felt while doing these things over the past week. Have you been working or going to school over the past week? Any volunteer work? Are you in a work-related treatment program?

If in a relevant role:
Tell me about what you do in your [insert role here]
How much time has this involved over the past week?

Behaviour:
Have you been able to complete tasks at [insert role here]?
In the past week has anyone raised any concerns with your [insert role here] performance?
Have you missed any days in the past week? Why?
Does someone need to remind you about [insert role here]? Why is that?
Were there things you meant to do or were supposed to do but just never got around to doing them?
Why?

Motivation:
How do you feel about [insert role here]?
Have you been motivated to do your [insert role here]?
What motivates you to do your [insert role here]?
Were there times during the past week when you just didn't feel like [insert role here]?
How important is your [insert role here] to you? What about it is important?

If no current role:

Is there a reason why you are not currently (work/school/volunteer)?

Has anything held you back from looking for (work/school/volunteer)?

How do you feel about working or going to school or volunteering?

Have you felt much interest in work/school/volunteer? {Tell me more}

Is working important to you? What about working/going to school/volunteering is important?

Have you tried to take any steps to start working/going to school/volunteering? What steps have you taken? How often have you looked into work/school/volunteer?

Item 5: Motivation for Work and School Activities _____

0 No impairment	Person is VERY MOTIVATED to seek out work or school, or new opportunities in work or school; initiates and persists in work, school, or job-seeking on a regular basis. Well within normal limits.
1 Mild deficit	Person is GENERALLY MOTIVATED to seek out work or school or new opportunities in work or school; a mild deficit in initiating and persisting; may report instances of initiating, but with moderate persistence.
2 Moderate deficit	Person is SOMEWHAT MOTIVATED to seek out work or school or new opportunities in work or school; notable deficit in initiating; may have initiated activities, but needed reminders on multiple occasions, and/or not initiated any new activities, and/or not persisted for very long.
3 Moderately severe deficit	Person is only SLIGHTLY MOTIVATED to seek out work or school or new opportunities in work or school; significant deficit in initiating; may have needed constant reminders, and/or initiated a few activities; did not persist for very long.
4 Severe deficit	Person is NOT AT ALL MOTIVATED to seek out work / school; nearly total lack of initiation and persistence in work, school, or job seeking.

Item 6: Frequency of Expected Pleasurable Work and School Activities – *Next Week*

Now I would like you to think ahead to NEXT week (next 7 days); thinking about work/volunteer/school.

If has a relevant role:
What do you think you will enjoy doing in the NEXT week at work/volunteer/school, etc.

If no relevant role:
Do you think you will enjoy anything related to seeking paid or volunteer work, or school?

For each answer provided:
What about it do you expect to enjoy?
How often do you think you will enjoy this in the next week?

Follow up:
Are there other work/school experiences you think you will enjoy in the next week?
[NOTE: Ratings are based on total NUMBER OF EXPECTED PLEASURABLE ACTIVITIES, regardless of days on which they are expected to occur].

Item 6: Frequency of Expected Pleasurable Work and School Activities – *Next Week* _____

0 No impairment	Expecting MANY (7 or more) pleasurable experiences.
1 Mild deficit	Expecting enjoyment from SEVERAL (5-6) pleasurable experiences.
2 Moderate deficit	Expecting enjoyment from a FEW (3-4) pleasurable experiences.
3 Moderately severe deficit	Expecting a COUPLE (1-2) pleasurable experiences.
4 Severe deficit	Expecting NO pleasurable experiences.

Recreation (Motivation and Pleasure)

Item 7: Motivation for Recreational Activities

In the next section, I am going to ask you some questions about what you do in your free time – any hobbies or recreational activities. I will ask about your motivation and feelings about the things that you have done in your free time over the past week.

What have you done in your free time in the past week?
Have you participated in any hobbies or leisure activities such as sports or games, going to church, TV, music, reading, internet, walking or other such activities during the past week?

If yes:
Behaviour:
Tell me about (activity). How much time has this involved over the past week?
Did you want to do (activity) more than that? Did it last longer than you had hoped? Why did it only last for (xx)?
Did anything get in the way of doing these activities over the past week? What was that?
Who initiated these activities? Did someone need to remind you to participate in these activities?

Motivation:
How has your motivation or drive to get involved in these activities been over the past week?
Did you ever feel like you just weren't very interested in these activities?
Are these types of activities important to you? Why? Have you been interested in these activities?
Did you ever feel that you would just as soon do nothing instead of getting involved in these types of activities?

If no:

Is there a reason why you haven't gotten involved in any hobbies or recreational activities in the past week? Have you wanted to or were you motivated to do something with your free time in the past week?

Did anything ever get in the way of doing these types of activities over the past week? What was that?

Item 7: Motivation for Recreational Activities _____

0 No impairment	Person is VERY MOTIVATED to seek out hobbies and recreational activities; initiates and persists in hobbies and recreational activities on a regular basis, well within normal limits.
1 Mild deficit	Person is GENERALLY MOTIVATED to seek out hobbies and recreational activities; a mild deficit in initiating and persisting; may report initiating hobbies, but with moderate persistence.
2 Moderate deficit	Person is SOMEWHAT MOTIVATED to seek out hobbies and recreational activities; notable deficit in initiating; may have initiated some activities and/or not persisted for very long. Others were somewhat more likely to initiate hobbies or activities.
3 Moderately severe deficit	Person is only SLIGHTLY MOTIVATED to seek out hobbies and recreational activities; significant deficit in initiating and persisting; may have initiated a few activities and not persisted for very long. Others were much more likely to initiate hobbies or prompt initiation.
4 Severe deficit	Person is NOT AT ALL MOTIVATED to seek out hobbies and recreational activities; nearly total lack of initiation and persistence in hobbies or recreational activities.

Item 8: Frequency of Pleasurable Recreational Activities – *Past Week*

Did you have any enjoyable (pleasurable) experience from things you did in your free time last week? You can include any of the activities we've talked about so far or any other leisure activities in the past week, including TV, sports or games, going to church, music, reading, internet, walking or other such activities?

What about [insert activity here] was enjoyable?

How many days did you enjoy/get pleasure from these experiences?

Ask about activities brought up in other sections that were described as enjoyable.

Follow up:

Any other enjoyable experiences from things you do in your free time or your hobbies?

Activity	Sun	Mon	Tues	Wed	Thurs	Fri	Sat

[NOTE: Rating is based on both VARIETY of pleasurable activities and DAILY FREQUENCY that these are experienced. When there are reports of several different activities occurring, need to clarify if these happened on same or different days.]

Item 8: Frequency of Pleasurable Recreational Activities – *Past Week* _____

0 No impairment	At least A FEW (3) different types of pleasurable experiences, experienced daily.
1 Mild deficit	At least A FEW (3) different types of pleasurable experiences, experienced more days than not.
2 Moderate deficit	1 or 2 different types of pleasurable experiences, experienced more days than not.
3 Moderately severe deficit	1 type of pleasurable experience, experienced on just a few days.
4 Severe deficit	No pleasurable experiences.

Item 9: Frequency of Expected Pleasurable Recreational Activities – *Next Week*

Now I would like you to think ahead to NEXT week (next 7 days), thinking about your free time/hobbies/ recreation. You can include any of the activities you have already talked about or anything else. What do you think you will enjoy doing in the NEXT WEEK in your recreational/free time?

For each answer provided:

What about it do you expect to enjoy?

How often do you think you will enjoy [activity] in the next week?

Follow up:
Are there other things you do in your free time like hobbies or recreational activities that you think you will enjoy in the next week?

[NOTE: Ratings are based on total NUMBER OF EXPECTED PLEASURABLE ACTIVITIES, regardless of days on which they are expected to occur]

Item 9: Frequency of Expected Pleasurable Recreational Activities – *Next Week* _____

0 No impairment	Expecting MANY (7 or more) pleasurable experiences.
1 Mild deficit	Expecting enjoyment from SEVERAL (5-6) pleasurable experiences.
2 Moderate deficit	Expecting enjoyment from a FEW (3-4) pleasurable experiences.
3 Moderately severe deficit	Expecting a COUPLE (1-2) pleasurable experiences.
4 Severe deficit	Expecting NO pleasurable experiences.

Expression

Item 10: Facial Expression

When making the facial expression rating, consider facial movements across all parts of the face, including in the eyes (e.g., raised brows when surprised), mouth (smiling or grimacing), and mid-face (e.g., wrinkled nose when disgusted).

Item 10: Facial Expression _____

0 No impairment	WITHIN NORMAL LIMITS; frequent expressions throughout the interview.
1 Mild deficit	MILD DECREASE in the frequency of facial expressions, with limited facial expressions during a few parts of the interview.
2 Moderate deficit	NOTABLE DECREASE in the frequency of facial expressions, with diminished facial expressions during several parts of the interview.
3 Moderately severe deficit	SIGNIFICANT LACK of facial expressions, with only a few changes in facial expression throughout most of the interview.
4 Severe deficit	NEARLY TOTAL LACK of facial expressions throughout the interview.

Item 11: Vocal Expression

This item refers to prosodic features of the voice. This item reflects changes in tone during the course of speech. Speech rate, amount, or content of speech is not assessed.

Item 11: Vocal Expression _____

0 No impairment	WITHIN NORMAL LIMITS. Normal variation in vocal intonation across interview. Speech is expressive and animated.
1 Mild deficit	MILD DECREASE in vocal intonation. Variation in intonation occurs with a limited intonation during a few parts of the interview.
2 Moderate deficit	NOTABLE DECREASE in vocal intonation. Diminished intonation during several parts of the interview. Much of speech is lacking variability in intonation but prosodic changes occur in several parts of the interview.
3 Moderately severe deficit	SIGNIFICANT LACK of vocal intonation with only a few changes in intonation throughout most of the interview. Most of speech is flat and lacking variability, only isolated instance of prosodic change.
4 Severe deficit	NEARLY TOTAL LACK OF change in vocal intonation with characteristic flat or monotone speech throughout the interview.

Item 12: Expressive Gestures

Expressive gestures are used to emphasize what is communicated verbally through gestures made with the hands, head (nodding), shoulders (shrugging), and trunk (leaning forward, leaning back).

Item 12: Expressive Gestures _____

0 No impairment	WITHIN NORMAL LIMITS; uses frequent gestures throughout the interview.
1 Mild deficit	MILD DECREASE in the frequency of expressive gestures, with limited gestures in a few parts of the interview.
2 Moderate deficit	NOTABLE DECREASE in the frequency of expressive gestures, with lack of gestures during several parts of the interview.
3 Moderately severe deficit	SIGNIFICANT LACK of expressive gestures, with only a few gestures throughout most of the interview.
4 Severe deficit	NEARLY TOTAL LACK of expressive gestures.

Item 13: Quantity of Speech

This item refers to the quantity of words spoken. Other speech abnormalities, such as disorganization, neologisms, or psychotic content are not rated here. For instance, a disorganized person may produce a large quantity of speech and have a low (normal) score on this item.

Item 13: Quantity of Speech _____

0 No impairment	NORMAL AMOUNT of speech throughout the interview. Replies provide sufficient information with frequent spontaneous elaboration.
1 Mild deficit	MILD DECREASE in the quantity of speech, with brief responses during a few parts of the interview.
2 Moderate deficit	NOTABLE DECREASE in speech output, with brief responses during several parts of the interview.
3 Moderately severe deficit	SIGNIFICANT LACK of speech, with very brief answers (only several words) in responses throughout most of the interview.
4 Severe deficit	All or nearly all replies are one or two words throughout the entire interview.

Temporal Experience of Pleasure Scale (TEPS)

DIRECTIONS: Please read each statement carefully and decide how true that statement is for you in general. Please respond to *all item*s. In the rare case where you have *never* had the experience described, think about the most similar experience you've had and make your response. Do *not* leave any blank. Choose only *one* response to each statement. Don't worry about being consistent in your responses. Choose from the following 6 response options and CIRCLE your response to the right of the item.

1	2	3	4	5	6
very false for me	moderately false for me	slightly false for me	slightly true for me	moderately true for me	very true for me

1. When I hear about a new movie starring my favourite actor, I can't wait to see it.	1 2 3 4 5 6
2. I enjoy taking a deep breath of fresh air when I walk outside.	1 2 3 4 5 6
3. The smell of freshly cut grass is enjoyable to me.	1 2 3 4 5 6
4. I look forward to a lot of things in my life.	1 2 3 4 5 6
5. I love it when people play with my hair.	1 2 3 4 5 6
6. Looking forward to a pleasurable experience is in itself pleasurable.	1 2 3 4 5 6
7. A hot cup of coffee or tea on a cold morning is very satisfying to me.	1 2 3 4 5 6
8. When I think of something tasty, like a chocolate chip cookie, I have to have one.	1 2 3 4 5 6
9. I appreciate the beauty of a fresh snowfall.	1 2 3 4 5 6
10. I get so excited the night before a major holiday I can hardly sleep.	1 2 3 4 5 6
11. When I'm on my way to an amusement park, I can hardly wait for ride the roller coasters.	1 2 3 4 5 6
12. I really enjoy the feeling of a good yawn.	1 2 3 4 5 6
13. I don't look forward to things like eating out at restaurants.	1 2 3 4 5 6
14. I love the sound of rain on the windows when I'm lying in my warm bed.	1 2 3 4 5 6
15. When I think about eating my favourite food, I can almost taste how good it is.	1 2 3 4 5 6
16. When ordering something off the menu, I imagine how good it will taste.	1 2 3 4 5 6
17. The sound of crackling wood in the fireplace is very relaxing.	1 2 3 4 5 6
18. When something exciting is coming up in my life, I really look forward to it.	1 2 3 4 5 6

Monetary Choice Questionnaire

For each of the next 27 choices, please indicate which reward you would prefer: the smaller reward today, or the larger reward in the specified number of days.

	Would you prefer……..	Smaller reward today	Larger reward in X days
1.	$54 today or $55 in 117 days?		
2.	$55 today or $75 in 61 days?		
3.	$19 today or $25 in 53 days?		
4.	$31 today or $85 in 7 days?		
5.	$14 today or $25 in 19 days?		
6.	$47 today or $50 in 160 days?		
7.	$15 today or $35 in 13 days?		
8.	$25 today or $60 in 14 days?		
9.	$78 today or $80 in 162 days?		
10.	$40 today or $55 in 62 days?		
11.	$11 today or $30 in 7 days?		
12.	$67 today or $75 in 119 days?		
13.	$34 today or $35 in 186 days?		
14.	$27 today or $50 in 21 days?		
15.	$69 today or $85 in 91 days?		
16.	$49 today or $60 in 89 days?		
17.	$80 today or $85 in 157 days?		
18.	$24 today or $35 in 29 days?		
19.	$33 today or $80 in 14 days?		
20.	$28 today or $30 in 179 days?		
21.	$34 today or $50 in 30 days?		
22.	$25 today or $30 in 80 days?		
23.	$41 today or $75 in 20 days?		
24.	$54 today or $60 in 111 days?		
25.	$54 today or $80 in 30 days?		
26.	$22 today or $25 in 136 days?		
27.	$20 today or $55 in 7 days?		

Behavioural Activation Scale

Each item of this questionnaire is a statement that a person may either agree with or disagree with. For each item, indicate how much you agree or disagree with what the item says. Please respond to all the items; do not leave any blank. Choose only one response to each statement. Please be as accurate and honest as you can be. Respond to each item as if it were the only item. That is, don't worry about being "consistent" in your responses. Choose from the following four response options:

1	2	3	4
Very true for me	Somewhat true for me	Somewhat false for me	Very false for me

1.	A person's family is the most important thing in life.	1	2	3	4
2.	Even if something bad is about to happen to me, I rarely experience fear or nervousness.	1	2	3	4
3.	I go out of my way to get things I want.	1	2	3	4
4.	When I'm doing well at something I love to keep at it.	1	2	3	4
5.	I'm always willing to try something new if I think it will be fun.	1	2	3	4
6.	How I dress is important to me.	1	2	3	4
7.	When I get something I want, I feel excited and energised.	1	2	3	4
8.	Criticism or scolding hurts me quite a bit.	1	2	3	4
9.	When I want something I usually go all-out to get it.	1	2	3	4
10.	I will often do things for no other reason than that they might be fun.	1	2	3	4
11.	It's hard for me to find the time to do things such as get a haircut.	1	2	3	4
12.	If I see a chance to get something I want I move on it right away.	1	2	3	4
13.	I feel pretty worried or upset when I think or know somebody is angry at me.	1	2	3	4
14.	When I see an opportunity for something I like I get excited right away.	1	2	3	4
15.	I often act on the spur of the moment.	1	2	3	4
16.	If I think something unpleasant is going to happen I usually get pretty "worked up".	1	2	3	4
17.	I often wonder why people act the way they do.	1	2	3	4
18.	When good things happen to me, it affects me strongly.	1	2	3	4
19.	I feel worried when I think I have done poorly at something important.	1	2	3	4
20.	I crave excitement and new sensations.	1	2	3	4

21.	When I go after something I use a “no holds barred” approach.	1	2	3	4
22.	I have very few fears compared to my friends.	1	2	3	4
23.	It would excite me to win a contest.	1	2	3	4
24.	I worry about making mistakes.	1	2	3	4

Appendix 3 – Clinical and cognitive effects of risk factors for mental illness study protocol

Clinical and Cognitive Effects of Risk Factors for Mental Illness

Recall Sample Protocol

Summary

Large, rare copy number variants (CNVs) occur in schizophrenia, autism spectrum disorder, intellectual disability and attention deficit hyperactivity disorder at higher rates than in controls and these associations cross traditional diagnostic boundaries. The range of phenotypes associated with the majority of neuropsychiatric CNVs is poorly defined.

The aims of this project are to identify CNV carriers and to establish the psychiatric, cognitive and physical phenotypes associated with neurodevelopmental CNVs. We hypothesise that neurodevelopmental CNV-carrying individuals will display more frequent and severe impairments than those without CNVs. The objectives of this project are i) to identify CNV carriers in two pre-existing clinical samples and, ii) to recall CNV carriers and matched controls and to carry out deep phenotypic assessment in psychiatric, cognitive and physical health domains. CNV carriers and matched non-carriers with psychiatric disorders will undergo deep phenotypic assessment in all three domains. Phenotypic features will then be compared between CNV carriers and non-carriers.

Study Information

Funder: Wellcome Trust

Investigator: Dr K Kendall
MRC Centre for Neuropsychiatric Genetics and Genomics
Hadyn Ellis Building
Cardiff University
Maindy Road
Cardiff University
CF24 4HQ
02920 688418
kendallkm@cardiff.ac.uk

Supervisors: Professor James Walters
Professor George Kirov
Professor Ian Jones

Background

Copy number variants (CNVs) are the deletion or duplication of greater than 1,000 DNA base pairs resulting in altered dosage of the affected sequence. Large, rare CNVs are associated with neurodevelopmental disorders, including schizophrenia, autism spectrum disorder, intellectual disability and attention deficit hyperactivity disorder and these associations cross traditional diagnostic boundaries. Until recently, information on CNV-associated phenotypes has largely come from syndromic presentations to clinical genetics services at the more severely affected end of the spectrum. Little has been known about the effects of these CNVs on apparently unaffected individuals from the general population or from psychiatric patient cohorts. We, and others, have now built on this information by showing that carriers of neurodevelopmental CNVs without severe neurodevelopmental disorders have i) impaired cognitive function relative to non-carriers, ii) a range of physical health phenotypes with the potential to affect an individual's quality of life and mortality risk, and iii) an increased risk of depression. What remains unknown is the effect of neurodevelopmental CNVs on individuals with schizophrenia and other severe mental health disorders.

Aims and Objectives

1. To identify neurodevelopmental CNV carriers in the Cardiff Cognition and NCMH samples.
2. To recall neurodevelopmental CNV carriers with psychiatric disorders and matched controls, carry out psychiatric, cognitive and physical phenotypic assessments for these individuals, and compare phenotypes between cases and controls.

Study Design

This is a recall case control study of individuals from existing psychiatric genetic datasets, which is expected to be active until at least January 2021. Sample sizes, pathogenic CNV rates and expected response rates are shown in Table 1.

Sample	Diagnosis	Sample Size	Rate of CNVs	Number with CNVs	Number Likely to Respond
Cardiff Cognition	Schizophrenia Schizoaffective disorder	1015	2.76%	28	14
NCMH	Affective disorders	4500	1.65%	58	29
				Total	~43

Table 1. Sizes, rates of 53 neurodevelopmental CNVs and expected response rates for the Cardiff Cognition and NCMH samples.

Inclusion criteria:

- 16 years of age or older.
- Able to understand written and spoken English.
- No uncorrected deficits in sight or hearing.

Exclusion criteria:

- Current inpatients.
- Individuals currently under the care of the Crisis Resolution and Home Treatment Team.
- Individuals deemed unsuitable to participate at that particular time.

Methodology

CNVs will be called in all samples using the standard PennCNV pipeline as previously used for the UK Biobank part of the study. Carriers of any of the 53 CNVs statistically associated with neurodevelopmental disorders will be identified along with controls matched according to age and sex (Appendix 1). Participants who have previously consented to being re-contacted for future research studies will be invited to participate via invitation letter and/or by telephone. This method of recruitment has been used successfully in existing studies and the parent studies have monitoring in place to ensure participants are not overburdened with such requests.

An invitation letter will be sent to participants. If no response is received within 2 weeks, this will be followed up by telephone. Individuals who agree to receive further information will be sent an information sheet and a copy of the consent form. Those who agree to participate will have a home visit 2 weeks prior to the assessment session when there will be an opportunity to ask questions, the consent procedure will take place and the actigraph and sleep diary will be delivered and explained. The assessment session will take place at the Hadyn Ellis Building, Cardiff University when there will be a further opportunity to ask questions and check consent.

Phenotypic Assessments

The phenotypic assessments have been selected to try to gain a broad overview of the phenotypes associated with neurodevelopmental CNVs and those encountered more generally across psychiatric disorders (Tables 2 and 3). There is a particular focus on negative symptoms and, where possible, assessments have been mapped to RDoC domains (Appendix 2).

Assessment		Phenotype	Information
Community Visit			
Actigraph and sleep diary	Psych	Negative symptoms	Activity monitoring for 2 weeks prior to assessment session.
Education and work	Cog	Education and work	Pre-fill with data last collected and double check with participant.
Participant's medical history	Psych Phys	Psychiatric history, medical history, medications	Physical: Combined disorders a) associated with CNVs, b) from NCMH assessment. Psychiatric: Same as NCMH assessment. Treatment/Treatment responsiveness: Current medications and medication review from notes.
Participant's family history	Psych Phys	Psychiatric history, medical history	As for participant's medical history above (excluding medications).
Developmental and Motor History Form	Phys	Development	Will likely need an informant's input (and possibly to shorten).
Time Use Diary	Psych Phys	Negative symptoms Functioning	Office of National Statistics Time Use Diary.

HEB Visit			
CANTAB Battery	Cog	Cognition Negative symptoms	See Table 3 for individual tests.
Test of Premorbid Functioning (TOPF)	Cog	Premorbid IQ	Similar to National Adult Reading Test (NART).
Clinical Assessment Interview for Negative Symptoms (CAINS)	Psych	Negative symptoms	Information can also be used to fill in SANS.
Premorbid Adjustment Scale	Psych	Premorbid adjustment	
Temporal Experience of Pleasure Scale (TEPS)	Psych	Negative symptoms	Anticipatory and Consummatory scales. Maps to RDoC positive valence system domains.
Monetary Choice Questionnaire	Psych	Negative symptoms	Delay within reward valuation (RDoC framework).
Behavioural Activation Scale	Psych	Negative symptoms	Drive subscale – effort within reward valuation. Also reward responsiveness (RDoC framework).
Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview	Psych	Psychiatric diagnosis	Adapted interview for validating diagnosis.
Assessment of movement	Phys	Movement disorder	Videotaped examination for movement disorder. Included some soft neurological signs.
Physical assessment and dysmorphology	Phys	Physical health	Height, weight, head circumference, photographs for dysmorphology.
Scales for the Assessment of Positive and Negative Symptoms (SAPS and SANS)	Psych	Positive and negative symptoms:	Should be able to rate these based on CAINS answers and answers to SCAN interview.

Table 2. Assessments.

CANTAB Task	Examines	Estimated Run Time
Cambridge Gambling Task	Decision-making and risk taking behaviour outside a learning context.	Up to 18 minutes
Stop Signal Task	Impulse control.	Up to 20 minutes
One Touch Stockings of Cambridge	Executive function.	10 minutes
Emotion Recognition Task	Emotion recognition.	6 minutes
Reaction Time Task	Motor and mental response speeds, movement and reaction time, response accuracy and impulsivity.	3 minutes
Spatial Working Memory	Spatial working memory, executive function.	4 minutes
Paired Associates Learning	Visual memory, new learning	8 minutes
Verbal Recognition Memory	Verbal memory, new learning	10 minutes

Table 3. CANTAB assessments. Total estimated run time of up to 79 minutes.

Data Management and Statistical Analysis

- Assessments carried out on paper will be identified anonymously by study ID and kept in a locked cupboard/drawer.
- Assessments carried out on the CANTAB iPads will be identified anonymously by study ID and uploaded automatically to the cloud.
- Assessment data will be recorded in SPSS and stored on the shared drive.
- Identifiable information (contact details and study ID) will be kept in a password-protected file on the shared drive.

The statistical methods used will be identified following data collection and advice from the Bioinformatics and Biostatistics Unit.

Safety Considerations

Standard Operating Procedure (SOP) for Home Visits

Participants will usually be psychiatrically well at the time of the research interview. This is important in terms of the risk assessment. It is also important to remember that the quality of information elicited at the interview is likely to be poor if a participant is mentally unwell or agitated for any reason. If there are any doubts or concerns regarding an individual's ability to participate in the research (either during screening or at interview) please discuss this with your line manager.

General Safety Issues

If team members ever feel threatened or at risk they should leave the situation. Team members should use their discretion in situations in which they feel uncomfortable, and leave where appropriate. Any such instances should be discussed with your line manager immediately.

Team members should raise any issues/potential problems with their line manager as soon as possible.

The line manager will ensure that any incidents that raise concern for the wellbeing of the participant or others is fed back to their clinical care team.

Field team members are responsible for ensuring they are familiar with this policy and refresh their knowledge of it as regularly as they need to.

Protocol for Home Visits

At the time of arranging home visits, relevant information should be obtained about safety issues. **The risk assessment must be completed for all potential participants within a week of the research interview (ideally within a day or two)**. If the visit is postponed for any reason, resulting in the risk assessment becoming out of date (i.e. the risk assessment is no longer within a week of the research interview) another risk assessment should be conducted prior to the rearranged research visit.

An attempt should always be made to try to talk to the participant on the telephone prior to a visit **(preferably the day before).** This builds a relationship, allows verification of mental state and also reminds the participant that you will be visiting them. This will aid in your risk assessment and will also result in a reduced number of "no show" interviews.

For systematically recruited participants, the responsible medical officer should refer only those patients who they feel do not pose any risk (in terms of current mental state and risk history).

For non-systematically recruited participants, agreement should be obtained from the participant (during the risk assessment) for us to inform their key worker (or relevant health professional) that they will be taking part in the research. If the participant agrees, their key-worker should be contacted by telephone and asked whether they feel there are any issues in terms of risk for a lone worker visiting the patient. If the participant does not agree the interviewer should discuss this with their line manager. Where a participant is not in contact with secondary care services and so does not have a key worker, participants should be visited in a pair.

Prior to the home visit, records of the location and timing of visits and contact arrangements should be entered onto the home visits diary. This should also include the make / model and registration number of the vehicle in which the interviewer is travelling to and from the visit, whether this may be the interviewer's personal car or a hire car.

During the visit safety guidelines should be followed including avoiding clothes or cases that may suggest you are carrying money, drugs etc.

Field team members should carry official University identification on every visit (photo ID).

Field workers should carry a mobile phone with the Stay Safe App installed. To access

the emergency services call 999 or 112. Prior to visits, it is the personal responsibility of each field worker to ensure that their mobile phone is credited and charged.

Field team members should carry an alternative means of communication when on visits if possible, e.g. a personal mobile phone as well as work mobile phone.

If there is **any** doubt over having sufficient mobile data reception at the location of a visit (for example, a rural area), it is possible to start the session in an area of reception (for example, the Haydn Ellis Building) and include sufficient time to reach the visit, conduct the interview and return to an area of known signal (e.g. the Haydn Ellis Building). If, when you arrive at the destination you do have a good signal, you can the cancel that session and start a new one if you prefer.

When arriving at the visit, the interviewer should start a session on the Stay Safe App by indicating the expected duration. Once the countdown timer has been set, the interviewer will be prompted to confirm their location on the map and move the pin if necessary, to indicate an accurate location. There is an additional option of providing a written description, which is necessary if the interviewer's specific location is unclear from GPS alone, for example if a visit is in a block of flats, at an office building, or in an area of densely packed buildings.

At the beginning of the interview, the interviewer should make his / her mobile phone visible to the interview participant and should explain the need to leave the phone switched on during the interview as a safety procedure is in operation. Mobile phones must be left **switched on** during interviews. If additional time is needed to complete the interview, select 'Extend' before the original session expires.

Having left the visit, select 'End' to terminate the session on the StaySafe App. If no response is given at the end of a session, Securitas will be alerted and the interviewer will receive a call to check they are safe. If the safety of the interviewer cannot be ascertained, calls will be made to nominated members of the NCMH team. If there is concern for the safety of the interviewer, the police will be contacted and sent to the

GPS location.

In the event of one of the nominated NCMH contacts being alerted of a possible problem by Securitas, they should:

a. attempt to contact the Researcher on their mobile;
b. if unable to contact the Researcher, contact the Participant on the number supplied (both landline and mobile if both are supplied);
c. if unable to contact either the Researcher or Participant, immediately inform the designated manager (in the first instance the Research Associate or, in her absence, the 1) NCMH Manager or 2) Data Manager or 3) Deputy Director or 4) Director) who will attempt to contact the Researcher again once five minutes have elapsed;
d. if it is still not possible to establish contact with the Researcher, the designated manager will contact the police.

If the interviewer feels under threat during the assessment, they **must** use the panic button on the Stay Safe App. If the device has no signal, then the interviewer should **telephone** the relevant contact person and use the amber code-word, **saliva pot** (e.g.. ”Can you check whether I’ve left my saliva pot in the office?”). The contact person must inform the line manager and call the participant back in **five** minutes and check whether the situation has escalated.

If the interviewer feels in danger and needs assistance, they **must** use the panic button on the Stay Safe App. If there is no signal, then the interviewer should either contact the emergency services directly or contact the office and use the red code-word, **blood pack** (e.g. “I have left my blood pack at the office.”). The contact person must then inform the NCMH Research Associate (or NCMH Manager or Director depending on who is available) and contact the emergency services.

If a visit by a single interviewer to the participant’s home is thought to involve potential risk, the following options should be considered: i) do not conduct assessment due to level of risk, ii) conduct the visit at the particpant’s home with 2

interviewers iii) invite the participant to the HEB clinic or other NHS clinic in order to conduct the interview (in accordance with the risk assessment). This should be discussed with the line manager before arranging the interview.

Some flexibility may be required with this protocol and there may be occasions when it is not necessary or appropriate to follow all of these guidelines. However, deviations from the guidelines should be discussed with your line manager. It is essential that field team members make an assessment of the safety issues relevant to each visit and take all sensible steps to look after their own safety and that of other persons.

Incident Reporting

It is important that staff report any accident, incident or near miss to their Line Manager at the earliest opportunity. If the incident occurs on NHS premises, the incident should also be reported to the NHS team/clinician.

The individual involved in the incident should detail the incident on an incident report form. This should be signed off by the individual involved in the incident and their Line Manager.

The employee will be offered advice and support from their Line Manager particularly with regard to any preventative action that may be taken in future.

If involved in a car accident or breakdown whilst conducting university business, the Line Manager should be informed.

Vehicle Security / Personal Safety

It is the duty of the field team member to ensure that the car that they are travelling in for work purposes is in good working order (MOT, Tax, Insurance, serviced regularly).

You must ensure that you have the correct type of motor insurance (to include any business usage if necessary).

Simple maintenance checks on personal cars being used for work related trips are a necessity: oil, water and tyre pressure before embarking on a trip; all lights are in full working order.

Please see Cardiff University Car Hire Policy for additional requirements when using a hire car for university business.

Always park close to the address being visited and in a well-lit area if possible.

Always have vehicle keys in hand when leaving premises/building (this saves time looking for keys whilst stood outside vehicle, thereby preventing a personal safety risk).

Never leave items on display in vehicles e.g. bags, brief cases, laptops etc. Be alert when transferring laptops, mobile phones to and from cars. When leaving your car ensure you place all items out of sight in the boot.

Do not leave work laptops or participant files in an unattended vehicle under any circumstances.

Keep your mobile telephone out of sight when not in use (apart from during a research interview).

If someone tries taking anything from you, let them take it rather than get into a confrontation and risk an injury. If this happens throw any bags, equipment away from you and run away.

Dealing with the Threat of Animals – All Types (Basic Guidelines)

If attending a visit and a member of staff is confronted by a dangerous animal or pet e.g. an aggressive dog, they must not put themselves at risk. The Line Manager should be contacted and informed of the situation and visit abandoned if necessary.

Safety Considerations for On-Site Assessment

Prior to commencing assessment:

- The assessor must ensure unimpeded access to the door.
- The assessor must familiarise themselves with the location of panic alarms and the procedure for using them (see assessment file).

During the assessment:

- Where possible, two assessors should be present. Others in the department should be made aware of the location of the assessors and their expected time of return to the department.
- Assessors should try to remain an arm's distance from the participant whenever possible.
- Assessors should aim to mirror the participant's positioning i.e. if the participant is sitting down, the assessors should be sitting down (do not stand over the participant).
- Assessors should try to start with less intrusive questions which are open ended.
- Assessors should be alert to changes in the participant's body language, which suggest that the participant is uncomfortable and must respond appropriately.
- If, at any point, the assessor feels they are under threat, they must leave the room and seek help immediately.
- If any safety related issues are raised by the participant (e.g. suicidal ideation/plans), a more detailed risk assessment should be carried out and discussed with senior members of staff followed by appropriate liaison with relevant individuals (e.g. the participant's GP).

Policy on the Return of Genotypes

(NCMH Sample Only)

During the NCMH consent procedure, individuals were asked the following:

I understand that very occasionally the researchers may discover genetic risk factors which may have important implications for my future health or for the health of my family. In these rare circumstances, the researchers will take advice from a clinical geneticist who may then advise the researchers to re-contact me and my GP and offer me the opportunity to seek further advice through a genetic counselling service.

I wish to be contacted about any findings that may have important implications for my future health or for the health of my family.

I do not wish to be contacted about any findings even though that may have implications for my future health or the health of my family.

By 2016, 93% of participants had indicated that they wished to be contacted in the above circumstances.

Multi-disciplinary team meetings, attended by clinical academic psychiatrists and medical genetics professionals, were held to discuss the procedure for offering the return of CNV genotypes and the ethical issues around this. Dr K Kendall also spent some time at the Centre for Health, Law and Emerging Technologies, Oxford University to gain additional perspectives on how to manage this process. A full discussion of the issues raised and addressed is in the thesis. To summarise, there are potential pros and cons to returning CNV genotypes to research participants. On the one hand, some of the phenotypes associated with neurodevelopmental CNVs are treatable and participants in the NCMH sample have been consented to receive genetic results. On the other hand, some of the CNV-phenotype associations are based on relatively small samples, the CNVs are of variable penetrance and there is a risk of distress when contacting someone about their potential CNV carrier status. In

order to balance respect for participant autonomy with the risks of disclosing genetic information with uncertain implications, we reached the decision to return pathogenic and likely pathogenic CNVs only. These classifications are based on those used in the clinical services in Medical Genetics and are derived from multiple sources of information. Variants of uncertain significance will not be returned. The procedure for returning results is shown in Figure 1.

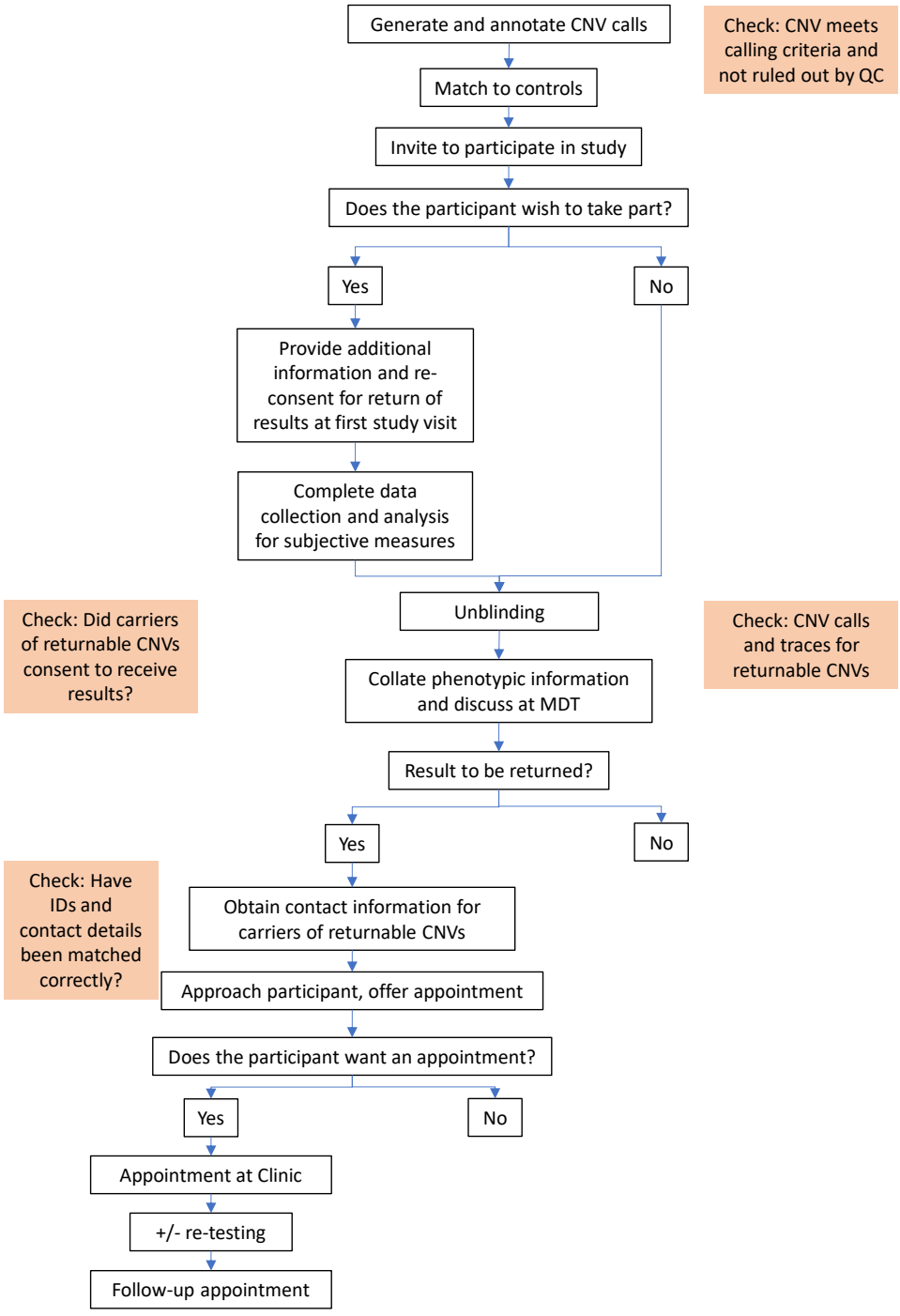

Figure 1. Procedure for the return of CNV genotypes.

Individuals who participate – individuals who participate will undergo the consent procedure as already established. In addition, they will be shown the above NCMH consent statement on the return of genetic results. The researcher, who at this stage is blind to CNV carrier status, will explain what form these genetic results may take and the potential implications of receiving such a result. They will also explain that only results which are of clinical relevance according to our knowledge at that time, will be returned (i.e. pathogenic or likely pathogenic results). Results not clinically relevant or of uncertain clinical relevance (i.e. variants of uncertain significance) will not be returned. The participant will then be asked whether their response remains the same. If the participating individual is later revealed to be a carrier of a returnable CNV and they indicated they wanted to receive this information, they will be contacted by the researcher offering them an appointment at the Psychiatric Genetics Clinic. Should they wish to have an appointment, this will take place with a medical genetics consultant and academic psychiatrist where the pros and cons of retesting for CNVs will be discussed. The course of action will be determined by the participant, in conjunction with these healthcare professionals at this point.

Individuals who are contactable but do not participate – if a participant is later revealed to be a carrier of a returnable CNV and they previously indicated they wanted to receive this information, they will be contacted by the researcher offering them an appointment at the Psychiatric Genetics Clinic.

Individuals who do not reply – the participant's address will be checked with their General Practitioner / Community Mental Health Team. The researcher will again attempt to make contact via letter and telephone where possible. If there is no reply after 4 weeks, a further, final letter will be sent. If there is no reply to this, further attempts will not be made to contact the participant with regards to the return of CNV results. If any individual expresses a wish to have genetic testing done, and they are not a carrier of a returnable CNV, the researcher will explain that they can seek referral to Medical Genetics via their GP.

Draft Letter for use in the Case of the Return of CNV Genotypes

[DATE]

Dear [PARTICIPANT'S NAME]

We are writing to you because you have previously taken part in mental health research with the National Centre for Mental Health (NCMH). At that time, you gave a blood sample for use in research. We used your blood sample to analyse the effects of genetic risk factors on mental illness.

When we asked for your consent to join NCMH one of the questions on the consent form was:
"I understand that very occasionally the researchers may discover genetic risk factors which may have important implications for my future health or for the health of my family. In these rare circumstances, the researchers will take advice from a clinical geneticist who may then advise the researchers to re-contact me and my GP and offer me the opportunity to seek further advice through a genetic counselling service."

You selected the following answer:
"I wish to be contacted about any findings that may have important implications for my future health or for the health of my family."

The reason for writing to you now is to inform you that we may have identified a possible change in the genetic material we prepared from your blood sample. It may have implications for your health, or possibly the health of your family.

We would like to offer you an appointment to discuss this further with a Medical Genetics doctor and Psychiatry doctor at the Psychiatric Genetics Clinic, Cardiff. They will explain the potential implications for you and/or your family, answer your

questions and invite you to have the genetic test repeated to confirm the initial findings. All of this will be completely up to you, it is your decision whether to attend the appointment or to have any tests. Any involvement you have with the Psychiatric Genetics Clinic would be fully confidential.

We would be most grateful if you could let us know whether you would like us to refer you to the Psychiatric Genetics Clinic – either e-mail us on INSERT E-MAIL ADDRESS HERE or complete the enclosed slip and sent it back to us in the stamped addressed envelope provided.

If we have not heard from you after 4 weeks from the date on this letter we will write to you again. If we do not hear from you after another 2 weeks we will assume you do not wish to proceed and we will not contact you again.

Reply Slip

I would / would not (please circle) like an appointment at the Psychiatric Genetics Clinic.

Signature _____

Name _____

Address _____

_____

Telephone number _____

Appendix 1

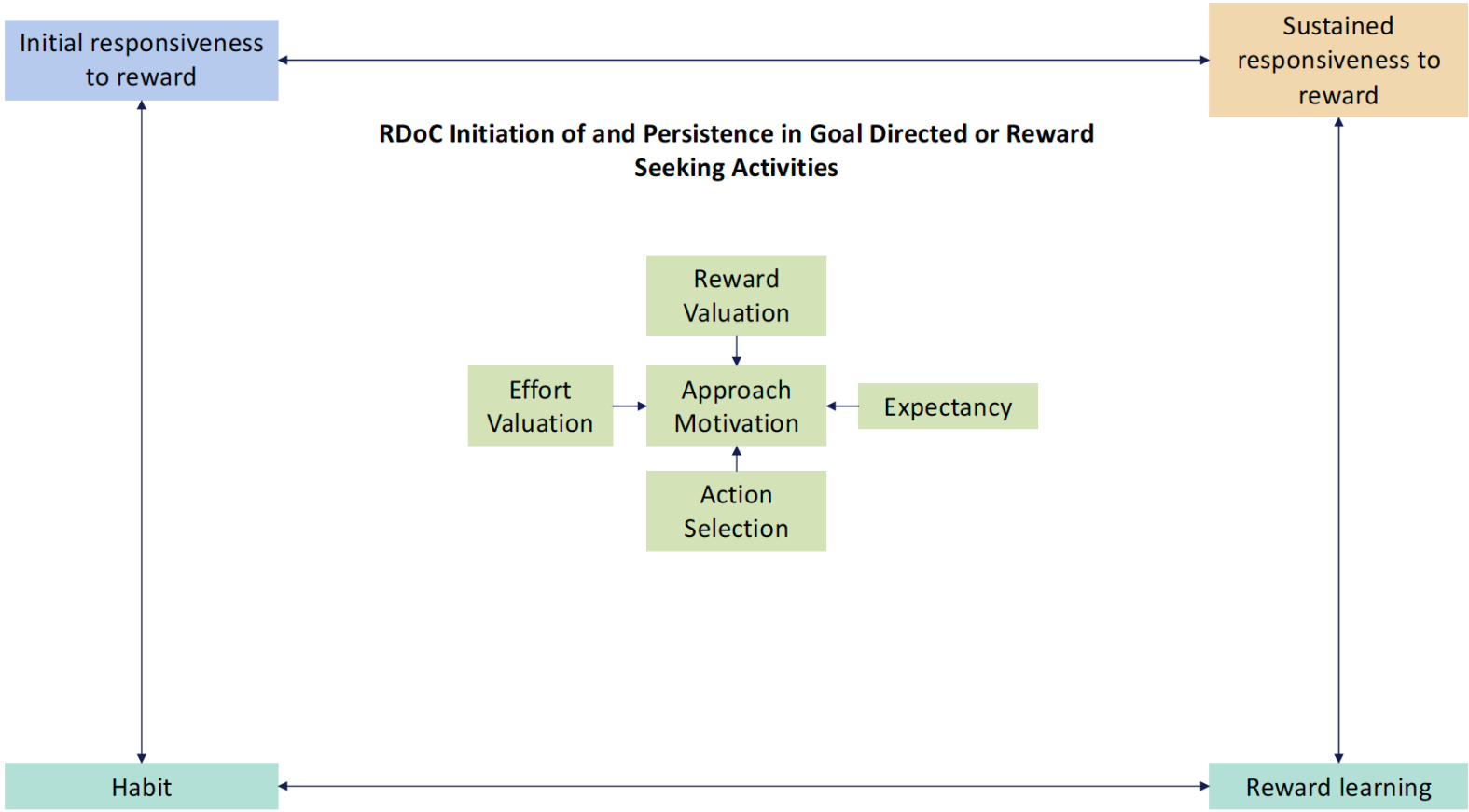

Figure 1. Schematic of the initiation of and persistence in goal directed or reward seeking activities using RDoC concepts and terminology.

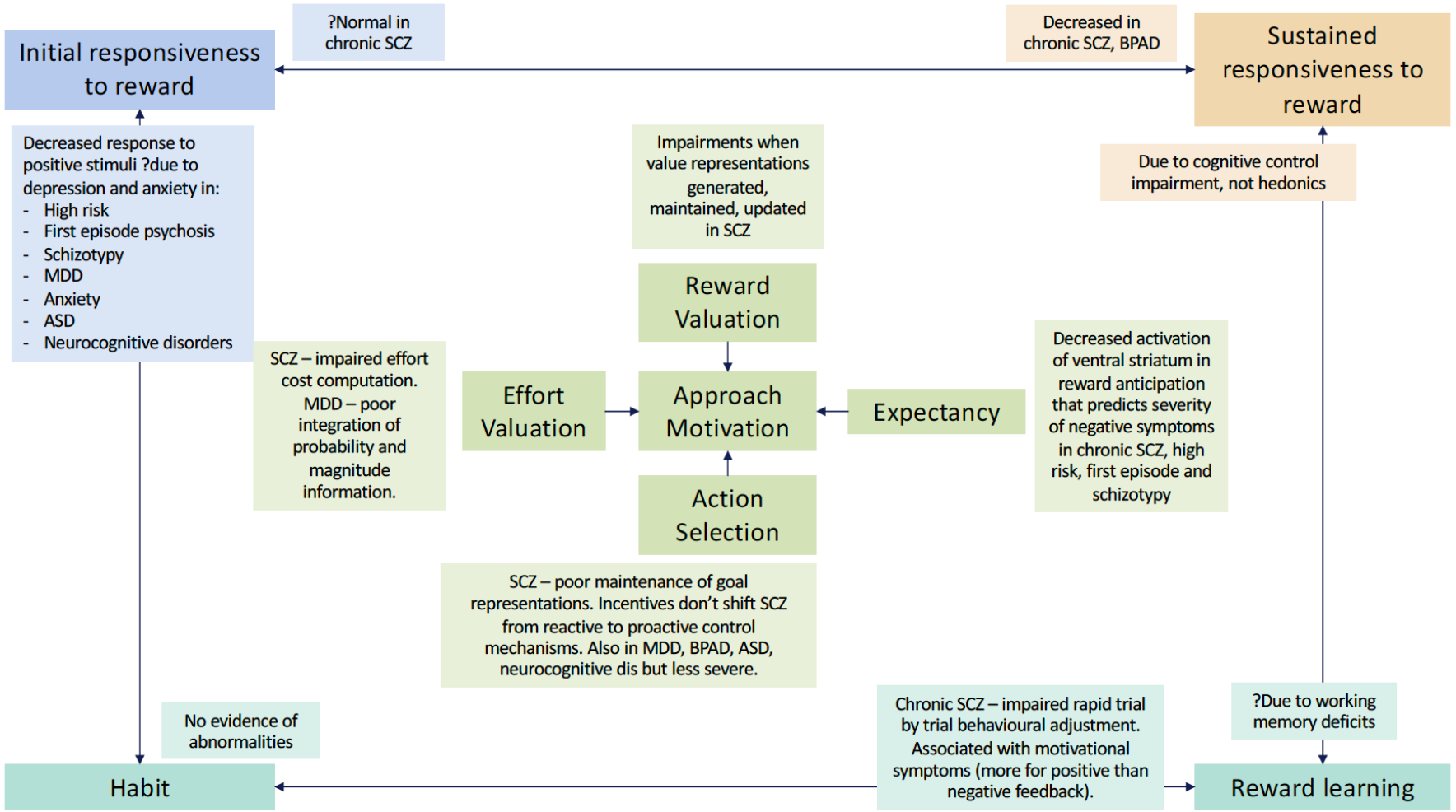

Figure 2. The same schematic in Figure 1 with the addition of a summary of the available evidence.

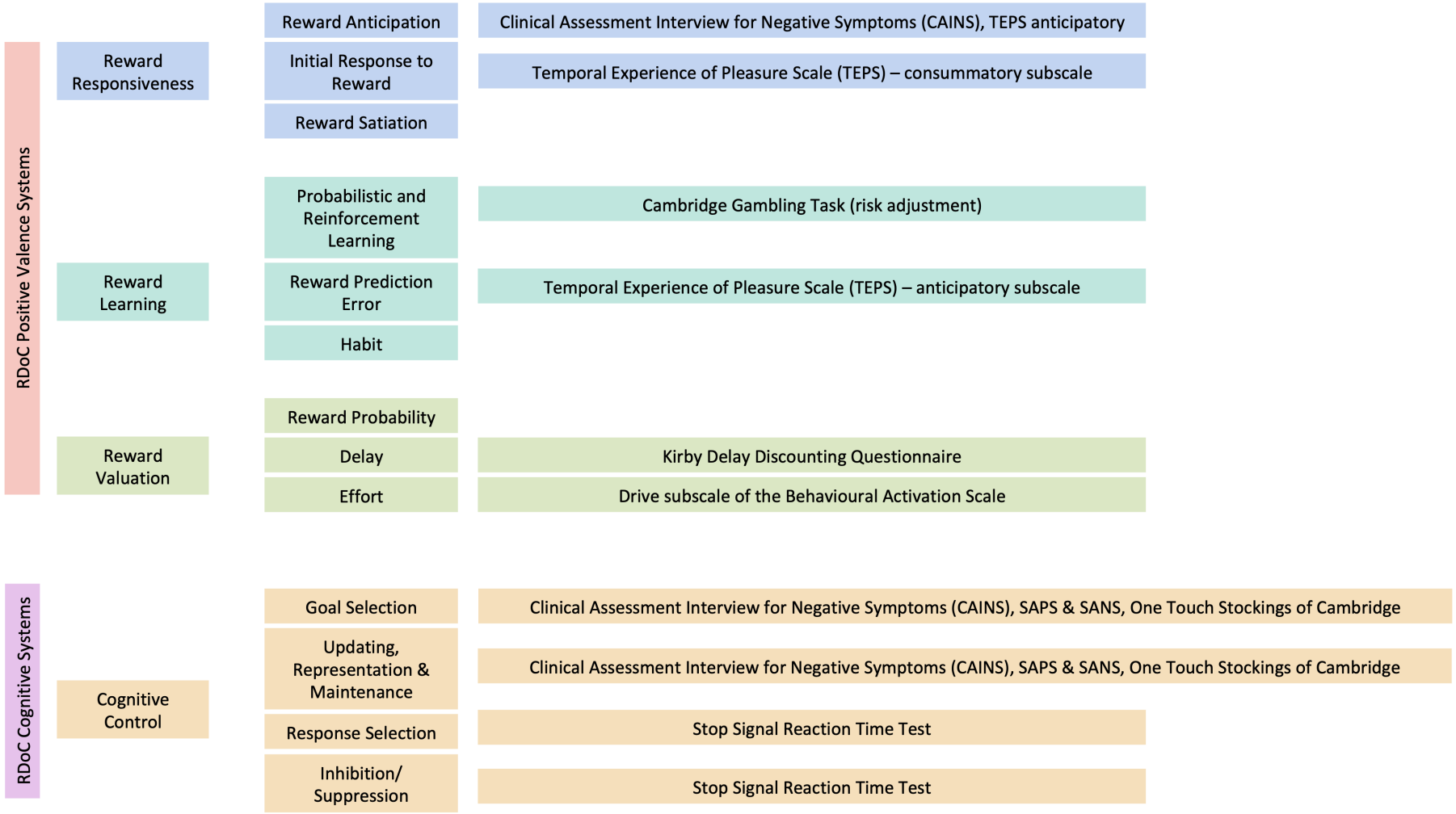

Figure 3. Mapping of phenotypic assessments to RDoC domains.

Appendix 4 – Power calculations for CNV recall study

Prior to commencing the recall of CNV carriers and CNV noncarriers described in Chapter 4, I carried out a power calculation to determine how many individuals I needed to assess. For this calculation, I chose to use cognitive data since it is this phenotype we know most about in the context of CNVs. These data came from a colleague and were subsequently published in *Biological Psychiatry*.(235) I carried out the calculation using the linear regression, two tailed setting in G*Power.(236)

Assessments in 13 CNV carriers and 27 CNV noncarriers have an 80% power, at an α of 0.05, to detect an association between CNV carrier status and cognitive function with an effect size (B) of -0.76 (main meta-analysis result in (235)).

We used this result, and what we thought was feasible in the time period, and concluded that I would try to recall 20 CNV carriers and 40 CNV noncarriers.

References

1. McGuffin P, Owen M, O'Donovan M, Thapar A, Gottesman I. Seminars in Psychiatric Genetics [Internet]. The Royal College of Psychiatrists; 1994. Available from: http://www.rcpsych.ac.uk/usefulresources/publications/seminarsseries/psychiatricgenetics.aspx

2. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry. 2016 May;57(5):585–95.

3. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60(12):1187–92.

4. Barnett JH, Smoller JW. The Genetics of Bipolar Disorder. Neuroscience. 2009 Nov 24;164(1):331–43.

5. Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. Mol Psychiatry. 2019 Apr;24(4):562–75.

6. Sullivan P, Neale M, Kendler K. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry. 2000;157(10):1552-62.

7. Lejeune J, Gautier M, Turpin R. [Study of somatic chromosomes from 9 mongoloid children]. Comptes Rendus Hebd Seances Acad Sci. 1959 Mar 16;248(11):1721–2.

8. Lubs HA. A marker X chromosome. Am J Hum Genet. 1969 May;21(3):231–44.

9. Jie X, Lonnie Z, Peter S, Stephen WS. Molecular Cytogenetics of Autism. Curr Genomics. 2004 May 1;5(4):347–64.

10. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016 Jan;17(1):9–18.

11. Bassett AS. Chromosomal aberrations and schizophrenia. Autosomes. Br J Psychiatry. 1992 Sep;161:323–34.

12. DeLisi LE, Friedrich U, Wahlstrom J, Boccio-Smith A, Forsman A, Eklund K, et al. Schizophrenia and sex chromosome anomalies. Schizophr Bull. 1994;20(3):495–505.

13. Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, et al. Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet. 2000 May 22;9(9):1415–23.

14. Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, et al. Evaluating historical candidate genes for schizophrenia. Mol Psychiatry. 2015 May;20(5):555–62.

15. MacDonald ME, Ambrose CM, Duyao MP, Myers RH, Lin C, Srinidhi L, et al. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell. 1993; 72(6): 971-83.

16. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerrière A, Vital A, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet. 2006; 38: 24-6.

17. Chartier-Harlin MC, Crawford F, Houlden H, Warren A, Hughes D, Fidani L, et al. Early-onset Alzheimer's disease caused by mutations at codon 717 of the β-amyloid precursor protein gene. Nature. 1991; 353: 844-6.

18. Goate A, Chartier-Harlin MC, Mullan M, Brown J, Crawford F, Fidani L, et al. Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. Nature. 1991; 349: 704-6.

19. Sherrington R, Rogaev EI, Liang Y, Rogaeva EA, Levesque G, Ikeda M, et al. Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. Nature. 1995 Jun 29;375(6534):754–60.

20. Rogaev EI, Sherrington R, Rogaeva EA, Levesque G, Ikeda M, Liang Y, et al. Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene. Nature. 1995 Aug 31;376(6543):775–8.

21. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008 May;32(4):381–5.

22. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018; 50: 381-9.

23. Stahl E, Breen G, Forstner A, McQuillin A, Ripke S, Consortium BDWG of the PG, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. Biorxiv [Internet]. 2018; Available from: https://www.biorxiv.org/content/early/2018/01/24/173062

24. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci [Internet]. 2019; Available from: https://doi.org/10.1038/s41593-018-0326-7

25. Meier SM, Trontti K, Purves KL, Als TD, Grove J, Laine M, et al. Genetic Variants Associated With Anxiety and Stress-Related Disorders: A Genome-Wide Association Study and Mouse-Model Study. JAMA Psychiatry. 2019 Sep 1;76(9):924–32.

26. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019 Mar;51(3):431–44.

27. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019 Jan;51(1):63–75.

28. Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45(9):984–94.

29. Common polygenic variation contributes to risk of schizophrenia that overlaps with bipolar disorder. Nature. 2009 Aug 6;460(7256):748–52.

30. Dennison CA, Legge SE, Pardiñas AF, Walters JTR. Genome-wide association studies in schizophrenia: Recent advances, challenges and future perspective. Schizophr Res [Internet]. 2019 Nov 25; Available from: http://www.sciencedirect.com/science/article/pii/S0920996419304839

31. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler AK, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet. 2013 Oct;45(10):1150–9.

32. Ni G, Moser G, Consortium SWG of the PG, Wray NR, Lee SH. Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. Am J Hum Genet. 2018 Jun 7;102(6):1185-94.

33. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236–41.

34. Allardyce J, Leonenko G, Hamshere M, Pardiñas AF, Forty L, Knott S, et al. Association Between Schizophrenia-Related Polygenic Liability and the Occurrence and Level of Mood-Incongruent Psychotic Symptoms in Bipolar Disorder. JAMA Psychiatry. 2018 Jan;75(1):28–35.

35. The Brainstorm Consortium, Anttila V, Bulik-Sullivan B, Finucane HK, Walters RK, Bras J, et al. Analysis of shared heritability in common disorders of the brain. Science [Internet]. 2018 Jun 22; 360(6395). Available from: https://science.sciencemag.org/content/360/6395/eaap8757

36. Lam M, Hill WD, Trampush JW, Yu J, Knowles E, Davies G, et al. Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia

Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. Am J Hum Genet. 2019 Aug;105(2):334–50.

37. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85–97.

38. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al. Copy number variation: New insights in genome diversity. Genome Res. 2006 Jan 8;16(8):949–61.

39. Zhang F, Gu W, Hurles ME, Lupski JR. Copy Number Variation in Human Health, Disease, and Evolution. Annu Rev Genomics Hum Genet. 2009;10:451–81.

40. Lafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004 Sep;36(9):949–51.

41. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. Am J Hum Genet. 2009 Feb;84(2):148–61.

42. Kirov G, Rees E, Walters J. What a psychiatrist needs to know about copy number variants. BJPsych Adv. 2015; 21(3): 157-63.

43. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet. 2006 Jul;7(7):552–64.

44. Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell. 2012 Mar 16;148(6):1223–41.

45. Currall BB, Chiang C, Talkowski ME, Morton CC. Mechanisms for Structural Variation in the Human Genome. Curr Genet Med Rep. 2013 Jun 1;1(2):81–90.

46. Simmons AD, Carvalho CMB, Lupski JR. What Have Studies of Genomic Disorders Taught Us About Our Genome? In: Feuk L, editor. Genomic Structural Variants: Methods and Protocols [Internet]. New York, NY: Springer; 2012. p. 1–27. (Methods in Molecular Biology). Available from: https://doi.org/10.1007/978-1-61779-507-7_1

47. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. PathoGenetics. 2008 Nov 3;1(1):4.

48. Zhang F, Seeman P, Liu P, Weterman MAJ, Gonzaga-Jauregui C, Towne CF, et al. Mechanisms for Nonrecurrent Genomic Rearrangements Associated with CMT1A or HNPP: Rare CNVs as a Cause for Missing Heritability. Am J Hum Genet. 2010 Jun 11;86(6):892–903.

49. Shprintzen RJ, Goldberg R, Golding-Kushner KJ, Marion RW. Late-onset psychosis in the velo-cardio-facial syndrome. Am J Med Genet. 1992 Jan 1;42(1):141–2.

50. Murphy KC, Jones LA, Owen MJ. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry. 1999;56(10):940–5.

51. Kirov G, Gumus D, Chen W, Norton N, Georgieva L, Sari M, et al. Comparative genome hybridization suggests a role for NRXN1 and APBA2 in schizophrenia. Hum Mol Genet. 2008 Feb 1;17(3):458–65.

52. Kirov G, Rujescu D, Ingason A, Collier DA, O'Donovan MC, Owen MJ. Neurexin 1 (NRXN1) deletions in schizophrenia. Schizophr Bull. 2009 Sep;35(5):851–4.

53. Rees E, Walters JT, Georgieva L, Isles AR, Chambert KD, Richards AL, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. Br J Psychiatry. 2014;204(2):108–14.

54. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet. 2017 Jan;49(1):27–35.

55. Rees E, Kendall K, Pardiñas AFF, Legge SEE, Pocklington A, Escott-Price V, et al. Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia. JAMA Psychiatry. 2016;73(9):963–9.

56. ISC. Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature. 2008;455(7210):237–41.

57. Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. Strong association of de novo copy number mutations with sporadic schizophrenia. Nat Genet. 2008 Jul;40(7):880–5.

58. Kirov G, Pocklington AJ, Holmans P, Ivanov D, Ikeda M, Ruderfer D, et al. De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol Psychiatry. 2012;17(2):142–53.

59. Malhotra D, McCarthy S, Michaelson JJ, Vacic V, Burdick KE, Yoon S, et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. Neuron. 2011 Dec 22;72(6):951–63.

60. Zhang D, Cheng L, Qian Y, Alliey-Rodriguez N, Kelsoe JR, Greenwood T, et al. Singleton deletions throughout the genome increase risk of bipolar disorder. Mol Psychiatry. 2009 Apr;14(4):376–80.

61. Bergen SE, O'Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, et al. Genome-wide association study in a Swedish population yields

support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol Psychiatry. 2012 Sep;17(9):880–6.

62. Priebe L, Degenhardt FA, Herms S, Haenisch B, Mattheisen M, Nieratschker V, et al. Genome-wide survey implicates the influence of copy number variants (CNVs) in the development of early-onset bipolar disorder. Mol Psychiatry. 2012 Apr;17(4):421–32.

63. Georgieva L, Rees E, Moran JL, Chambert KD, Milanova V, Craddock N, et al. De novo CNVs in bipolar affective disorder and schizophrenia. Hum Mol Genet. 2014 Dec 15;23(24):6677–83.

64. Grozeva D, Kirov G, Ivanov D, Jones IR, Jones L, Green EK, et al. Rare copy number variants: a point of rarity in genetic risk for bipolar disorder and schizophrenia. Arch Gen Psychiatry. 2010 Apr;67(4):318–27.

65. McQuillin A, Bass N, Anjorin A, Lawrence J, Kandaswamy R, Lydall G, et al. Analysis of genetic deletions and duplications in the University College London bipolar disorder case control sample. Eur J Hum Genet EJHG. 2011 May;19(5):588–92.

66. Grozeva D, Kirov G, Conrad DF, Barnes CP, Hurles M, Owen MJ, et al. Reduced burden of very large and rare CNVs in bipolar affective disorder. Bipolar Disord. 2013;15(8):893–8.

67. Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D, et al. Copy number variation in bipolar disorder. Mol Psychiatry. 2016;21(1):89–93.

68. Charney AW, Stahl EA, Green EK, Chen C-Y, Moran JL, Chambert K, et al. Contribution of Rare Copy Number Variants to Bipolar Disorder Risk Is Limited to Schizoaffective Cases. Biol Psychiatry. 2019 Jul 15;86(2):110–9.

69. Glessner JT, Wang K, Sleiman PM, Zhang H, Kim CE, Flory JH, et al. Duplication of the SLIT3 locus on 5q35.1 predisposes to major depressive disorder. PLoS One. 2010;5(12):e15463.

70. O'Dushlaine C, Ripke S, Ruderfer DM, Hamilton SP, Fava M, Iosifescu DV, et al. Rare copy number variation in treatment-resistant major depressive disorder. Biol Psychiatry. 2014;76(7):536–41.

71. Degenhardt F, Priebe L, Herms S, Mattheisen M, Mühleisen TW, Meier S, et al. Association between copy number variants in 16p11.2 and major depressive disorder in a German case-control sample. Am J Med Genet B Neuropsychiatr Genet. 2012;159B(3):263–73.

72. Rucker JJ, Breen G, Pinto D, Pedroso I, Lewis CM, Cohen-Woods S, et al. Genome-wide association analysis of copy number variation in recurrent depressive disorder. Mol Psychiatry. 2013;18(2):183–9.

73. Rucker JJ, Tansey KE, Rivera M, Pinto D, Cohen-Woods S, Uher R, et al. Phenotypic Association Analyses With Copy Number Variation in Recurrent Depressive Disorder. Biol Psychiatry. 2016;79(4):329–36.

74. Zhang X, Abdellaoui A, Rucker J, de Jong S, Potash JB, Weissman MM, et al. Genome-wide Burden of Rare Short Deletions Is Enriched in Major Depressive Disorder in Four Cohorts. Biol Psychiatry. 2019 Jun 15;85(12):1065–73.

75. Perlis RH, Ruderfer D, Hamilton SP, Ernst C. Copy number variation in subjects with major depressive disorder who attempted suicide. PLoS One. 2012;7(9):e46315.

76. Tansey KE, Rucker JJ, Kavanagh DH, Guipponi M, Perroud N, Bondolfi G, et al. Copy number variants and therapeutic response to antidepressant medication in major depressive disorder. Pharmacogenomics J. 2014;14(4):395–9.

77. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J Affect Disord. 2015 Feb 1;172:96–102.

78. Cai N, Choi KW, Fried EI. Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. Hum Mol Genet [Internet]. 2020; Available from: https://academic.oup.com/hmg/article/doi/10.1093/hmg/ddaa115/5860824

79. Cai N, Revez JA, Adams MJ, Andlauer TFM, Breen G, Byrne EM, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. Nat Genet. 2020 Apr;52(4):437–47.

80. Williams NM, Franke B, Mick E, Anney RJ, 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.

81. 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 Lond Engl. 2010 Oct 23;376(9750):1401–8.

82. Girirajan S, Brkanac Z, Coe BP, Baker C, Vives L, Vu TH, et al. Relative burden of large CNVs on a range of neurodevelopmental phenotypes. PLoS Genet. 2011;7(11):e1002334.

83. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015 Sep 23;87(6):1215–33.

84. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011;43(9):838–46.

85. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46(10):1063–71.

86. McDonald-McGinn DM, Sullivan KE, Marino B, Philip N, Swillen A, Vorstman JAS, et al. 22q11.2 deletion syndrome. Nat Rev Dis Primer. 2015 19;1:15071.

87. Hoeffding LK, Trabjerg BB, Olsen L, Mazin W, Sparsø T, Vangkilde A, et al. Risk of Psychiatric Disorders Among Individuals With the 22q11.2 Deletion or Duplication: A Danish Nationwide, Register-Based Study. JAMA Psychiatry. 2017 01;74(3):282–90.

88. Boot E, Bassett AS, Marras C. 22q11.2 Deletion Syndrome-Associated Parkinson's Disease. Mov Disord Clin Pract. 2019 Jan;6(1):11–6.

89. Cox DM, Butler MG. The 15q11.2 BP1–BP2 Microdeletion Syndrome: A Review. Int J Mol Sci. 2015 Feb 13;16(2):4068–82.

90. Jønch AE, Douard E, Moreau C, Van Dijck A, Passeggeri M, Kooy F, et al. Estimating the effect size of the 15Q11.2 BP1-BP2 deletion and its contribution to neurodevelopmental symptoms: recommendations for practice. J Med Genet. 2019;56(10):701–10.

91. Allach El Khattabi L, Heide S, Caberg J-H, Andrieux J, Doco Fenzy M, Vincent-Delorme C, et al. 16p13.11 microduplication in 45 new patients: refined clinical significance and genotype-phenotype correlations. J Med Genet. 2020 May;57(5):301–7.

92. Nagamani SCS, Erez A, Bader P, Lalani SR, Scott DA, Scaglia F, et al. Phenotypic manifestations of copy number variation in chromosome 16p13.11. Eur J Hum Genet. 2011 Mar;19(3):280–6.

93. Kuang S-Q, Guo D-C, Prakash SK, McDonald M-LN, Johnson RJ, Wang M, et al. Recurrent chromosome 16p13.1 duplications are a risk factor for aortic dissections. PLoS Genet. 2011 Jun;7(6):e1002118.

94. Tropeano M, Ahn JW, Dobson RJB, Breen G, Rucker J, Dixit A, et al. Male-biased autosomal effect of 16p13.11 copy number variation in neurodevelopmental disorders. PloS One. 2013;8(4):e61365.

95. Kirov G, Rees E, Walters JT, Escott-Price V, Georgieva L, Richards AL, et al. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry. 2014;75(5):378–85.

96. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. 2012 Oct 4;367(14):1321–31.

97. Olsen L, Sparsø T, Weinsheimer SM, Dos Santos MBQ, Mazin W, Rosengren A, et al. Rearrangements in the 22q11.2 Region: Prevalence and Population-Based Risk for Neuropsychiatric and Developmental Disorders. Lancet Psychiatry. 2018 Jul;5(7):573–80.

98. Fahed AC, Wang M, Homburger JR, Patel AP, Bick AG, Neben CL, et al. Polygenic background modifies penetrance of monogenic variants conferring risk for coronary artery disease, breast cancer, or colorectal cancer. medRxiv. 2019 Jan 1;19013086.

99. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL, et al. Common alleles contribute to schizophrenia in CNV carriers. Mol Psychiatry. 2016;21(8):1085–9.

100. Cleynen I, Engchuan W, Hestand MS, Heung T, Holleman AM, Johnston HR, et al. Genetic contributors to risk of schizophrenia in the presence of a 22q11.2 deletion. Mol Psychiatry. 2020 Feb 3;1–15.

101. Capute AJ, Palmer FB. A pediatric overview of the spectrum of developmental disabilities. J Dev Behav Pediatr JDBP. 1980 Jun;1(2):66–9.

102. Lilienfeld AM, Pasamanick B, Rogers M. Relationship Between Pregnancy Experience and the Development of Certain Neuropsychiatric Disorders in Childhood. Am J Public Health Nations Health. 1955 May;45(5 Pt 1):637–43.

103. Weinberger DR. Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry. 1987 Jul;44(7):660–9.

104. Murray RM, Lewis SW. Is schizophrenia a neurodevelopmental disorder? Br Med J Clin Res Ed. 1988 Jan 2;296(6614):63.

105. Craddock N, Owen MJ. The Kraepelinian dichotomy - going, going... but still not gone. Br J Psychiatry J Ment Sci. 2010 Feb;196(2):92–5.

106. Owen MJ. Intellectual disability and major psychiatric disorders: a continuum of neurodevelopmental causality. Br J Psychiatry J Ment Sci. 2012 Apr;200(4):268–9.

107. Moreno-De-Luca A, Myers SM, Challman TD, Moreno-De-Luca D, Evans DW, Ledbetter DH. Developmental brain dysfunction: revival and expansion of old concepts based on new genetic evidence. Lancet Neurol. 2013;12(4):406–14.

108. Owen MJ, O'Donovan MC. Schizophrenia and the neurodevelopmental continuum:evidence from genomics. World Psychiatry. 2017 Oct;16(3):227–35.

109. Owen MJ, O'Donovan MC, Thapar A, Craddock N. Neurodevelopmental hypothesis of schizophrenia. Br J Psychiatry. 2011;198(3):173–5.

110. Rees E, Moskvina V, Owen MJ, O'Donovan MC, Kirov G. De novo rates and selection of schizophrenia-associated copy number variants. Biol Psychiatry. 2011;70(12):1109–14.

111. Doherty JL, Owen MJ. Genomic insights into the overlap between psychiatric disorders: implications for research and clinical practice. Genome Med. 2014 Apr 28;6(4):29.

112. Hilborn JV, Strauss E, Hultsch DF, Hunter MA. Intraindividual variability across cognitive domains: investigation of dispersion levels and performance profiles in older adults. J Clin Exp Neuropsychol. 2009 May;31(4):412–24.

113. Harada CN, Natelson Love MC, Triebel KL. Normal cognitive aging. Clin Geriatr Med. 2013 Nov;29(4):737–52.

114. Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015 Jul;47(7):702–9.

115. Chabris CF, Hebert BM, Benjamin DJ, Beauchamp J, Cesarini D, van der Loos M, et al. Most reported genetic associations with general intelligence are probably false positives. Psychol Sci. 2012;23(11):1314–23.

116. Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK, Brion MJ, et al. Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Mol Psychiatry. 2014;19(2):253–8.

117. Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, et al. Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry. 2011 Oct;16(10):996–1005.

118. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018 Jul;50(7):912–9.

119. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nat Commun. 2018 29;9(1):2098.

120. WHO. International Classification of Diseases 10th Revision [Internet]. 2016. Available from: http://apps.who.int/classifications/icd10/browse/2010/en

121. Kodituwakku PW. Neurocognitive Profile In Children With Fetal Alcohol Spectrum Disorders. Dev Disabil Res Rev. 2009;15(3):218–24.

122. Strømme P, Diseth TH. Prevalence of psychiatric diagnoses in children with mental retardation: data from a population-based study. Dev Med Child Neurol. 2000 Apr;42(4):266–70.

123. Pellicano E. The Development of Core Cognitive Skills in Autism: A 3-Year Prospective Study. Child Dev. 2010;81(5):1400–16.

124. Velikonja T, Fett A-K, Velthorst E. Patterns of Nonsocial and Social Cognitive Functioning in Adults With Autism Spectrum Disorder: A Systematic Review and Meta-analysis. JAMA Psychiatry. 2019 Feb 1;76(2):135–51.

125. Bucaille A, Grandgeorge M, Degrez C, Mallégol C, Cam P, Botbol M, et al. Cognitive profile in adults with Asperger syndrome using WAIS-IV: Comparison to typical adults. Res Autism Spectr Disord. 2016;21:1–9.

126. Coghill DR, Seth S, Matthews K. A comprehensive assessment of memory, delay aversion, timing, inhibition, decision making and variability in attention deficit hyperactivity disorder: advancing beyond the three-pathway models. Psychol Med. 2014 Jul;44(9):1989–2001.

127. Willcutt EG, Doyle AE, Nigg JT, Faraone SV, Pennington BF. Validity of the executive function theory of attention-deficit/hyperactivity disorder: a meta-analytic review. Biol Psychiatry. 2005 Jun 1;57(11):1336–46.

128. Sergeant JA. Modeling attention-deficit/hyperactivity disorder: a critical appraisal of the cognitive-energetic model. Biol Psychiatry. 2005 Jun 1;57(11):1248–55.

129. Luman M, Tripp G, Scheres A. Identifying the neurobiology of altered reinforcement sensitivity in ADHD: a review and research agenda. Neurosci Biobehav Rev. 2010 Apr;34(5):744–54.

130. Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Primer. 2015 Aug 6;1(1):1–23.

131. Fett A-KJ, Viechtbauer W, Dominguez M-G, Penn DL, van Os J, Krabbendam L. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: a meta-analysis. Neurosci Biobehav Rev. 2011 Jan;35(3):573–88.

132. Heinrichs RW, Zakzanis KK. Neurocognitive deficit in schizophrenia: a quantitative review of the evidence. Neuropsychology. 1998 Jul;12(3):426–45.

133. Aleman A, Hijman R, de Haan EHF, Kahn RS. Memory Impairment in Schizophrenia: A Meta-Analysis. Am J Psychiatry. 1999 Sep 1;156(9):1358–66.

134. Fioravanti M, Carlone O, Vitale B, Cinti ME, Clare L. A meta-analysis of cognitive deficits in adults with a diagnosis of schizophrenia. Neuropsychol Rev. 2005 Jun;15(2):73–95.

135. Fatouros-Bergman H, Cervenka S, Flyckt L, Edman G, Farde L. Meta-analysis of cognitive performance in drug-naïve patients with schizophrenia. Schizophr Res. 2014 Sep;158(1–3):156–62.

136. Bora E, Yücel M, Pantelis C. Theory of mind impairment: a distinct trait-marker for schizophrenia spectrum disorders and bipolar disorder? Acta Psychiatr Scand. 2009 Oct;120(4):253–64.

137. Milev P, Ho B-C, Arndt S, Andreasen NC. Predictive values of neurocognition and negative symptoms on functional outcome in schizophrenia: a longitudinal first-episode study with 7-year follow-up. Am J Psychiatry. 2005 Mar;162(3):495–506.

138. Tabarés-Seisdedos R, Balanzá-Martínez V, Sánchez-Moreno J, Martinez-Aran A, Salazar-Fraile J, Selva-Vera G, et al. Neurocognitive and clinical predictors of functional outcome in patients with schizophrenia and bipolar I disorder at one-year follow-up. J Affect Disord. 2008 Aug;109(3):286–99.

139. Malhi GS, Mann JJ. Depression. The Lancet. 2018 Nov 24;392(10161):2299–312.

140. Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. BMC Med. 2011;9:90.

141. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. Annu Rev Public Health. 2013;34:119–38.

142. Rice SM, Fallon BJ, Aucote HM, Möller-Leimkühler A, Treeby MS, Amminger GP. Longitudinal sex differences of externalising and internalising depression symptom trajectories: Implications for assessment of depression in men from an online study. Int J Soc Psychiatry. 2015 May;61(3):236–40.

143. Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. Int J Methods Psychiatr Res. 2003;12(1):3–21.

144. Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, et al. Cross-national epidemiology of major depression and bipolar disorder. JAMA. 1996 Jul 24;276(4):293–9.

145. Bromet EJ, Gluzman SF, Paniotto VI, Webb CPM, Tintle NL, Zakhozha V, et al. Epidemiology of psychiatric and alcohol disorders in Ukraine: findings from the Ukraine World Mental Health survey. Soc Psychiatry Psychiatr Epidemiol. 2005 Sep;40(9):681–90.

146. Kessler RC, Birnbaum HG, Shahly V, Bromet E, Hwang I, McLaughlin KA, et al. Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: results from the WHO World Mental Health Survey Initiative. Depress Anxiety. 2010 Apr;27(4):351–64.

147. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care. 2001 Jun;24(6):1069–78.

148. Buist-Bouwman MA, de Graaf R, Vollebergh W a. M, Ormel J. Comorbidity of physical and mental disorders and the effect on work-loss days. Acta Psychiatr Scand. 2005 Jun;111(6):436–43.

149. Chapman DP, Perry GS, Strine TW. The vital link between chronic disease and depressive disorders. Prev Chronic Dis. 2005 Jan;2(1):A14.

150. Derogatis LR, Morrow GR, Fetting J, Penman D, Piasetsky S, Schmale AM, et al. The prevalence of psychiatric disorders among cancer patients. JAMA. 1983 Feb 11;249(6):751–7.

151. McWilliams LA, Cox BJ, Enns MW. Mood and anxiety disorders associated with chronic pain: an examination in a nationally representative sample. Pain. 2003 Nov;106(1–2):127–33.

152. Nemeroff CB, Musselman DL, Evans DL. Depression and cardiac disease. Depress Anxiety. 1998;8 Suppl 1:71–9.

153. Van der Kooy K, van Hout H, Marwijk H, Marten H, Stehouwer C, Beekman A. Depression and the risk for cardiovascular diseases: systematic review and meta analysis. Int J Geriatr Psychiatry. 2007 Jul;22(7):613–26.

154. Wulsin LR, Singal BM. Do depressive symptoms increase the risk for the onset of coronary disease? A systematic quantitative review. Psychosom Med. 2003 Apr;65(2):201–10.

155. Pratt LA, Ford DE, Crum RM, Armenian HK, Gallo JJ, Eaton WW. Depression, psychotropic medication, and risk of myocardial infarction. Prospective data from the Baltimore ECA follow-up. Circulation. 1996 Dec 15;94(12):3123–9.

156. Scherrer JF, Virgo KS, Zeringue A, Bucholz KK, Jacob T, Johnson RG, et al. Depression increases risk of incident myocardial infarction among Veterans Administration patients with rheumatoid arthritis. Gen Hosp Psychiatry. 2009 Aug;31(4):353–9.

157. Ohira T, Iso H, Satoh S, Sankai T, Tanigawa T, Ogawa Y, et al. Prospective study of depressive symptoms and risk of stroke among japanese. Stroke. 2001 Apr;32(4):903–8.

158. Carnethon MR, Kinder LS, Fair JM, Stafford RS, Fortmann SP. Symptoms of depression as a risk factor for incident diabetes: findings from the National Health and Nutrition Examination Epidemiologic Follow-up Study, 1971-1992. Am J Epidemiol. 2003 Sep 1;158(5):416–23.

159. Gross AL, Gallo JJ, Eaton WW. Depression and cancer risk: 24 years of follow-up of the Baltimore Epidemiologic Catchment Area sample. Cancer Causes Control CCC. 2010 Feb;21(2):191–9.

160. Gillen R, Tennen H, McKee TE, Gernert-Dott P, Affleck G. Depressive symptoms and history of depression predict rehabilitation efficiency in stroke patients. Arch Phys Med Rehabil. 2001 Dec;82(12):1645–9.

161. Mancuso CA, Rincon M, McCulloch CE, Charlson ME. Self-efficacy, depressive symptoms, and patients' expectations predict outcomes in asthma. Med Care. 2001 Dec;39(12):1326–38.

162. Peyrot M, Rubin RR. Levels and risks of depression and anxiety symptomatology among diabetic adults. Diabetes Care. 1997 Apr;20(4):585–90.

163. Breitbart W, Rosenfeld B, Pessin H, Kaim M, Funesti-Esch J, Galietta M, et al. Depression, hopelessness, and desire for hastened death in terminally ill patients with cancer. JAMA. 2000 Dec 13;284(22):2907–11.

164. Cluley S, Cochrane GM. Psychological disorder in asthma is associated with poor control and poor adherence to inhaled steroids. Respir Med. 2001 Jan;95(1):37–9.

165. Ziegelstein RC, Fauerbach JA, Stevens SS, Romanelli J, Richter DP, Bush DE. Patients with depression are less likely to follow recommendations to reduce cardiac risk during recovery from a myocardial infarction. Arch Intern Med. 2000 Jun 26;160(12):1818–23.

166. Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J, et al. Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. Biol Psychiatry. 2017;82(2):103–10.

167. Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V, et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. Br J Psychiatry. 2019; 214(5): 297-304.

168. Dittwald P, Gambin T, Szafranski P, Li J, Amato S, Divon MY, et al. NAHR-mediated copy-number variants in a clinical population:

mechanistic insights into both genomic disorders and Mendelizing traits. Genome Res. 2013;23(9):1395–409.

169. Lal D, Ruppert AK, Trucks H, Schulz H, de Kovel CG, Kasteleijn-Nolst Trenité D, et al. Burden analysis of rare microdeletions suggests a strong impact of neurodevelopmental genes in genetic generalised epilepsies. PLoS Genet. 2015;11(5):e1005226.

170. Bassett AS, Chow EW, Husted J, Weksberg R, Caluseriu O, Webb GD, et al. Clinical features of 78 adults with 22q11 Deletion Syndrome. Am J Med Genet A. 2005;138(4):307–13.

171. Jawad AF, McDonald-Mcginn DM, Zackai E, Sullivan KE. Immunologic features of chromosome 22q11.2 deletion syndrome (DiGeorge syndrome/velocardiofacial syndrome). J Pediatr. 2001;139(5):715–23.

172. Burnside RD, Pasion R, Mikhail FM, Carroll AJ, Robin NH, Youngs EL, et al. Microdeletion/microduplication of proximal 15q11.2 between BP1 and BP2: a susceptibility region for neurological dysfunction including developmental and language delay. Hum Genet. 2011;130(4):517–28.

173. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature. 2014;505(7483):361–6.

174. Männik K, Mägi R, Macé A, Cole B, Guyatt AL, Shihab HA, et al. Copy number variations and cognitive phenotypes in unselected populations. JAMA. 2015;313(20):2044–54.

175. Wain LV, Shrine N, Miller S, Jackson VE, Ntalla I, Soler Artigas M, et al. Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. Lancet Respir Med. 2015;3(10):769–81.

176. Affymetrix. Affymetrix Power Tools [Internet]. Vol. 2016. 2016. Available from: http://media.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx

177. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665–74.

178. Rousseeuw PJ, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T et al. Robustbase: basic robust statistics. 2009. URL http://CRAN. R-project. org/package= robustbase.

179. Hubert M, Rousseeuw PJ, Verdonck T. A deterministic algorithm for robust location and scatter. J Comput Graph Stat. 2012; 21(3): 618-37.

180. Legge SE, Jones HJ, Kendall KM, Pardiñas AF, Menzies G, Bracher-Smith M, et al. Association of Genetic Liability to Psychotic Experiences With Neuropsychotic Disorders and Traits. JAMA Psychiatry. 2019 Dec 1;76(12):1256-65.

181. ONS. Standard occupational classification 2000. London: The Stationery Office; 2000.

182. Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S, et al. The role of rare copy number variants in depression. JAMA Psychiatry. 2019; 76(8): 818-25.

183. Rees E, Walters JT, Chambert KD, O'Dushlaine C, Szatkiewicz J, Richards AL, et al. CNV analysis in a large schizophrenia sample implicates deletions at 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. Hum Mol Genet. 2014;23(6):1669–76.

184. Consortium C-DG of the PG. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381(9875):1371–9.

185. Cooper SA, McLean G, Guthrie B, McConnachie A, Mercer S, Sullivan F, et al. Multiple physical and mental health comorbidity in adults with intellectual disabilities: population-based cross-sectional analysis. BMC Fam Pr. 2015;16:110.

186. Buckley PF, Miller BJ, Lehrer DS, Castle DJ. Psychiatric comorbidities and schizophrenia. Schizophr Bull. 2009;35(2):383–402.

187. Howard DM, Adams MJ, Shirali M, Clarke TK, Marioni RE, Davies G, et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nat Commun. 2018; 9(1): 1470.

188. Smith DJ, Nicholl BI, Cullen B, Martin D, Ul-Haq Z, Evans J, et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. PLoS One. 2013;8(11):e75362.

189. Spijker J, de Graaf R, Bijl R, Beekman A, Ormel J, Nolen W. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry. 2002;181(3):208–13.

190. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

191. Lorant V, Deliège D, Eaton W, Robert A, Philippot P, Ansseau M. Socioeconomic inequalities in depression: a meta-analysis. Am J Epidemiol. 2003;157(2):98–112.

192. Egede LE. Major depression in individuals with chronic medical disorders: prevalence, correlates and association with health resource utilization, lost productivity and functional disability. Gen Hosp Psychiatry. 2007;29(5):409–16.

193. Fluharty M, Taylor AE, Grabski M, Munafò MR. The association of cigarette smoking with depression and anxiety: A systematic review. 2017; 19(1): 3-13.

194. Boden JM, Fergusson DM. Alcohol and depression. 2011; 106(5): 906-14.

195. Davies G, Marioni RE, Liewald DC, Hill WD, Hagenaars SP, Harris SE, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). Mol Psychiatry. 2016;21(6):758–67.

196. Crawford K, Bracher-Smith M, Kendall KM, Rees E, Pardinas AF, Einon M, et al. Medical consequences of pathogenic CNVs in adults: Analysis of the UK Biobank. J Med Genet. 2018 Jan; 56(3): 131-8.

197. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw. 2010;48(2):1–36.

198. Han J, Walters JT, Kirov G, Pocklington A, Escott-Price V, Owen MJ, et al. Gender differences in CNV burden do not confound schizophrenia CNV associations. Sci Rep. 2016;6:25986.

199. Martin J, Tammimies K, Karlsson R, Lu Y, Larsson H, Lichtenstein P, et al. Copy number variation and neuropsychiatric problems in females and males in the general population. Am J Med Genet B Neuropsychiatr Genet. 2019;180(6):341–50.

200. Feighan S-M, Hughes M, Maunder K, Roche E, Gallagher L. A profile of mental health and behaviour in Prader–Willi syndrome. J Intellect Disabil Res. 2020;64(2):158–69.

201. Soni S, Whittington J, Holland AJ, Webb T, Maina E, Boer H, et al. The course and outcome of psychiatric illness in people with Prader–Willi syndrome: implications for management and treatment. J Intellect Disabil Res. 2007;51(1):32–42.

202. Whittington J, Holland A. A review of psychiatric conceptions of mental and behavioural disorders in Prader-Willi syndrome. Neurosci Biobehav Rev. 2018 Dec 1;95:396–405.

203. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Heal. 2016;9:211–7.

204. Maier W, Gänsicke M, Gater R, Rezaki M, Tiemens B, Urzúa RF. Gender differences in the prevalence of depression: a survey in primary care. J Affect Disord. 1999;53(3):241–52.

205. Fusar-Poli P, Papanastasiou E, Stahl D, Rocchetti M, Carpenter W, Shergill S, et al. Treatments of Negative Symptoms in Schizophrenia: Meta-Analysis of 168 Randomized Placebo-Controlled Trials. Schizophr Bull. 2015 Jul;41(4):892–9.

206. Strauss GP, Cohen AS. A Transdiagnostic Review of Negative Symptom Phenomenology and Etiology. Schizophr Bull. 2017 Jul;43(4):712–9.

207. Strauss GP, Nuñez A, Ahmed AO, Barchard KA, Granholm E, Kirkpatrick B, et al. The Latent Structure of Negative Symptoms in Schizophrenia. JAMA Psychiatry. 2018 Dec 1;75(12):1271–9.

208. NIMH » Domain: Positive Valence Systems [Internet]. Available from: https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/positive-valence-systems.shtml

209. Association AP. Diagnostic and Statistical Manual of Mental Disorders. Washington DC: American Psychiatric Press; 1994.

210. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. Arch Gen Psychiatry. 1990;47(6):589–93.

211. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry. 1998;59 Suppl 20:22-33.

212. IBM Corp. Released 2019. IBM SPSS for Mac, Version 26.0. Armonk, NY: IBM Corp.

213. Reiersen AM. Collection of developmental history in the evaluation of schizophrenia spectrum disorders. Scand J Child Adolesc Psychiatry Psychol. 2016;4(1):36–43.

214. Horan WP, Kring AM, Gur RE, Reise SP, Blanchard JJ. Development and psychometric validation of the Clinical Assessment Interview for Negative Symptoms (CAINS). Schizophr Res. 2011 Nov;132(2–3):140–5.

215. Gard DE, Gard MG, Kring AM, John OP. Anticipatory and consummatory components of the experience of pleasure: A scale development study. J Res Personal. 2006 Dec 1;40(6):1086–102.

216. Kirby KN, Petry NM, Bickel WK. Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. J Exp Psychol Gen. 1999;128(1):78–87.

217. Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. J Pers Soc Psychol. 1994;67(2):319–33.

218. Foley C, Heron EA, Harold D, Walters J, Owen M, O'Donovan M, et al. Identifying schizophrenia patients who carry pathogenic genetic copy number variants using standard clinical assessment: retrospective cohort study. Br J Psychiatry. 2020 May;216(5):275–9.

219. Jones HJ, Stergiakouli E, Tansey KE, Hubbard L, Heron J, Cannon M, et al. Phenotypic Manifestation of Genetic Risk for Schizophrenia During Adolescence in the General Population. JAMA Psychiatry. 2016;73(3):221–8.

220. Xavier RM, Dungan JR, Keefe RSE, Vorderstrasse A. Polygenic signal for symptom dimensions and cognitive performance in patients with chronic schizophrenia. Schizophr Res Cogn. 2018 Jun;12:11–9.

221. Jonas KG, Lencz T, Li K, Malhotra AK, Perlman G, Fochtmann LJ, et al. Schizophrenia polygenic risk score and 20-year course of illness in psychotic disorders. Transl Psychiatry. 2019 14;9(1):300.

222. Foss-Feig JH, McPartland JC, Anticevic A, Wolf J. Re-conceptualizing ASD Within a Dimensional Framework: Positive, Negative, and Cognitive Feature Clusters. J Autism Dev Disord. 2016 Jan;46(1):342–51.

223. Zarrei M, Burton CL, Engchuan W, Young EJ, Higginbotham EJ, MacDonald JR, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. NPJ Genomic Med. 2019;4:26.

224. Tammimies K, Li D, Rabkina I, Stamouli S, Becker M, Nicolaou V, et al. Association between Copy Number Variation and Response to Social Skills Training in Autism Spectrum Disorder. Sci Rep. 2019 Jul 8;9(1):9810.

225. Rikhy S, Tough S, Trute B, Benzies K, Kehler H, Johnston DW. Gauging knowledge of developmental milestones among Albertan adults: a cross-sectional survey. BMC Public Health. 2010 Apr 8;10(1):183.

226. Rice F, Riglin L, Thapar AK, Heron J, Anney R, O'Donovan MC, et al. Characterizing Developmental Trajectories and the Role of Neuropsychiatric Genetic Risk Variants in Early-Onset Depression. JAMA Psychiatry. 2019 Mar 1;76(3):306–13.

227. Davis KAS, Coleman JRI, Adams M, Allen N, Breen G, Cullen B, et al. Mental Health in UK Biobank Revised. medRxiv. 2019 Aug 5;19001214.

228. Lee C-H, Liu C-M, Wen C-C, Chang S-M, Hwu H-G. Genetic copy number variants in sib pairs both affected with schizophrenia. J Biomed Sci. 2010 Jan 11;17(1):2.

229. Bergen SE, Ploner A, Howrigan D, CNV Analysis Group and the Schizophrenia Working Group of the Psychiatric Genomics Consortium, O'Donovan MC, Smoller JW, et al. Joint Contributions of Rare Copy

Number Variants and Common SNPs to Risk for Schizophrenia. Am J Psychiatry. 2019 Jan;176(1):29–35.

230. Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, et al. A recurrent 16p12.1 microdeletion suggests a two-hit model for severe developmental delay. Nat Genet. 2010 Mar;42(3):203–9.

231. Pizzo L, Jensen M, Polyak A, Rosenfeld JA, Mannik K, Krishnan A, et al. Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. Genet Med Off J Am Coll Med Genet. 2019;21(4):816–25.

232. Mazina V, Gerdts J, Trinh S, Ankenman K, Ward T, Dennis MY, et al. Interactive effects of copy number variation and maternal infection on autism impairment. J Dev Behav Pediatr JDBP. 2015;36(2):61–7.

233. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 2017;186(9):1026–34.

234. Batty GD, Gale CR, Kivimäki M, Deary IJ, Bell S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. BMJ [Internet]. 2020 Feb 12; 368. Available from: https://www.bmj.com/content/368/bmj.m131

235. Hubbard L, Rees E, Morris DW, Lynham AJ, Richards AL, Pardiñas AF, et al. Rare Copy Number Variants Are Associated With Poorer Cognition in Schizophrenia. Biol Psychiatry. 2020 Dec 19; S0006-3223(20)32117-X. doi: 10.1016/j.biopsych.2020.11.025.

236. Faul F, Erdfelder E, Buchner A, Lang A-G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods. 2009 Nov;41(4):1149–60.