Evaluating the Joint Effects of Recurrent Copy Number Variants and Polygenic Scores on the Risk of Psychiatric Disorders in the iPSYCH2015 Case-Cohort Sample

Morteza Vaez[1, 2], Simone Montalbano[1,2], Ryan Waples[1,2] , Morten Dybdahl Krebs[1,2], Kajsa-Lotta Georgii Hellberg[1,2], Jesper Gådin[1,2], Jonas Bybjerg-Grauholm[2,3], Preben B. Mortensen[2,4], Anders D. Børglum[2,5,6], Merete Nordentoft[2,7], Daniel H. Geschwind[8-12], Dorte Helenius[1,2], Thomas Werge[1,2,13,14], Andrew J. Schork[1,2,13,15], Andrés Ingason[1,2,13]

1. Institute of Biological Psychiatry, Mental Health Services, Copenhagen University Hospital, DK-4000 Roskilde, Denmark
2. The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH), Copenhagen and Aarhus, Denmark
3.Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, DK-2300 Copenhagen, Denmark
4. National Centre for Register-based Research, Aarhus University, DK-8210 Aarhus, Denmark
5. Department of Biomedicine—Human Genetics and the iSEQ Center, Aarhus University, DK-8000 Aarhus, Denmark
6. Center for Genomics and Personalized Medicine, DK-8000 Aarhus, Denmark
7. Mental Health Center Copenhagen, Copenhagen University Hospital, DK-2400 Copenhagen, Denmark
8. Department of Neurology, University of California, Los Angeles, CA, USA
9. Department of Human Genetics, University of California, Los Angeles, CA, USA
10. Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA
11. Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine at UCLA, University of California, Los Angeles, CA, USA
12. Center for Human Development, University of California, San Diego, CA, USA.
13. Lundbeck Foundation Center for GeoGenetics, GLOBE Institute, University of Copenhagen, DK-1350 Copenhagen, Denmark
14. Department of Clinical Medicine, University of Copenhagen, DK-2200 Copenhagen, Denmark.
15. Neurogenomics Division, The Translational Genomics Research Institute (TGEN), Phoenix, AZ, USA

Abstract:

The impact of rare recurrent copy number variants (rCNVs) and polygenic background attributed to common variants, on the risk of psychiatric disorders is well-established in separate studies. However, it remains unclear how polygenic background modulates the effect of rCNVs. Using the population-representative iPSYCH2015 case-cohort sample (N=96,599), we investigated the association between absolute risk of psychiatric disorders and carriage of rCNVs and polygenic scores (PGS), as well as the interaction effect between the two on disease risk.

Carriers of rCNVs with higher gene constraint scores had an increased absolute risk for autism, ADHD, and schizophrenia, but not depression, whereas an increase in PGS for each respective disorder was associated with higher absolute risk across all four disorders. Similarly, elevated absolute risks were observed with the increase of both PGS and gene constraints of rCNVs except in the case of depression. In contrast to some previous case-control studies, our joint analysis of rCNV groups and PGS revealed no indication of significant interactive effect between these two factors on disease risk. Also, we found no significant interactions of PGS with any of the most common individual rCNVs, except in the case of 16p13.11 duplication, which was found to attenuate the effect of ADHD-PGS on the absolute risk of ADHD.

This study advances our understanding of the interplay between rare and common important genetic risk factors for major psychiatric disorders and sheds light on the importance of population-based samples in implementing precision medicine.

Introduction:

Genetic studies have consistently shown substantial heritability for psychiatric disorders involving both common and rare variants across the human genome. [1-3] The ongoing advances in genome-wide association studies (GWAS) have identified many common single nucleotide polymorphisms (SNPs) associated with an increased risk of common psychiatric disorders, such as schizophrenia (SCZ), autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), and major depressive disorder (MDD). [4-7] While most of the identified risk alleles for these disorders have small effect sizes (typically an associated odds ratio (OR) <1.2), the aggregated effect of these alleles (along with other common alleles with effect sizes that fall short of genome-wide significant association) can be used to derive polygenic scores (PGS) that in turn can be used to predict case status with low to moderate accuracy. [8-10]

In addition to common SNPs, some rarer variants have also been identified that are associated with risk of psychiatric disorders. These variants include both single nucleotide variants that disrupt protein-coding sequences [11] [12] [13] [14] and structural variants that either disrupt exons of single genes [15] (such as NRXN1) or change the dosage of entire genes through deletion or duplication of large genomic regions (such as rare recurrent copy number variants; rCNVs).[16,17] Such rare protein-disrupting and/or gene-dosage-altering variants that have been found to associate with risk of psychiatric disorders, usually have larger effect sizes (OR>1.5) [16,18] than risk alleles of common SNPs and typically have a pleiotropic risk profile, i.e. are also associated with an increased risk of intellectual disability, epilepsy, and some somatic disorders.[18-19]

While early case-control studies indicated a very high, or even complete, penetrance of psychiatric or other neurodevelopmental disorders for some rCNVs (most prominently deletions at the velo-cardio facial syndrome (VCFS) region on 22q11)[20], these studies often relied on case samples enriched with long-stay patients and/or patients referred to genetic analysis, and controls screened for any history of mental illness. In recent years, studies of large community- and population-based samples have challenged this perception and derived estimates indicating a more moderately increased risk associated with 22q11 deletions and other rCNVs, especially for schizophrenia.[18,21-24] This emerging evidence of reduced penetrance of rCNVs underscores the critical role of additional etiological factors, such as polygenic background, and environmental influences, in modulating disease susceptibility.[16, 24-25] Understanding the complex relationship between these etiological factors is important for improved diagnostic and treatment initiatives.[26]

Some earlier studies have shown the penetrance of high-risk variants to be modified by the polygenic background in both human [27-28] and animal disease models [29-30] , and a few recent studies have also explored the combined effect of rCNV and PGSs on the risk of psychiatric disorders.[31-34] For example, among 481 children affected with ADHD, Martin et al.[32] found that carriers of any rare recurrent or non-recurrent CNV (N=60) had a significantly lower PGS for ADHD than non-carriers. Comparing the affected with 4670 control individuals, they found a corresponding significant interaction effect between carriage of CNVs and PGS with respect to risk of ADHD in a multiplicative logistic

regression model.[32] In another study, Bergen et al. evaluated the combined effects of CNVs and a PGS for schizophrenia on disease risk in a large SCZ case-control study.[33] They observed that among affected individuals, carriers of rCNVs had significantly lower PGS for SCZ than non-carriers and that those carrying rCNVs associated with higher risk of SCZ tended to have lower PGS than those carrying rCNVs associated with less increased risk of SCZ (although, surprisingly, the same was true for controls). When fitting a model to predict case status among all rCNV carriers, they also found evidence of a significant interaction between the odds ratio for SCZ of a carried rCNV and the PGS for SCZ, in such a way that the higher the odds ratio of the carried rCNV, the less added effect of PGS was found on the prediction of case status.[33] The results of both these studies, as well as two other studies comparing SCZ-PGS in affected carriers with unaffected controls[31], and PGS for SCZ and intelligence quotient, and the prevalence of SCZ and intellectual disability among 22q11.2 deletion carriers[35], are mostly consistent with a liability threshold model of disease, where PGS and CNVs contribute additively to the risk of ADHD and SCZ. In contrast to these four studies, one study found no significant difference in SCZ-PGS between affected carriers and non-carriers. [34]

Despite these efforts, the question of how common variants modify the effects of rCNVs on the risk of psychiatric disorders remains unclear. In fact, the above-mentioned studies are conducted on case-control samples that often are biased toward cases with long-term or severe illness, and healthy controls screened for family history of mental illness. Thus, their results are not population-representative and unlikely to be generalizable to the overall population.

In this study, we benefit from the unique design of the Danish iPSYCH2015 case-cohort sample[36] to examine the combined effect of rCNV and PGSs on the risk of psychiatric disorders in a population-representative setting. The iPSYCH2015 sample includes 140,016 individuals, consisting of a random subcohort (n=50,615) nested within a source population of all individuals who were born in Denmark between 1981-2008, and 92,531 cases encompassing all individuals from the same source population, who were diagnosed with ASD, ADHD, schizophrenia spectrum disorder (SSD), MDD or bipolar disorder (BPD) by the end of 2015. We have previously identified carriers of rCNVs at 27 genomic loci and estimated their population-based prevalence and associated risk of psychiatric disorders,[18] and in the current study we sought to:

1. Estimate absolute risk of ASD, ADHD, SSD, and MDD associated with rCNVs and PGS until the age of 31 years, both separately and combined using survival analysis;
2. Investigate potential interactions between rCNVs and PGSs with respect to the associated risk of the four psychiatric disorders using generalized linear models (GLMs), and
3. Compare PGS of psychiatric disorders between rCNV carriers and non-carriers within cases and controls separately for the above-mentioned diagnoses.

Results:

This study analyzed data from 96,599 unrelated individuals of European ancestry (Method) within the iPSYCH2015, comprising 61,890 cases and 34,709 randomly drawn samples (i.e., sub-cohort), with an overlap of 2,292 individuals (cases that were a part of the cohort). Of the total sample, 51,507 (53%) were male and 45,092 (47%) were female. The average age at the end of follow-up was 22.01 years (mean [SD], 22.01 [7.0]), with ages ranging from 7.0 to 34.6 years. After quality control, the number of remaining verified rCNV calls totaled 2,956 at 25 rCNV loci (sTable 1). We excluded BPD from our analyses due to the small sample size.

As individual rCNVs are rare in the population (prevalence <0.5%), we needed to group rCNVs to gain sufficient power for conducting analyses of the absolute risk of ASD, ADHD, SSD, and MDD associated with rCNVs. Meanwhile, we wanted to avoid using the simple and biased approach of "any rCNV vs. none" to categorize rCNV carriers, as it has been previously demonstrated that rCNV penetrance varies across different rCNVs and outcomes.[16-18] Hence, we considered dividing the rCNVs into various groups based on multiple approaches, including their associated odds ratios (ORs) with each diagnosis derived from iPSYCH2015[18] and previous literature,[17,22,37- 40] their locus-wide loss-of-function observed/expected upper-bound fraction (LOEUF) score,[41] and their pathogenicity score in ClinGen,[42] as well as composite risk scores based on a combination of these factors (see supplementary text for details). Overall, comparisons of these grouping strategies revealed consistency in distinguishing penetrance of the different rCNV groups. Since the LOEUF score is the most objective of these measures we chose it as our primary source of grouping, and divided rCNVs into three groups based on the summed LOEUF score of their contained genes (Method) as follows: rCNVs with LOEUF < 10, 10 < LOEUF < 25, and LOEUF > 25 (sFigure 1, sTable 2-6).

Absolute Risk of ASD, ADHD, SSD, and MDD Associated with rCNV Groups in iPSYCH2015 Case-Cohort

We derived the absolute risks of ADHD, ASD, SSD, and MDD associated with each rCNV group in carriers relative to non-carriers by integrating an inverse probability weighting scheme into the survival data (Method). Overall, rCNV carriers within all three LOEUF groups had a higher absolute risk than non-carriers of being diagnosed with ASD, ADHD, and SSD, but less so for MDD (Figure 1). For ASD and ADHD, the high LOEUF group (> 25) exhibited the highest absolute risk of diagnosis, with rates of 5.5% and 6.5%, respectively, followed by the medium and low LOEUF groups, as expected from our previous study[18]. However, in the case of SSD, both the medium and high LOEUF groups displayed highly similar increases in absolute risk, at 3.8% and 4.2%, respectively. In contrast, the low LOEUF group had an increased absolute risk of SSD diagnosis by only 0.5% compared to non-carriers (Figure 1). For MDD, the observed increase in the absolute risk only amounted to 0.7-1.1% increase for any of the rCNV carrier groups compared to noncarriers (Figure 1). Having observed the rise of absolute risks with the increase in LOEUF score of the rCNV groups, we first fitted nested GLM

models with and without rCNV carrier status as the independent linear covariate and then assessed the differences using likelihood ratio tests (methods). In concordance with the observations from Figure 1, the results revealed a significant increase in risk with the increase in LOEUF score levels of the rCNV groups for ASD, ADHD, and SSD ($\beta>0.24$, $P<6.67\times10^{-10}$), but not MDD ($\beta=0.04$, $P= 0.21$) (sTable 7).

Absolute Risk of ASD, ADHD, SSD, and MDD associated with PGS in iPSYCH2015 Case-Cohort

We first computed the polygenic scores (PGS) of the four disorders for the entire sample using external GWASs[4,6,43,44] through the SBayesR pipeline[45]. Subsequently, samples were stratified into three groups based on their PGS: low (PGS < 20th percentile), medium (20th percentile < PGS < 80th percentile), and high (PGS > 80th percentile). Absolute risk estimates for the four diagnoses were then derived in a similar way as for the rCNVs groups using the PGS group as the predictor in the survival model (methods). As expected, increased PGS was associated with an increased risk of being diagnosed with ASD, ADHD, SSD, and MDD (Figure 2). Among all groups, individuals with high PGS (> 80th percentile) exhibited the most substantial increase in absolute risk, with notable increases to 7.9% and 4.3% for MDD and ADHD, respectively (Figure 2). To formally assess the increase in the absolute risk of diagnoses associated with PGS groups, we employed GLM models similar to those in the previous step, although using PGS as the linear covariate in each model (methods). The overall results revealed that PGS was a significant positive predictor of the four diagnoses ($\beta>0.13$ and $P>3.88\times10^{-45}$ for each) (sTable 7).

Absolute Risk of ASD, ADHD, SSD, and MDD Associated with rCNVs and PGS Combined in iPSYCH2015 Case-Cohort

Finally, we derived the absolute risks of the four diagnoses associated jointly with both rCNVs and PGS groups using survival data, as previously described (see Methods). Overall, the absolute risk curves demonstrated that higher levels of PGS and LOEUF scores were associated with an elevated risk of diagnosis for ASD, ADHD, and SSD relative to non-carriers, with some notable exceptions within specific strata (Figure 3). In general, the absolute risk curves tended to follow the pattern observed separately for rCNV and PGS groups in additive manner. For instance, no clear distinctions in SSD diagnosis were observed between the middle and high LOEUF groups across the three PGS levels (Figure 3). Furthermore, for MDD, as previously noted, increases in PGS resulted in higher absolute risk but did not differ among carriers of the three LOEUF groups and non-carriers (Figure 3).

Model fitting results for ASD, ADHD, SSD, and MDD and investigating rCNV×PGS interactions

Following the findings from the absolute risk plots derived from the combination of rCNV and PGS groups, we investigated whether rCNVs and PGSs are associated with each

diagnosis in both an additive and interactive manner using GLM models. First, we constructed the base models for predicting each diagnosis using age, sex, and origin of the sample (i.e., iPSYCH2012 and iPSYCH2015i) as the covariates and then stepwise added rCNV groups and PGS separately, and then combined (i.e., additive model) as the predictive variables. The ultimate model included the rCNV×PGS interaction as the covariate. Likelihood ratio tests were performed to evaluate the difference between each model and its corresponding nested model (Methods, sTable 8a).

The results of fitted models using only rCNVs (i.e., in the form of aggregated, rCNV groups, and rCNVs individually) were significant for ASD, ADHD, and SSD (χ^2>38.09, P_{LRT}>1.14×10^{-5}), but not for MDD (χ^2<23.62, P_{LRT}>0.22). When using the PGS of each disorder to predict the corresponding disorder in the fitted models, the PGSs were positively associated with the four outcomes (χ^2>199.75, P_{LRT}<2.36×10^{-45}). Since there were no significant associations of rCNVs with MDD, it was excluded from further additive and interactive analyses (sTable 8a).

Furthermore, the fitted additive models (i.e., rCNV + PGS) were statistically significant when including rCNVs in the aggregated manner, groups, for ASD, ADHD, and SSD (χ^2<36.30, P_{LRT}<6.71×10^{-6}) (sTable 8a). However, the interaction analyses of rCNV×PGS (i.e., rCNVs in the form of groups and aggregated manner) across ASD, ADHD, and SSD did not show any significant results (χ^2<3.55, P_{LRT}>0.29) (sTable 8a & 10). As a sensitivity analysis, we repeated similar additive and interactive analyses replacing SSD with schizophrenia (SCZ) as the outcome including both rCV groups and aggregated rCNVs, in which the results consistently were significant for the additive models (χ^2>25.41, P_{LRT}<1.27 x 10^{-5}) but not the interactive models (χ^2<3.80, P_{LRT}>0.14) (sTable 9).

In addition to the main analyses, we repeated all the analyses assessing the 9 most frequent rCNVs (overall carrier frequency >1⁄1000) in the iPSCYH2015 dataset for ADHD, ASD, and SSD (see sTables 8b & 10-11 and method). Similarly, our findings revealed no significant rCNV×PGS interactions when assessing individual rCNVs across the three diagnoses (χ^2<3.39, P_{LRT}>0.06), except for the 16p13.11 duplication, which showed a significant negative interaction with PGS in predicting ADHD (χ^2=7.66, P_{LRT}>0.0056) (sTables 10).

We also calculated Nagelkerke's pseudo-R^2 for all the fitted models to assess the explained variance. Results confirmed that the additive inclusion of rCNVs and PGS improved the explained variance for ASD, ADHD, and SSD (Figure 4, sTable 8a). However, the interactive models did not yield significant improvements in explained variance except for 16p1311 duplication in ADHD.

Comparison of the psychiatric PGS profiles Between CNV Carriers and Non-Carriers Within Cases and Controls for ASD, ADHD, SSD, and MDD

Based on the liability threshold model, we hypothesized that carriers of rCNVs would require less additional additive risk from common variants (i.e., lower PGS) to be diagnosed with psychiatric disorders compared to non-carriers. Therefore, we compared the average PGS for each of the corresponding disorders between CNV carriers and non-carriers across rCNV groups in ADHD, ASD, SSD, and MDD cases and controls, separately (Figure 5, sTable 12) (see Methods).

Although we observed a trend of reduced mean PGS for ASD and ADHD among cases who were rCNV carriers in the three LOEUF groups, none differed significantly from non-carriers ($P_{welch's\text{-}FDR}$>0.35) (Figure 5). Moreover, there was no clear trend for diminished mean PGS among SSD and MDD cases among the rCNV groups, and all tests were insignificant ($P_{welch's\text{-}FDR}$ >0.58). Within the controls, results were not significantly different between carriers of rCNV groups and non-carriers for each diagnosis. Similarly, the comparison of aggregated rCNV carriers with non-carriers yielded insignificant results both among cases and controls ($P_{welch's\ FDR}$>0.11). Among the three LOEUF groups, a nominally significant decrease in PGS for ADHD was observed for unaffected carriers in the highest LOEUF group compared with the middle LOEUF group, which did not remain significant after FDR correction (P_{FDR}=0.35), (Figure 5, sTable 13).

Discussion:

Using 96,599 samples from the population-based iPSYCH2015 case-cohort,[36] we investigated the absolute risk of four major psychiatric disorders—ASD, ADHD, SSD, and MDD—associated with previously identified rCNVs and common SNPs, both independently and in combination. The large sample size, comprising cases diagnosed with ASD (n=16,964), ADHD (n=20,002), SSD (n=9,749), and MDD (n=25,693), along with a random subcohort (n=37,001), provided a robust framework for evaluating the interplay between rCNVs and common SNPs concerning the risk of these psychiatric disorders. To our knowledge, this is the first study to examine the interactive effects of common and rare variants (i.e., rCNVs) on ASD, and it represents the most extensive analysis of this kind for ADHD and MDD. Our findings show that the absolute risk of ASD, ADHD, and SSD was elevated in rCNV carriers compared to non-carriers, with variations observed among different subsets of rCNVs grouped by the summed LOEUF score of the genes they affect. Similarly, an increase in PGS was associated with a higher absolute risk for all four diagnoses, and the additive effects of these two genetic predictors were confirmed for ASD, ADHD, and SSD (no rCNV-associated effect was found for the risk of MDD). Analyses of interaction between rCNVs and PGS did not reveal any significant interactive effects for ASD, ADHD and SSD, except for 16p13.11 duplication which showed a negative significant interaction with ADHD-PGS.

First, we utilized time-to-event data related to each disorder to determine the absolute risk associated with rCNVs until the age of 31 years, subdividing the rCNVs into three groups based on the summed LOEUF scores[41] of the genes within each rCNV locus. Consistent with previous studies,[17,37,39] we confirmed that the absolute risk for ASD,

ADHD, and SSD was higher among rCNV carriers compared to non-carriers, with higher LOEUF scores of rCNVs generally indicating a greater risk than lower ones.[18] In contrast, rCNV carriers did not exhibit an increased absolute risk of MDD, neither overall nor across groups of increasing LOEUF scores. This aligns with our previous research, which found no overall evidence for an increased rCNV-associated risk of MDD, even though this study included 9 additional rare rCNV loci.[18] The negative MDD results differ from those of previous studies on internalizing disorders (i.e., depression and anxiety) in the UK Biobank (UKB)[22] and two US-based healthcare community samples, [21,46] which reported small to moderate effects of a few rCNVs the risk of anxiety and depression. This discrepancy may be attributed to the design differences between iPSYCH2015 (its young sample and population representativeness) compared to the other samples, which has been discussed in detail elsewhere.[18]

Second, using polygenic scores as predictors in the survival analysis, we demonstrated that an increase in PGS for each diagnosis was associated with a higher absolute risk of being diagnosed with the corresponding disorder. After stratifying individuals into low- to high-risk PGS groups, we replicated the finding that PGS, indexed by the aggregated effect of SNPs, is a significant predictor for each diagnosis, particularly for ADHD, SSD, and MDD.[33,47-48,49] However, the performance of PGS in differentiating the sample based on their PGS was suboptimal, highlighting the insufficient predictive power of current PGS and its potential utility in clinical psychiatry, which has been shown previously in other studies.[33,47-49] Among the four assessed diagnoses, the lack of predictability power was more pronounced for PGS of ASD, with the highest PGS group exhibiting only a slight increase in the absolute risk of ASD compared to the lowest PGS group. This is in line with earlier studies that have also observed a lower predictive accuracy for ASD-PGS compared to ADHD, SSD, and MDD.[6,43,44,50]

Although the impact of both rare and common genetic variants on psychiatric disorders is well established, their combined influence, particularly in population-based samples, has not been adequately explored. Our joint analysis of rCNVs and common variants revealed that rCNVs and PGS independently influence the risk of ASD, ADHD, and SSD in an additive manner. In other words, summing the effects of both rCNVs and PGS enhances the variation explained for these disorders (but not for MDD, as discussed in the previous section). However, we found no significant interaction effects when investigating the multiplicative interaction between rCNVs and PGS to predict each diagnosis, neither when considering overall carrier status (any rCNV vs no CNV) nor when considering groups of increasing dosage effect on constrained genes. Thus, we could not confirm the hypothesis that rCNV carriers in general would require a lower PGS according to the liability threshold model to be diagnosed with psychiatric disorders than non-carriers. The comparison of the average PGS between aggregated rCNV carriers (i.e., LOEUF groups and overall carriers) and non-carriers, stratified by cases and controls for each disorder, also did not support this hypothesis, suggesting that the two risk factors combined are not sufficiently impactful to saturate a liability threshold model of risk for any of the tested disorders, at least not as estimated in this study.

For SSD, our results contrast with some previous studies.[33,35] However, our findings from interaction analyses may not be directly comparable with the observations by studies such as Bergen et al., who reported significant interactions between odds ratios attributed to several rCNVs with SCZ-PGS among carriers of rCNVs.[33] Also, the inconsistency between our findings and previous observations may be due to differences in study design. Clinical case-control studies (such as Bergen et al.[33]) typically include case individuals with severe and/or long-term illness, often with multiple comorbidities, and healthy controls screened for any family history of mental illness, and are therefore likely enriched with individuals representing both ends of the distribution of the genetic load of disease-relevant SNPs and rCNVs. This can result in an overestimation of the risk magnitude associated with the tested variants. For instance, our previous studies within the iPSYCH2015 case-cohort sample [18,25,24] have found considerably lower risk estimates for several rCNVs (most prominently the 22q11.2 deletion) for SSD than those reported in clinical case-control studies.[37,38,40]

Also for ADHD, our results contrast with those of a previous study.[32] In this case, the failure to replicate previous observations could be attributed to differences in the study design, sample size, and/or tested rCNVs, as the previous study of Martin et al., included a much smaller case sample and only considered several large deletions.[32] Another reason could be that the ADHD-PGS in our study is based on less well-powered summary statistics of allelic risk estimates, as ADHD cases from iPSYCH2015 constitute almost half of all cases in the most recent GWAS by the Psychiatric Genomics Consortium (PGC),[6] and the leave-one-out summary statistics that we use are correspondingly less powered than when applying those of the full PGC-ADHD GWAS. Nonetheless, when examining the interaction of ADHD-PGS with the common individual rCNVs, we observed a significant negative interaction of the 16p13.11 duplication with PGS. This suggests an attenuation of the PGS effect on ADHD risk among carriers of this duplication, possibly due to the oversaturation of the risk among the carriers once they surpass the liability threshold. While the observed significant interaction in this rCNV may be due to distinct underlying biological processes that are currently unknown, it raises the question of why other tested rCNVs did not exhibit similar interactions. This discrepancy warrants further investigation in future studies.

For ASD, the interaction between rCNV and PGS has to our knowledge not been explored in previous studies, and we therefore have no prior studies to compare our results with. However, as with ADHD, iPSYCH2015 samples constitute a substantial part of all case samples in the largest available GWAS meta-analysis for ASD, and therefore the ASD-PGS may lack sensitivity in estimating the effect of common SNPs on ASD risk. Regarding MDD, it is unsurprising that we do not observe an interaction between the effect of common variants and rCNVS, as we do not see any effect of rCNVs on MDD risk to begin with. However, a recent study by Mollon et al.[51] explored the impact of rCNVs and PGS of anxiety and major depressive disorder (MDD) on psychopathology scores in the UKB and discovered only a few nominally significant interactions between some individual rCNVs and PGS on the mood and anxiety

scores.[51] Consequently, the lack of strong evidence for rCNV×PGS interaction in MDD, observed in both UKB and iPSYCH, suggests that such interaction would be unlikely to be detected if investigated in other samples as well.

This study has a few limitations. First, the rarity of many rCNVs limits the power to test their interactive effects with PGS on the risk of targeted diagnoses. We tried to overcome this by grouping rCNVs by the summed constraint of the genes within each rCNV locus (summed locus LOEUF score), as we had in our previous study[18] found evidence that increased locus LOEUF scores correlated with risk of disease, whereas the dosage change (i.e., deletion or duplication) did not. However, the LOEUF score may not optimally reflect the pathogenicity of individual rCNVs; thus, grouping rCNVs based on LOEUF score thresholds may have influenced our findings. Nonetheless, the only indication of a PGS-rCNV interaction we found, was when we looked for rCNV-specific effects among the 9 most common rCNVs, from which only duplication at 16p13.11 showed a significant interaction effect on ADHD. Second, diagnostic information was retrieved from hospital records, indicating that diagnoses made outside of the hospital setting were not accessible. Also, our use of SSD as the main measure of psychotic illness may differ from diagnostic criteria used to define schizophrenia in some previous studies; however, a sensitivity analysis of SCZ cases applying stricter criteria (ICD10:F20) found no indication of a difference from our overall results for SSD. Lastly, given the relatively young age of participants in the iPSYCH, our power to study late-onset disorders, such as bipolar disorder, was restricted. Future studies could mitigate these limitations by incorporating larger samples from broader age ranges, including diagnoses made outside of hospital settings.

Figure legends:

Figure 1: <u>Absolute risks of ASD, ADHD, SSD, and MDD associated with rCNV groups in iPSCYH2015</u>[36]. Absolute risks associated with rCNV groups were derived by fitting the *surv.fit* function using the survival package in R for each disorder separately. rCNVs were categorized into 3 groups based on their corresponding locus' LOEUF score. Absolute risk curves associated with rCNVs with LOEUF below 10, LOEUF greater than 10 and lower than 25, and greater than 25, are presented by blue, green, and red colors (full), respectively. Standard errors (SE) of each rCNV group are also shown by semi-transparent colors, accordingly. All the absolute risk curves are smoothed by the *smooth.spline* function from the Spline package in R.[52]

Figure 2: <u>Absolute risks of ASD, ADHD, SSD, and MDD associated with polygenic score (PGS) groups in iPSCYH2015</u>[36]. Absolute risks associated with PGS groups were derived by fitting the *surv.fit* function using the survival package[53] in R for each disorder separately. Individuals were categorized into 3 groups based on their PGS for each corresponding disorder (i.e., PGS<20%, 20%<PGS<80%, and PGS>80%), where their associated absolute risk curves are presented by light brown, green, and dark gray colors (full), respectively. Standard errors (SE) of each PGS group are also shown by semi-transparent colors, accordingly. All the absolute risk curves are smoothed by the *smooth.spline* function from the Spline package in R.[52]

Figure 3: <u>Absolute risks of ASD, ADHD, SSD, and MDD associated with rCNV and polygenic score (PGS) groups combined in ISPCYH2015</u>[36]. Joint absolute risks associated with rCNV and PGS group were derived by fitting the *surv.fit* function using the survival package in R for each disorder separately. For enhancing the visualization, absolute risk curves associated with each disorder are shown for non-carriers, and carriers within each of rCNV groups separately in 4 different panels, stratified by the PGS groups with light brown, green, and dark gray colors. Standard errors (SE) of each PGS group are also shown by transparent colors, accordingly. Due to low carrier counts across the strata at the follow-up age above 31, absolute risk curves are shown until the follow-up age of 31. All the curves were smoothed by the *smooth.spline* function from the Spline package in R.[52]

Figure 4: <u>Comparison of Variance explained by fitted GLM models using rCNV groups and PGS as the explanatory genetic exposures for predicting ASD, ADHD, SSD, and MDD in iPSYCH2015</u>. GLM models were fitted stepwise to investigate the effect of rCNV groups and PGS groups individually, additively, and interactively on each diagnosis, separately. All models included age, sex, and the genotyping array of the samples. Nagelkerke's R-squared was computed for each model utilizing the *PseudoR2* function from the DescTools package. Each bar is colored according to the genetic covariate used in the models across the diagnoses. Dark green, light green, dark gray, and red bars indicate R2 that are derived from fitted GLMs using rCNV status, and PGS, in additive form and interactively for predicting each diagnosis, respectively. Error bars indicate standard errors (SE) of Nagelkerke's R-squared obtained from each fitted model also were derived by bootstrapping the sample and fitting each model 1000 times (See method, sTable 5a). The results showed rCNV groups and PGS contribute significantly to the risk of ASD, ADHD, and SSD only in the additive manner and no significant interaction effects were found between rCNV groups and PGS. For MDD, there was no indication of significant rCNV effect.

Figure 5: <u>Comparison of average PGS among carriers and non-carriers across ASD, ADHD, SSD, and MDD by affection status.</u> Mean of PGS for each diagnosis are compared among non-carriers, carriers of rCNV groups as well as carriers of any rCNV

for cases and controls of ASD, ADHD, SSD, and MDD separately in different panels. Error bars indicate the SE of the PGS mean (sTable 12).

Method:

Sample and study design:

iPSYCH2015[36] is a Danish case-cohort study composed of iPSYCH2012 case-cohort and its extension; iPSYCH2015i case-cohort with 140,116 samples from the birth cohort of 1,657,449 singletons born between May 1, 1981, and December 31, 2008, who were residing in Denmark at the age of one and had a known registered mother.[36] The iPSYCH2015 case sample comprises 92,531 psychiatric patients with inpatient or outpatient hospital diagnoses, including major depressive disorder (MDD; ICD-10: F32-F33, ICD-8: 296.09, 296.29, 298.09, 300.49, n=37,555), affective disorder (AFF; ICD-10: F30-F39, n=40,482), autism spectrum disorder (ASD; ICD-10: F84, n=24,975), bipolar disorder (BPD; ICD-10: F30-F31, n=3,819), schizophrenia spectrum disorder (SSD; ICD-10: F20-F29, n=16,008), schizophrenia (SCZ; ICD-10: F20, n=8,113), and attention-deficit/hyperactivity disorder (ADHD; ICD-10: F90, n=29,668). The cohort component includes 50,615 individuals randomly selected from the same birth cohort. All diagnostic codes and dates of diagnoses were available until December 31, 2015, which were obtained through the Psychiatric Central Research Register (PCRR)[54].

Genotyping, phasing, and imputation:

All the individuals had their dried blood spots stored in the National Neonatal Screening Biobank (DSNB)[55], from which the DNA samples were retrieved. Genotyping was performed on the Infinium Psych Chip v1.0 (Illumina, San Diego CA, USA) and Illumina Global Screening Array v2.0 (Illumina, San Diego CA, USA) for the individuals in iPSYCH2012 and iPSYCH2015i, respectively. Furthermore, GenomeStudio software (Illumina) was used to extract B-allele frequency and probe intensities (through log-R Ratio; LRR). Only samples with a genotyping call rate of over 95% were retained. Further details regarding the quality control and genotyping call rate are described elsewhere. [36]

Samples genotyped on both arrays were first phased and imputed using BEAGLE5.4 separately, with HRCv1.1[56] as the reference. Afterwards, SNPs from two separately imputed datasets were selected if passed the following criteria: 1) Genotyped at least on one of the two arrays; 2) Minimum imputation quality above 0.99 within both datasets, and 3) Maximum of 1.5% different minor allele frequency (MAF) between the two datasets. The selected SNPs from both datasets were ultimately phased and imputed again using BEAGLE5.4 aligning with HRCv1.1[56] resulting in a total 6,518,119 SNPs with a minimum MAF>1% after performing quality control.

Identifying genetically homogenous samples, relatedness, and PCs:

We utilized a set of 20 principal components (PCs) and KING-related kinship coefficients previously obtained[57] to identify a list of individuals of European ancestry from iPSYCH2015i, pruned for relatedness beyond the third degree. For more details, see the Supplementary text.

CNV calling and quality control:

Briefly, the selection of rCNVs in this study largely overlapped with the 54 CNVs identified in the UKB study.[19] We excluded rCNV loci that were not listed as recurrent CNVs on ClinGen[43] or labeled as "No Evidence" for either haploinsufficiency or triplosensitivity, resulting in 27 rCNV loci (2 of which dropped out during subsequent QC subsetting). CNV calling was performed by PennCNV[59] and samples with an LRR-SD of 0.35 or greater, B-allele frequency drift of 0.005 or greater, and/or GC-wave factor of 0.02 or greater were removed from further analysis. PennCNV calls were validated using QCtreeCNV pipeline [60] providing a list of putative CNV calls, which were then visually inspected by three raters. Further details regarding the CNV calling process, quality control, and visual inspection are described thoroughly elsewhere.[18]

CNV grouping:

We evaluated whether categorizing rCNVs based on different approaches and estimating odds ratios associated with rCNV groups across the groups and approaches would result in stratification of rCNVs pathogenicity in a similar manner. The 6 grouping strategies were to group rCNVs based on; their internal odds ratios (ORs) for each disorder (i.e., rCNV-associated OR estimated previously within iPSYCH2015 data)[18]; external odds ratios (i.e., rCNV-associated odds found in the literature); simple approach of any rCNV vs no rCNV; whether rCNVs were identified to have some degrees of haploinsufficiency or triplosensitivity (HI/TS) on ClinGen[43]; LOEUF score of rCNVs; as well as composite score constructed from the three latter scores (see more details in Supplement).

LOEUF score calculation:

The LOEUF score of each rCNV locus was derived by summing the inverted LOEUF scores (1/LOEUF) of all the genes that overlapped at least 50% with the corresponding locus, obtained from the GnomADdatabase (v2.1.1)[42] as described in our previous study.[18] In this study, rCNV loci were categorized into three levels based on LOEUF score thresholds: <10, 10 to 25, and >25.

Polygenic score:

To compute polygenic scores, we used the latest external GWAS summary statistics of ASD[45], ADHD[6], SCZ[4], and MDD[44] obtained from the Psychiatric Genomics Consortium (PGC) after excluding iPSYCH individuals from the discovery sample). The number of SNPs used for PGS generation were 1,041,920 after intersecting SNPs in iPSYCH2015 datasets with SNPs in HapMap3 as the reference. The p-value threshold for SNPs in each GWAS was set at 0.9. SNP effects reported in each GWAS were rescaled using SBayesR[46], based on Bayesian multiple regression with the scaling factor (γ) and the number of mixture components (π) set to the default settings (--pi 0.95,0.02,0.02,0.02,0.02,0.01; --gamma 0,0.01,0.1,1). After rescaling the SNP's effect, PGSs were generated by applying the *--score* in *PLINKv2.00a2.3LM* to the *SBayesR's* output files. Scores were normalized by subtracting the mean score of the control group from each individual's score and dividing by the standard deviation of the control group's mean score.

Absolute Risk of ASD, ADHD, SSD, and MDD Associated with rCNV and PGS Groups in iPSYCH2015 Case-Cohort

The rCNV-associated absolute risk for each diagnosis was derived using the survfit function from the R survival package[61], with age at first diagnosis or censoring age- depending on the event proximity- as the outcome and rCNV carrier status as the explanatory variable in each model. Censorship was implemented such that if there was a death, emigration, no diagnosis of any iPSYCH index disorder by the end of the follow-up period (December 31st, 2015), or if the participant was lost to follow-up. rCNV status was defined as a categorical variable with "no CNV" as the reference in each model. Similar models were also used when investigating absolute risk associated with PGS groups as the explanatory variable, where the lowest PGS group (i.e., PGS<20%) was the reference for this categorical variable. The definition of variables was furthermore the same when using both rCNV and PGS groups as independent variables for estimating absolute risks. All models accounted for the inverse probability of sampling (IPS) weights, following the weighting method developed by Barlow et al[62]. Due to legal restrictions in sharing individual-level data, we smoothened absolute risk curves applying *smooth.spline* function from the Spline package[53] in R to the derived survival data in each fitted model.
To account for population stratification effects, all PGS calculations were linearly regressed against 20 principal components derived from unrelated individuals of European ancestry, and the residuals of fitted models were subsequently preserved for analysis.

Fitting generalised linear models (GLMs):

We constructed a series of logistic regression models by GLM to investigate the effect of rCNV and PGS on the risk of each outcome, separately. All models were adjusted for age, sex, and sample origin (i.e., iPSYCH2012 or iPSYCH2015i) of subjects where the model outcome was the status for the target diagnosis coded as a binary variable.

Sequentially, we introduced rCNV and PGS as predictors into each model, assessing them individually, in combination (i.e., rCNV + PGS) as well as in the interactive form of rCNV×PGS. In each step, a likelihood ratio test was employed to compare the fitted model with the nested model. For each diagnosis, 54 rCNVs (i.e., corresponding to del/dup at each locus) were analyzed individually, in the form of LOEUF-groups and aggregate (i.e., any rCNV) as a categorical variable where the carrier status was labeled with regards to the reference level as having no rCNV. Furthermore, to test the effect of levels of LOEUF groups on each outcome, we fitted similar GLMs, where rCNV groups (i.e., LOEUF groups) were included as a numeric covariate after converting the LOEUF groups' levels to 0,1,2 and 3, representing "having no rCNV", rCNVs with "LOEUF<10", "10<LOEUF<25" and "LOEUF>25", respectively. When assessing the PGS effect using PGS as a linear variable as well as the categorial variable with three strata (i.e., less than 20%, between 20% and 80%, and over 80%) for all the analyses, except for the interaction analysis with the individual rCNVs. To adjust for the population stratification, all PGS scores were regressed on 20 principal components from unrelated European individuals, and the residuals were retained for further analysis. We derived the Nagelkerke's R-squared for each model, by subtracting the R-squared of the fitted model from the R-squared of the null model utilizing the *PseudoR2* function from the DescTools package[63]. Subsequently, to compute the standard errors (SE) of Nagelkerke's R-squared obtained from each fitted model, we first bootstrapped the sample 1000 times followed by fitting the model each time and deriving their corresponding Nagelkerke's R-squared. We then estimated the SE of R-squared by deriving the standard deviation (SD) of the R-squareds vector corresponding to the bootstrapped models divided by the square root of 1000.

Comparison of PGS profile between rCNV carriers and non-carriers:

Welch's t-tests were used to test the differences in the mean of PGS scores between carriers and non-carriers within cases and controls of each outcome separately.

Ethical permissions:

Research permission and access to the iPSYCH data were granted by the Danish Scientific Ethics Committee, Danish Health Authority, and the Danish Neonatal Screening Biobank Committee (https://ipsych.dk/en/data-security/health-research-and-ethicalapproval/). Following Danish legislation, constructing iPSYH data was exempted from the need for informed consent from human research participants. This study was conducted on the data stored on a secure server at the Aarhus Genome Data Center (https://genome.au.dk/).

Acknowledgements:

The iPSYCH team was supported by grants from the Lundbeck Foundation (R102-A9118, R155-2014-1724, and R248-2017-2003), NIH/NIMH (1R01MH124851-01 to

A.D.B.), EU's Horizon Europe program under grant agreement no. 101057385 (R2D2-MH; to A.D.B.) and the Universities and University Hospitals of Aarhus and Copenhagen. The Danish National Biobank resource was supported by the Novo Nordisk Foundation. High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Center for Genomics and Personalized Medicine and the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to A.D.B.).

COIs:

A.D.B. has received speaker fee from Lundbeck.

References:

1. The Brainstorm Consortium, Anttila V, Bulik-Sullivan B, et al. Analysis of shared heritability in common disorders of the brain. *Science*. 2018;360(6395):eaap8757. doi:10.1126/science.aap8757

2. Visscher PM, Goddard ME, Derks EM, Wray NR. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Molecular Psychiatry*. 2012;17(5):474-485. doi:10.1038/mp.2011.65

3. Buxbaum JD, Baron-Cohen S, Devlin B. Genetics in psychiatry: common variant association studies. *Molecular Autism*. 2010;1(1):6. doi:10.1186/2040-2392-1-6

4. Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature*. 2022;604(7906):502-508. doi:10.1038/s41586-022-04434-5

5. Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444. doi:10.1038/s41588-019-0344-8

6. Demontis D, Walters GB, Athanasiadis G, et al. Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet*. 2023;55(2):198-208. doi:10.1038/s41588-022-01285-8

7. Meng X, Navoly G, Giannakopoulou O, et al. Multi-ancestry genome-wide association study of major depression aids locus discovery, fine mapping, gene prioritization and causal inference. *Nat Genet*. Published online January 4, 2024. doi:10.1038/s41588-023-01596-4

8. Patron J, Serra-Cayuela A, Han B, Li C, Wishart DS. Assessing the performance of genome-wide association studies for predicting disease risk. Devaney J, ed. *PLoS ONE*. 2019;14(12):e0220215. doi:10.1371/journal.pone.0220215

9. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):44. doi:10.1186/s13073-020-00742-5

10. Bogdan R, Baranger DAA, Agrawal A. Polygenic Risk Scores in Clinical Psychology: Bridging Genomic Risk to Individual Differences. *Annu Rev Clin Psychol*. 2018;14(1):119-157. doi:10.1146/annurev-clinpsy-050817-084847

11. Nagahama K, Sakoori K, Watanabe T, et al. Setd1a Insufficiency in Mice Attenuates Excitatory Synaptic Function and Recapitulates Schizophrenia-Related Behavioral Abnormalities. *Cell Reports*. 2020;32(11):108126. doi:10.1016/j.celrep.2020.108126

12. Semesta KM, Garces A, Tsvetanova NG. The psychosis risk factor RBM12 encodes a novel repressor of GPCR/cAMP signal transduction. *Journal of Biological Chemistry*. 2023;299(9):105133. doi:10.1016/j.jbc.2023.105133

13. Fu JM, Satterstrom FK, Peng M, et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat Genet*. 2022;54(9):1320-1331. doi:10.1038/s41588-022-01104-0

14. Singh T, Poterba T, Curtis D, et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature*. 2022;604(7906):509-516. doi:10.1038/s41586-022-04556-w

15. Ching MSL, Shen Y, Tan W, et al. Deletions of *NRXN1* (neurexin-1) predispose to a wide spectrum of developmental disorders. *American J of Med Genetics Pt B*. 2010;153B(4):937-947. doi:10.1002/ajmg.b.31063

16. Rees E, Kirov G. Copy number variation and neuropsychiatric illness. *Current Opinion in Genetics & Development*. 2021;68:57-63. doi:10.1016/j.gde.2021.02.014

17. Malhotra D, Sebat J. CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. *Cell*. 2012;148(6):1223-1241. doi:10.1016/j.cell.2012.02.039

18. Vaez M, Montalbano S, Calle Sánchez X, et al. Population-Based Risk of Psychiatric Disorders Associated With Recurrent Copy Number Variants. *JAMA Psychiatry*. Published online June 26, 2024. doi:10.1001/jamapsychiatry.2024.1453

19. Crawford K, Bracher-Smith M, Owen D, et al. Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *J Med Genet*. 2019;56(3):131-138. doi:10.1136/jmedgenet-2018-105477

20. Gothelf D, Frisch A, Michaelovsky E, Weizman A, Shprintzen RJ. Velo-Cardio-Facial Syndrome. *J Ment Health Res Intellect Disabil*. 2009;2(2):149-167. doi:10.1080/19315860902756136

21. Martin CL, Wain KE, Oetjens MT, et al. Identification of Neuropsychiatric Copy Number Variants in a Health Care System Population. *JAMA Psychiatry*. 2020;77(12):1276. doi:10.1001/jamapsychiatry.2020.2159

22. Kendall KM, Rees E, Bracher-Smith M, et al. Association of Rare Copy Number Variants With Risk of Depression. *JAMA Psychiatry*. 2019;76(8):818. doi:10.1001/jamapsychiatry.2019.0566

23. Olson H, Shen Y, Avallone J, et al. Copy number variation plays an important role in clinical epilepsy: CMA in Epilepsy. *Ann Neurol*. 2014;75(6):943-958. doi:10.1002/ana.24178

24. Calle Sánchez X, Helenius D, Bybjerg-Grauholm J, et al. Comparing Copy Number Variations in a Danish Case Cohort of Individuals With Psychiatric Disorders. *JAMA Psychiatry*. 2022;79(1):59. doi:10.1001/jamapsychiatry.2021.3392

25. Olsen L, Sparsø T, Weinsheimer SM, et al. Prevalence of rearrangements in the 22q11.2 region and population-based risk of neuropsychiatric and developmental disorders in a Danish population: a case-cohort study. *The Lancet Psychiatry*. 2018;5(7):573-580. doi:10.1016/S2215-0366(18)30168-8

26. Schork AJ, Schork MA, Schork NJ. Genetic risks and clinical rewards. *Nature Genetics*. 2018;50(9):1210-1211. doi:10.1038/s41588-018-0213-x

27. Fahed AC, Wang M, Homburger JR, et al. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nat Commun*. 2020;11(1):3635. doi:10.1038/s41467-020-17374-3

28. iPSYCH-Broad Autism Group, Psychiatric Genomics Consortium Autism Group, Weiner DJ, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet*. 2017;49(7):978-985. doi:10.1038/ng.3863

29. Neumann PE, Frankel WN, Letts VA, Coffin JM, Copp AJ, Bernfield M. Multifactorial inheritance of neural tube defects: localization of the major gene and recognition of modifiers in ct mutant mice. *Nat Genet*. 1994;6(4):357-362. doi:10.1038/ng0494-357

30. Letts VA, Schork NJ, Copp AJ, Bernfield M, Frankel WN. A Curly-Tail Modifier Locus,mct1,on Mouse Chromosome 17. *Genomics*. 1995;29(3):719-724. doi:10.1006/geno.1995.9946

31. Tansey KE, Rees E, Linden DE, et al. Common alleles contribute to schizophrenia in CNV carriers. *Mol Psychiatry*. 2016;21(8):1085-1089. doi:10.1038/mp.2015.143

32. Martin J, O'Donovan MC, Thapar A, Langley K, Williams N. The relative contribution of common and rare genetic variants to ADHD. *Transl Psychiatry*. 2015;5(2):e506-e506. doi:10.1038/tp.2015.5

33. Bergen SE, Ploner A, Howrigan D, et al. Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. *AJP*. 2019;176(1):29-35. doi:10.1176/appi.ajp.2018.17040467

34. Taniguchi S, Ninomiya K, Kushima I, et al. Polygenic risk scores in schizophrenia with clinically significant copy number variants. *Psychiatry Clin Neurosci*. 2020;74(1):35-39. doi:10.1111/pcn.12926

35. International 22q11.2 Brain and Behavior Consortium, Davies RW, Fiksinski AM, et al. Using common genetic variation to examine phenotypic expression and risk prediction in 22q11.2 deletion syndrome. *Nat Med*. 2020;26(12):1912-1918. doi:10.1038/s41591-020-1103-1

36. Bybjerg-Grauholm J, Pedersen CB, Pedersen G, et al. Title: The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. :20.

37. Marshall CR, Howrigan DP, Merico D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017;49(1):27-35. doi:10.1038/ng.3725

38. Rees E, Kendall K, Pardiñas AF, et al. Analysis of Intellectual Disability Copy Number Variants for Association With Schizophrenia. *JAMA Psychiatry*. 2016;73(9):963. doi:10.1001/jamapsychiatry.2016.1831

39. Gudmundsson OO, Walters GB, Ingason A, et al. Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Transl Psychiatry*. 2019;9(1):258. doi:10.1038/s41398-019-0599-y

40. Rees E, Walters JTR, Georgieva L, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry*. 2014;204(2):108-114. doi:10.1192/bjp.bp.113.131052

41. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

42. Rehm HL, Berg JS, Brooks LD, et al. ClinGen — The Clinical Genome Resource. *N Engl J Med*. 2015;372(23):2235-2242. doi:10.1056/NEJMsr1406261

43. Howard DM, Adams MJ, Clarke TK, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7

44. Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics

Consortium, et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*. 2019;51(3):431-444. doi:10.1038/s41588-019-0344-8

45. Lloyd-Jones LR, Zeng J, Sidorenko J, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun*. 2019;10(1):5086. doi:10.1038/s41467-019-12653-0

46. Birnbaum R, Mahjani B, Loos RJF, Sharp AJ. Clinical Characterization of Copy Number Variants Associated With Neurodevelopmental Disorders in a Large-scale Multiancestry Biobank. *JAMA Psychiatry*. 2022;79(3):250. doi:10.1001/jamapsychiatry.2021.4080

47. Østergaard SD, Trabjerg BB, Als TD, et al. Polygenic risk score, psychosocial environment and the risk of attention-deficit/hyperactivity disorder. *Transl Psychiatry*. 2020;10(1):335. doi:10.1038/s41398-020-01019-6

48. Ni G, Zeng J, Revez JA, et al. A Comparison of Ten Polygenic Score Methods for Psychiatric Disorders Applied Across Multiple Cohorts. *Biological Psychiatry*. 2021;90(9):611-620. doi:10.1016/j.biopsych.2021.04.018

49. De Wit MM, Morgan MJ, Libedinsky I, et al. A Systematic Review and Meta-Analysis: Research Using the Autism Polygenic Score. Published online March 9, 2024. doi:10.1101/2024.03.08.24303918

50. Antaki D, Guevara J, Maihofer AX, et al. A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nat Genet*. 2022;54(9):1284-1292. doi:10.1038/s41588-022-01064-5

51. Mollon J, Schultz LM, Huguet G, et al. Impact of Copy Number Variants and Polygenic Risk Scores on Psychopathology in the UK Biobank. *Biological Psychiatry*. 2023;94(7):591-600. doi:10.1016/j.biopsych.2023.01.028

52. Wang W, Yan J. Shape-Restricted Regression Splines with R Package splines2. *Journal of Data Science*. Published online 2021:498-517. doi:10.6339/21-JDS1020

53. Therneau T. Published online 2024. https://cran.r-project.org/package=survival

54. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. *Scand J Public Health*. 2011;39(7_suppl):54-57. doi:10.1177/1403494810395825

55. Nørgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish Newborn Screening Biobank. *J of Inher Metab Disea*. 2007;30(4):530-536. doi:10.1007/s10545-007-0631-x

56. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

57. Appadurai V, Bybjerg-Grauholm J, Krebs MD, et al. Accuracy of haplotype estimation and whole genome imputation affects complex trait analyses in complex biobanks. *Commun Biol*. 2023;6(1):101. doi:10.1038/s42003-023-04477-y

Figures:

Figure 1:

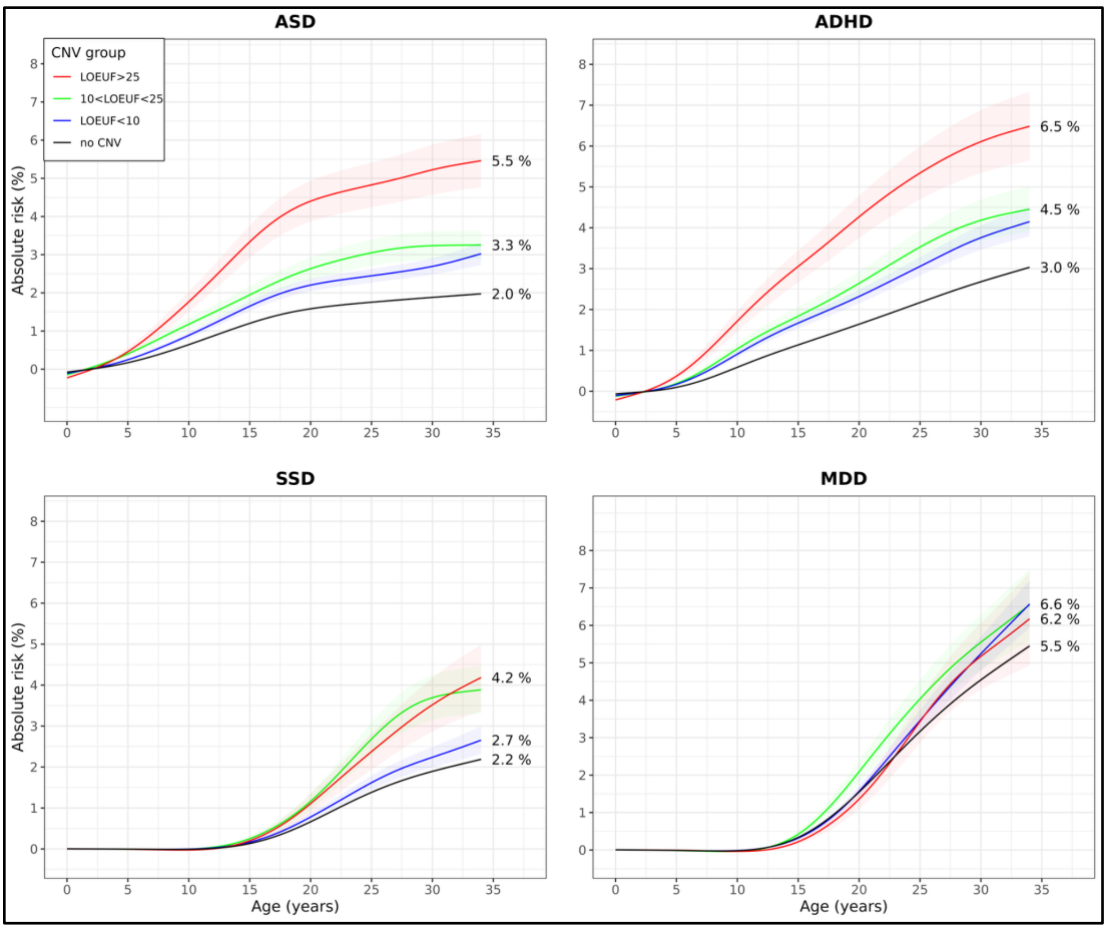


Figure 2:

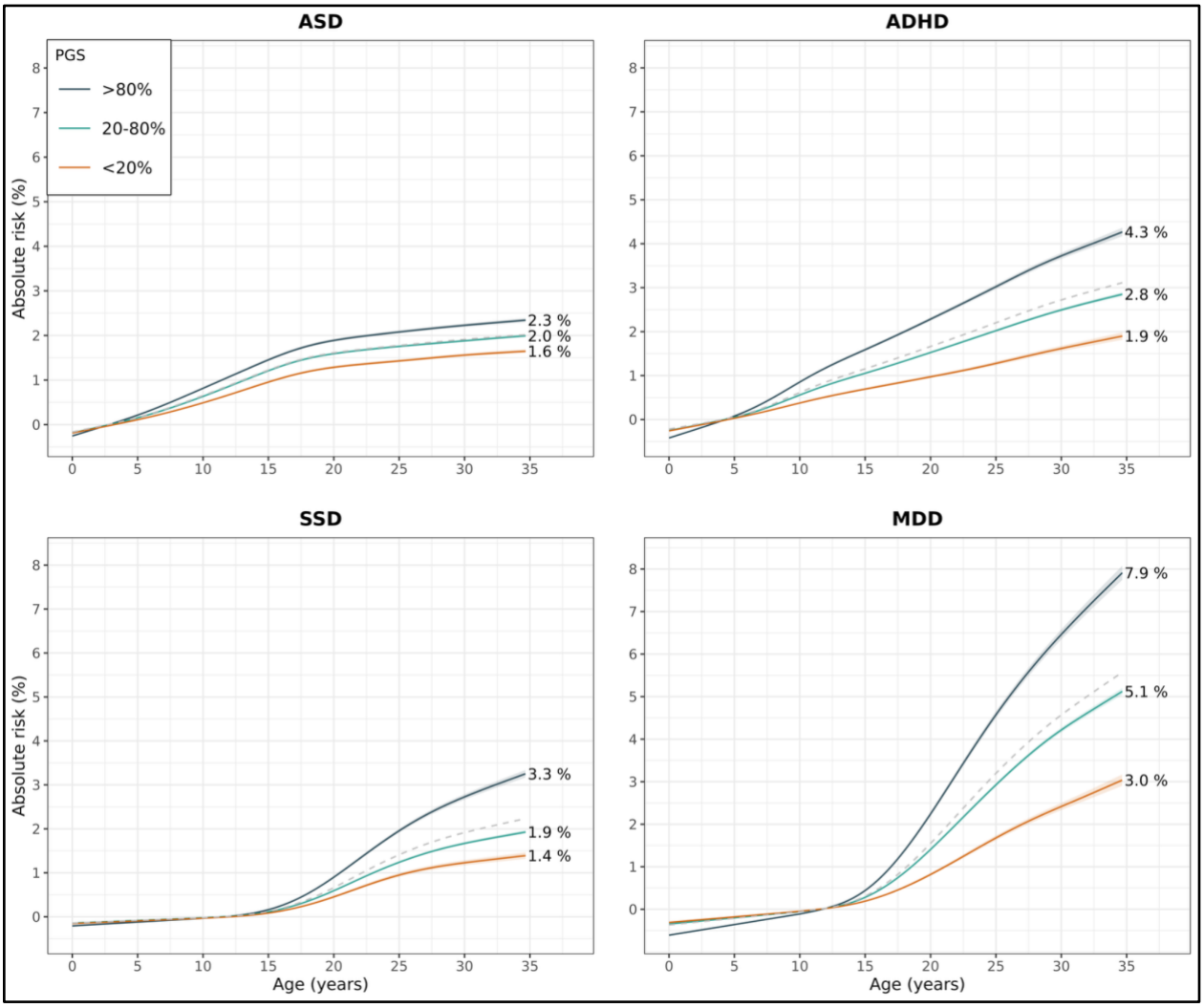

ASD
ADHD
SSD
MDD
PGS
>80%
20-80%
<20%
Absolute risk (%)
Age (years)
2.3 %
2.0 %
1.6 %
4.3 %
2.8 %
1.9 %
3.3 %
1.9 %
1.4 %
7.9 %
5.1 %
3.0 %

Figure 3:

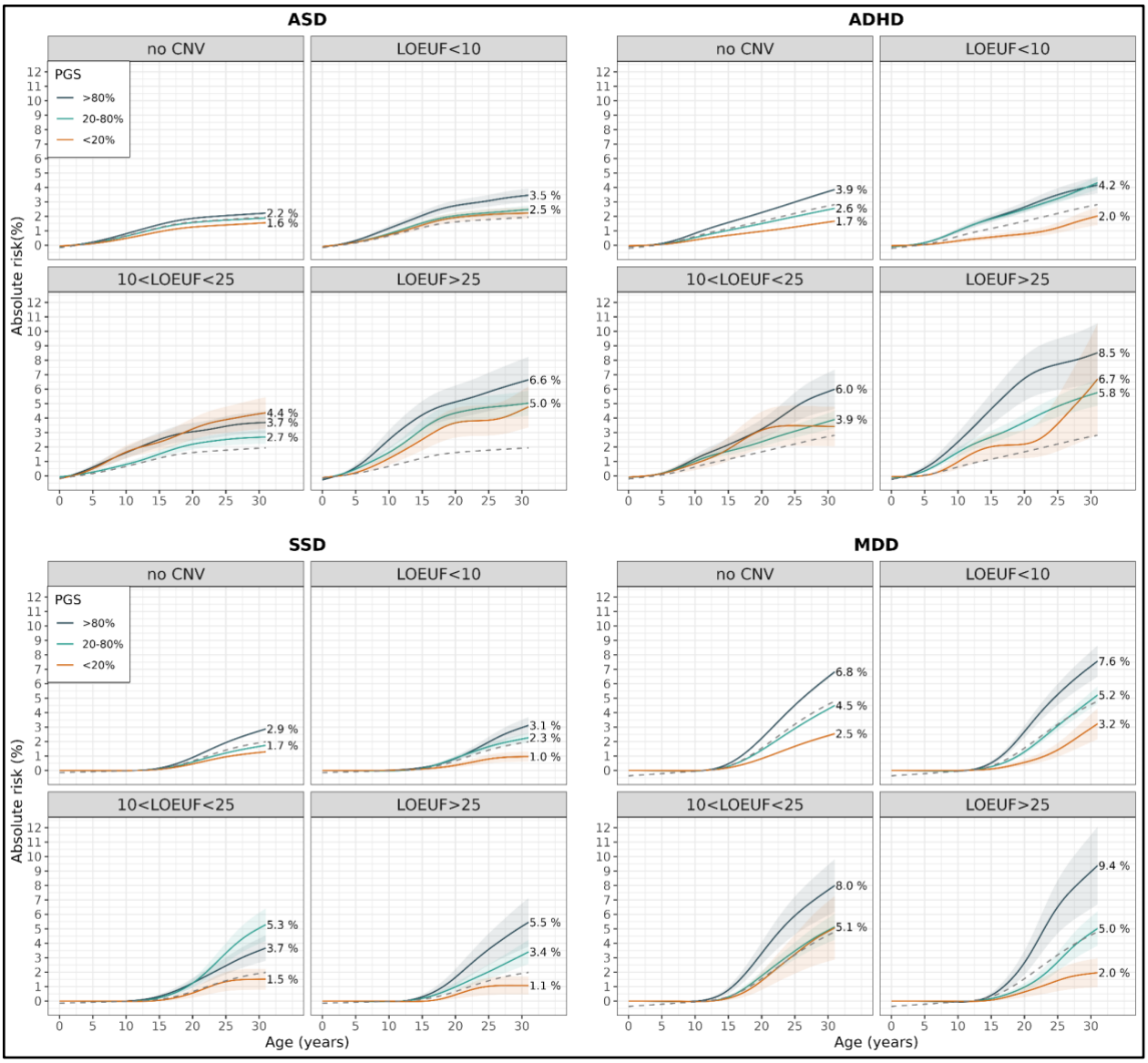

ASD
ADHD
SSD
MDD
no CNV
LOEUF<10
10<LOEUF<25
LOEUF>25
PGS
>80%
20-80%
<20%
Absolute risk(%)
Absolute risk (%)
Age (years)
2.2 %
1.6 %
3.5 %
2.5 %
4.4 %
3.7 %
2.7 %
6.6 %
5.0 %
3.9 %
2.6 %
1.7 %
4.2 %
2.0 %
6.0 %
3.9 %
8.5 %
6.7 %
5.8 %
2.9 %
1.7 %
3.1 %
2.3 %
1.0 %
5.3 %
3.7 %
1.5 %
5.5 %
3.4 %
1.1 %
6.8 %
4.5 %
2.5 %
7.6 %
5.2 %
3.2 %
8.0 %
5.1 %
9.4 %
5.0 %
2.0 %

Figure 4:

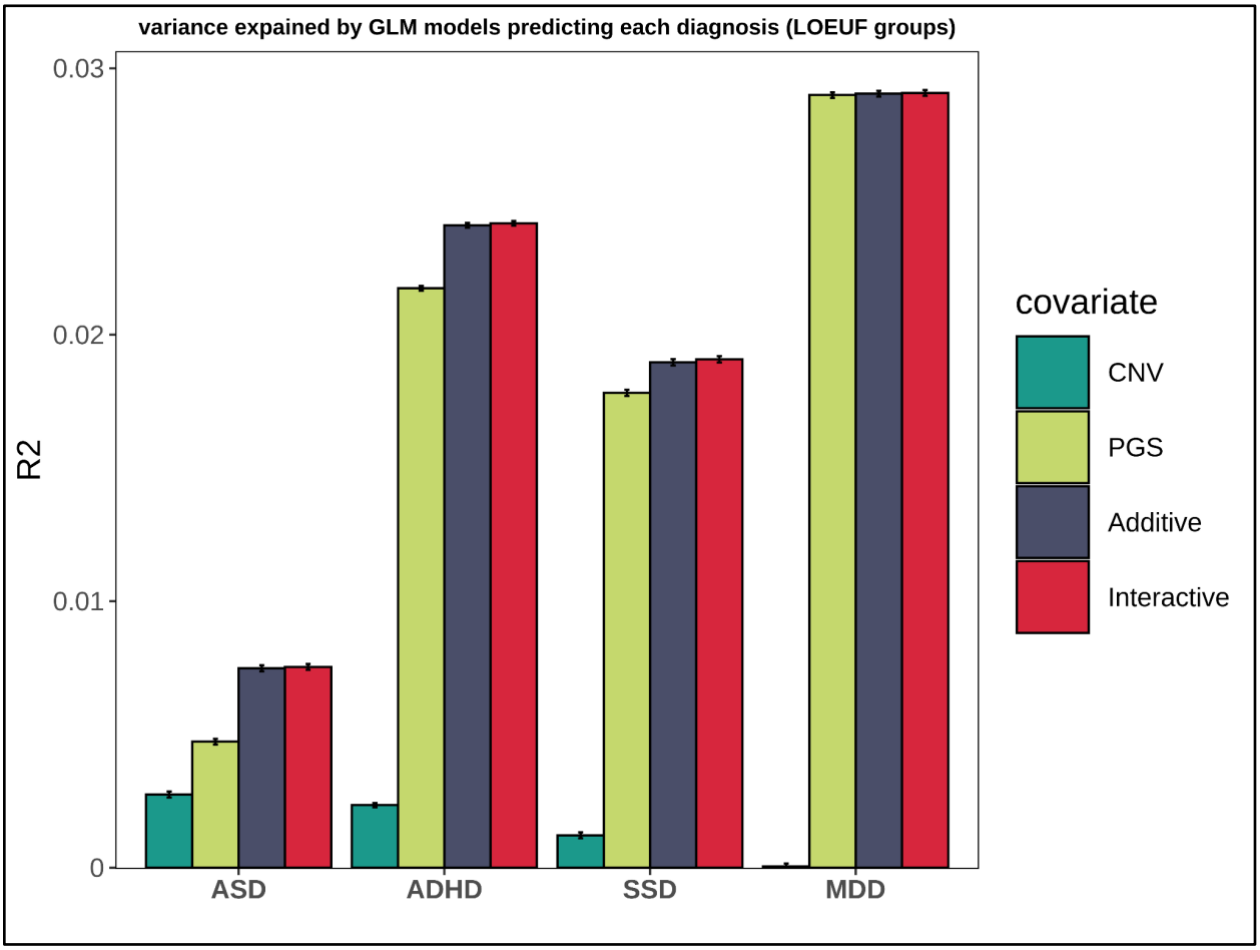
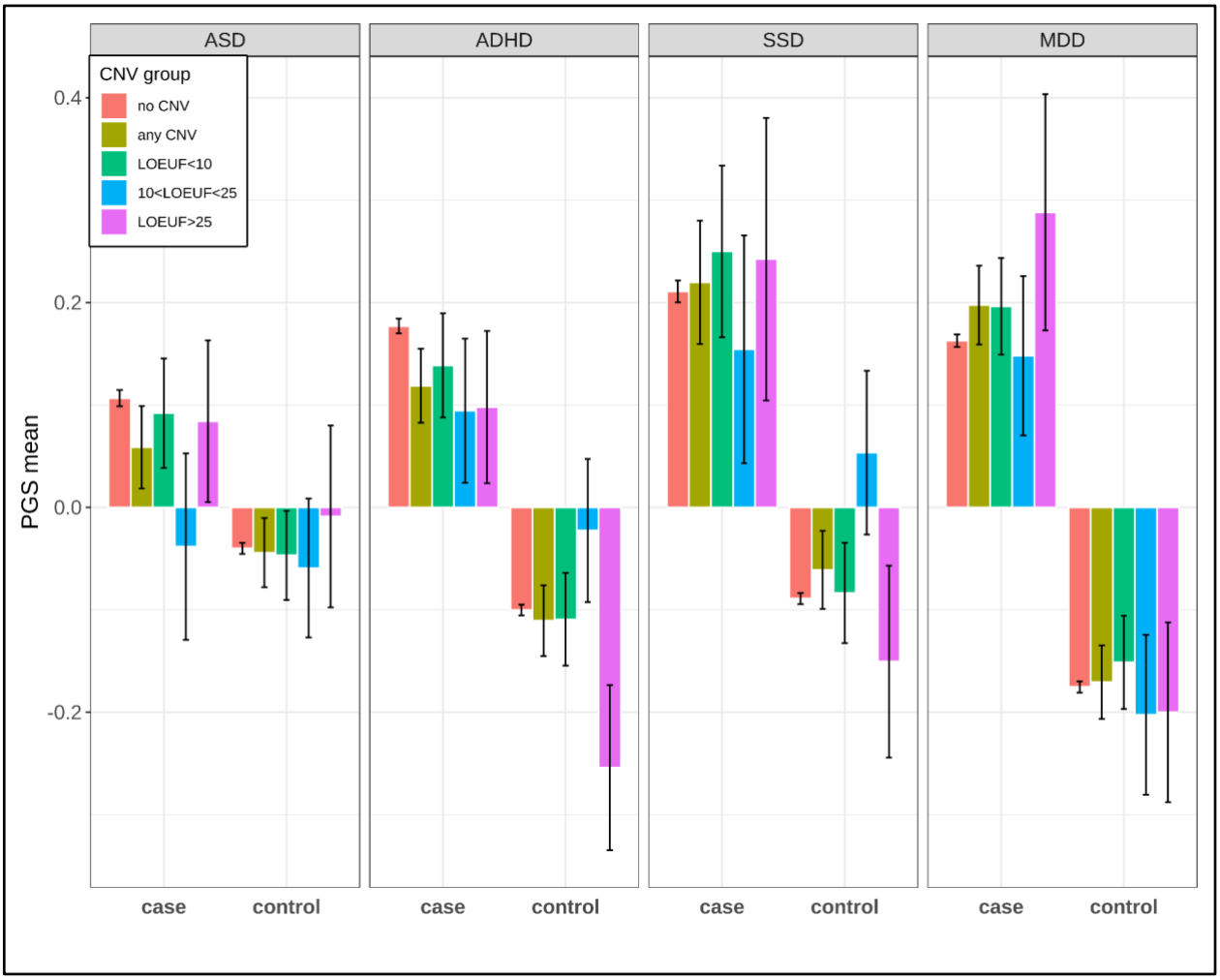

variance expained by GLM models predicting each diagnosis (LOEUF groups)
R2
0.03
0.02
0.01
0
ASD
ADHD
SSD
MDD
covariate
CNV
PGS
Additive
Interactive

Figure 5:


ASD
ADHD
SSD
MDD
CNV group
no CNV
any CNV
LOEUF<10
10<LOEUF<25
LOEUF>25
PGS mean
0.4
0.2
0.0
-0.2
case
control
case
control
case
control
case
control