

University of Groningen

Data-driven subphenotypic dissection of the clinical heterogeneity of schizophrenia spectrum disorders

Habtewold, Tesfa

DOI:
[10.33612/diss.156108872](https://doi.org/10.33612/diss.156108872)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Habtewold, T. (2021). *Data-driven subphenotypic dissection of the clinical heterogeneity of schizophrenia spectrum disorders*. University of Groningen. <https://doi.org/10.33612/diss.156108872>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

CHAPTER 5

Association of schizophrenia polygenic risk score with data-driven cognitive subtypes: a six-year longitudinal study in patients, siblings and controls

Tesfa Dejenie Habtewold, Edith J. Liemburg, Md Atiqul Islam, Sonja M. C. de Zwarte, H. Marike Boezen, GROUP Investigators (Jurjen J. Luykx, Bart P.F. Rutten, Ruud van Winkel, Therese van Amelsvoort, Agna A. Bartels-Velthuis, Wiepke Cahn, Lieuwe de Haan, Rene S. Kahn, Frederike Schirmbeck, Claudia J.P. Simons, Jim van Os), Richard Bruggeman[‡], Behrooz Z. Alizadeh[‡]

[‡]= Contributed equally

Schizophrenia Research 2020; 223: 135-147

ABSTRACT

Cross-sectional studies have shown that the polygenic risk score for schizophrenia (PRS_{SCZ}) may influence heterogeneity in cognitive performance although evidence from family-based longitudinal study is limited. This study aimed to identify trajectories of cognitive function and assess whether the PRS_{SCZ} is associated with baseline cognitive performance and predicted six-year trajectories. We included 1 119 patients with a schizophrenia spectrum disorder, and 1 059 unaffected siblings and 586 unrelated controls who are eligible at baseline. Genotype data were collected at baseline, whereas clinical and sociodemographic data were collected at baseline, three and six years. Group-based trajectory modeling was applied on a weighted standardized composite score of general cognition to unravel cognitive subtypes and explore trajectories over time. We followed a standard procedure to calculate the polygenic risk score. A random-effects ordinal regression model was used to investigate the association between PRS_{SCZ} and cognitive subtypes. Five cognitive subtypes with variable trajectories were found in patients, four in siblings and controls, and six in all combined samples. PRS_{SCZ} significantly predicted poor cognitive trajectories in patients, siblings and all samples. After Bonferroni correction and adjustment for non-genetic factors, only the results in all combined sample remained significant. Cognitive impairment in schizophrenia is heterogeneous and may be linked with high PRS_{SCZ} . Our finding confirmed at least in all combined samples the presence of genetic overlap between schizophrenia and cognitive function and can give insight into the mechanisms of cognitive deficits.

Keywords: Psychosis; Schizophrenia; Cognition; Cognitive trajectory; Polygenic risk score; Heterogeneity

INTRODUCTION

Cognitive impairment is one of the core clinical manifestations of schizophrenia and can contribute to poor daily functioning and quality of life in patients (1). This impairment is highly heterogeneous and variable over time in a domain-selective as well as a general manner and affects the majority of patients and their siblings (2,3). General cognition is a highly heritable phenotype with estimated single-nucleotide polymorphisms (SNPs) based heritability of up to 30% (4,5). To date, more than 100 loci have been associated with general cognitive function. Of these loci, most of the associated variants are located in the intronic or intergenic regions (6). Similarly, family-based heritability of up to 85% is observed in specific cognitive domains, such as verbal fluency, early visual perception, spatial working memory and IQ (7). Findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) Study showed that cognitive impairment clusters frequently within families, whereby the cognitive performance of first-degree relatives of patients with schizophrenia is lower than performance in the general population, but less severe compared to their affected proband (8). This evidence is further supported by the significant familial transmission of attention, memory, and verbal learning impairment in schizophrenia (9).

Given the highly polygenic nature of cognition and schizophrenia, and as the cognitive impairment is defined as one of the features of schizophrenia, the endophenotype hypothesis states that alleles associated with the low cognitive ability also increase the risk for schizophrenia. Within the framework of the Consortium on the Genetics of Schizophrenia (COGS) study, genome-wide association analysis of 12 quantitative neurophysiological and neurocognitive endophenotypes for schizophrenia found more than seven genome-wide significant regions containing several genes of interest for seven endophenotypes related to executive function, inhibition, episodic memory, attention, learning, vigilance, working memory and social cognition domains (10-12). Other reports from the B-SNIP Consortium and Cognitive Genomics consortium (COGENT) showed a bidirectional genetic association between cognition and schizophrenia in both non-clinical and clinical samples. Polygenic risk scores for schizophrenia (PRS_{scz}) was associated with low general cognition and polygenic risk score for low cognition was associated with greater risk for schizophrenia (4,13,14). Additionally, the COGENT reported SNPs that have previously been strongly

associated with schizophrenia also had nominally significant associations with cognition in healthy individuals (14). A meta-analysis of genome-wide association studies of COGENT, Psychiatric Genomics Consortium, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium and UK Biobank cohort also revealed that 21 loci associated to both schizophrenia and verbal-numerical reasoning, reaction time and general cognitive ability (15).

Cognitive function is a suitable objectively measurable (endo)phenotype (16,17). However, results suggest that endophenotypes are equally complex as the disease, which makes it difficult to conclude about the extent of heterogeneity, variability over time and their predictors. To overcome the clinical heterogeneity, molecular complexity and facilitate gene discovery of schizophrenia, the use of endophenotypes and subtyping approaches were advocated as a promising alternative strategy in psychiatric research (7,18). To explore the heterogeneity of cognitive function in a Dutch population, we first cross-sectionally examined cognitive heterogeneity in non-affected siblings of patients (19) and subsequently explored cognitive trajectories over six years in patients and their siblings (20). To date, up to five clinically and statistically meaningful homogenous cognitive subtypes have been identified (21). The course of cognitive functioning over time is characterized by stability, progressive degeneration, relapsing and progressive improvement trajectory (21,22). Our study and others report showed that trajectory of cognitive impairment can be associated with exposure to multifactorial clinical and sociodemographic factors, including aging, male gender, low educational status, low IQ, high dose second-generation antipsychotics, late age of illness onset, long duration of illness, severe positive and/or negative symptoms, and poor functioning and quality of life (21). Nevertheless, none of the previous studies determined the effect of underlying genetic susceptibility either of a single or an aggregated effect of schizophrenia-associated SNPs on cognitive trajectories. Identification of the sum effects of SNPs on homogeneous trajectories may help to accurately predict the clinical course of schizophrenia.

Here, we extended previous knowledge and investigated the trajectories of general cognitive function among patients with schizophrenia spectrum disorders, their unaffected siblings and healthy controls using longitudinal family-based data. We further examined the effect of PRS_{SCZ} on cognitive performance at baseline and over the six years.

METHODS AND MATERIALS

Study procedure and population

Data were extracted from the Genetic Risk and Outcome of Psychosis (GROUP) cohort (data release 7.00), a six-year longitudinal multi-center national study in the Netherlands (23). In total, 2 764 individuals were eligible at baseline: 1 119 patients with a schizophrenia spectrum disorder, 1 059 unaffected first-degree relatives and 586 unrelated healthy controls. Outpatients or inpatients presenting at the mental health service centers were consecutively recruited. Patients with the age range of 16 to 50 years (both extremes included), a diagnosis of schizophrenia spectrum disorder based on the 4th edition Diagnostic and Statistical Manual of Mental Disorders (DSM IV) criteria, good command of the Dutch language, and willing and capable of giving written informed consent were included. Similar criteria, excluding a diagnosis of psychosis, were applied to the siblings (more than one per family is expected). Controls were included if they had no known lifetime psychotic disorder and no first-degree family member with a lifetime psychotic disorder. The main reason to use the age range of 16 to 50 years was that psychosis is presumed to be prevalent to occur between in this age range. This age limit was also used to minimize the effect of age-related cognitive decline. Genotype data were collected at baseline, whereas clinical and demographic data were collected at baseline, third year and sixth year using various standard tools. This study was conducted in line with the published study protocol (23).

Sociodemographic and clinical variables

Age, gender, ethnicity, years of education and marital status were included sociodemographic variables. Premorbid IQ, premorbid adjustment, age of illness onset, duration of illness, psychotic episode, use of antipsychotics, psychotic symptoms, schizotypy, psychotic-like experience, social and occupational functioning, and health-related quality of life were included clinical variables. Schizophrenia spectrum disorders were ascertained using Schedules for Clinical Assessment in Neuropsychiatry (SCAN) or Comprehensive Assessment of Symptoms and History (CASH) (24) instrument (25,26). The Premorbid Adjustment Scale (PAS) was used to assess premorbid functioning (27). Severity of positive and negative symptoms in the patients was assessed by the positive and negative syndrome scale (PANSS), which is a 30-item clinician-rated questionnaire and currently recognized

as the most widely used scale (28). Positive and negative schizotypy in siblings and healthy controls were assessed using a reliable scale known as the Structured Interview for Schizotypy revised (SISR) (29-31). Community Assessment of Psychic Experiences (CAPE), a 42-item reliable and valid self-report questionnaire, was used to assess the psychotic-like experiences in the community and in individuals at an increased risk for developing psychosis (32). The World Health Organization Quality of Life BREF (WHOQOLBREF 1998) questionnaire, which has high construct validity and reliability, was used to assess health-related quality of life (33), whereas Social Functioning Scale (SFS) was used to measure functioning (34).

Cognitive function assessment

Cognitive function was assessed using Word Learning Task (WLT) (immediate recall, delayed recall and retention rate), Continuous Performance Test HQ (CPTHQ) (CPT performance index and CPT variability), Wechsler Adult Intelligence Scale (WAISIII) (digit symbol substitution, information, calculation, block design), Response Set shifting Task (RST), Benton Facial Recognition Test (BFT), Degraded Affect Recognition Task (DFAR) and Hinting Task (HT) (35-37). HT, RST, and BFT were excluded due to the absence of data at the second and/or third measurement. DFAR was excluded because it measures social cognition. The retention rate component of the Word Learning Task was excluded due to high correlation with immediate and delayed recall. Finally, immediate recall, delayed recall, CPT performance index and variability, digit symbol substitution, information, calculation and block design tests were retained to use for principal component analysis (PCA). Details on the scoring system were published in our previous article (20).

Principal component analysis (PCA)

Unlike our previous studies (19,20), we applied PCA to obtain a weighted standardized composite score from the eight cognitive tests. PCA was done on the correlation matrix using varimax with the Kaiser Normalization rotation method separately in patients, siblings, control and all combined samples. We chose the correlation matrix to standardize cognitive tests given that each cognitive function was measured on a different scale. Principal components were extracted using a two-step approach combining the Kaiser criterion (i.e., Eigenvalue ≥ 1) with the visual inspection of scree plots and their component scores were saved for further analysis. The composite score for general neurocognition was obtained by summing up component scores

and used in the trajectory modeling. PCA has been previously applied in several studies examining the heterogeneity of schizophrenia symptoms (37-43).

Trajectory modeling

A trajectory represents people subgroups with a homogenous symptom profile within a group and a heterogeneous profile between groups over time (44). Distinct clinically and statistically meaningful trajectories in patients, siblings, controls and all samples that have unique cognitive profiles over time were identified using censored normal group-based trajectory modeling (GBTM) using PROC TRAJ in SAS version 9.4 for Windows (45). The drop-out model, which includes a logistic model of drop-out probability per wave of measurement, was used to investigate whether the attrition rates significantly biased group membership probabilities of trajectories (46). For each group, we assumed that the drop-out probability depends on the two previous responses (i.e. baseline and three-year observation). A complete theoretical explanation of concepts, functions, and applications of GBTM is published elsewhere (47-49). We started with one group and repeatedly increased the number of groups with a quadratic polynomial order until the number of groups that best fit to the data were identified. The optimal number of trajectories was selected based on sample size adjusted Bayesian information criterion (BICn), logged Bayes factor (i.e. $2 \cdot \Delta \text{BICn}$) and authors' clinical judgment. ΔBICn calculated by subtracting the absolute value of BICn of less complex model from the absolute value of BICn of the complex model. The model with the lowest BICn and logged Bayes factor > 10 was selected as the best-fitting model. Then, the polynomial order was adjusted until it became significant at the confidence level $\alpha = 0.05$ for all trajectories. To evaluate the model classification accuracy, we calculated the average group posterior probability (AvePP). An AvePP > 0.7 indicates that the model has good accuracy in the classification of individuals (50).

Genotyping and quality control (QC)

Genotype data of 2 812 individuals were generated on a customized Illumina, IPMCN array with 570k SNPs comprising of ~250k common tagging SNPs, 250k exome SNPs (rare, nonsynonymous with minor allele frequency (MAF) $< 1\%$), and ~50k psychiatric-related SNPs. QC procedures were performed using PLINK v1.9 (51). SNPs and samples with call rates below 95% and 98% were removed, respectively. A strict SNP QC involved a MAF threshold $> 10\%$ and Hardy-Weinberg equilibrium

(HWE) p -value $> 1e-05$ that followed by linkage disequilibrium (LD) based SNP pruning ($R^2 < 0.2$). This resulted in ~58k SNPs to assess sex errors, heterozygosity (at F -value < 3 standard deviation (SD)), homozygosity (F -value $> 3SD$) and relatedness by pairwise identity by descent (IBD) values. Duplicate samples (at $\text{pihat} > 0.8$) were removed and remaining pairs were manually checked since this dataset contains family members. After removing failing samples, a regular SNP QC was performed considering SNP call rate of $> 98\%$, HWE $p > 1e-06$ and MAF $> 1\%$. Multidimensional scaling clustering with Hapmap Phase 3 individuals was done to check ancestry and individuals deviated more than $3SD$ from the mean of our sample were removed ($n = 91$). Besides, the first 20 genetic principal components of qualified samples were determined by EIGENSTRAT software using genetic information from the strict SNP QC list (52). Next, strand ambiguous SNPs and duplicate SNPs were removed. Mendelian errors were set to missing followed by another missingness check at a 2% threshold for samples ($n = 8$ excluded) and SNPs, and SNPs with differential missingness between cases and controls were removed. In total, 2 505 individuals and 275k SNPs passed QC steps. SNPs were imputed on the Michigan server using the Haplotype Reference Consortium (HRC) version r1.1 2016 reference panel (53) after phasing with Eagle v2.3. Post-imputation QC involved removing SNPs with INFO score < 0.3 , MAF < 0.01 and discordant MAF compared to the reference panel. Strand ambiguous AT/CG and multi-allelic SNPs were also removed.

Polygenic risk score calculation

We followed a standard approach to calculate the polygenic risk score for schizophrenia. Overlapping SNPs between the Psychiatric Genomics Consortium (PGC) genome-wide association study (training dataset), 1000 reference Genome (reference dataset) and the present sample (target dataset) were selected. Then, insertion/deletion or ambiguous SNPs, SNPs with MAF < 0.01 and imputation quality (R^2) < 0.8 both in the training and target dataset, and SNPs located in complex-LD regions were excluded (54), leaving 2 950k SNPs. Next, these SNPs were clumped in two rounds using PLINK: in the first round with the default parameters (physical distance threshold 250kb and LD threshold (R^2) < 0.5 and with a physical distance threshold of 5,000kb and LD threshold (R^2) < 0.2 in the second round, which resulted in 194k independent SNPs for PRS calculation. Odds ratios for autosomal SNPs were computed in PGC after excluding the present Dutch samples (55) and log-converted to beta values. Finally, four PRS_{scz} were constructed using the PLINK

score function separately for 2 505 individuals at four p-value thresholds (P_{\uparrow}) of $5e-08$ (81 SNPs included), 0.05 (35k SNPs), 0.1 (54k SNPs) and 0.5 (147k SNPs) while adjusting for the 20 genetic principal components to control for any undetected population stratification. To simplify the interpretation, we standardized all the PRSs to a standard normal distribution with a mean of 0 and a standard deviation of 1 (56). Each threshold resulted in a different score for every subject. Our main reason to use the four thresholds was to compare the predictive value of different PRSs that were constructed based on the aggregate effects of SNPs selected at different p-value thresholds. In general, as the number of associative p-value threshold becomes less stringent, the number of informative/predictive SNPs included in the risk estimation becomes larger, and hence the predictive value of PRS of the desired outcome improves as well. A p-value threshold (PT) of ≤ 0.05 was considered the most predictive threshold for schizophrenia (55).

Statistical analysis

A linear mixed-effects model using PROC MIXED in SAS version 9.4 for Windows was used to explore the differences in PRSs and all continuous environmental factors between patients, siblings and controls. Since individuals in the present sample were clustered into families, it is very likely that observations from the same family are statistically correlated and dependent. Therefore, familial clustering was used as a random effect. Using family as a random effect sets up a common correlation among all individuals within the same family. We ignored family as a random effect if the G-matrix was not positive definite. The Maximum Likelihood (ML) estimation method was used to estimate the model parameters and Type-III (overall) tests of difference in fixed-effects (i.e. sample baseline sociodemographic and clinical characteristics) were interpreted. Pair-wise comparison was done if the overall difference was significant. For categorical variables, chi-square tests were used to examine the difference between patients, siblings, and controls. Pearson's correlation test was applied to assess the bivariate association between baseline cognitive function and PRS_{schz} across all sample groups as well as all combined samples. A random-effects ordinal regression analysis using PROC NLMIXED in SAS version 9.4 for Windows (57-59) was done to determine whether PRS_{schz} (each PRS separately modeled) predicted cognitive trajectories by setting 'family' as a random-effect and adjusting covariables under the proportional and non-proportional odds assumption. The covariables adjusted were age, sex, years of education, ethnicity,

premorbid IQ, premorbid adjustment, positive and negative symptoms/schizotypy, and psychotic experiences that reported in our previous study (20). Random-effects ordinal regression analysis was used because cognitive trajectories had more than two ordered categories and study participants were clustered within families. We reported results based on the proportional and non-proportional odds assumption because the proportional odds assumption was not consistently fulfilled for all four PRSs across samples. During modeling, data missingness was handled automatically by using full information maximum likelihood that provides parameter estimates by maximizing the likelihood function of the incomplete data. It is believed the ML estimation method produces unbiased results compared with other methods (60). Moreover, all tests were two-tailed and Bonferroni correction (i.e., correction for multiple comparisons) was applied to the non-proportional odds regression model. To obtain the Bonferroni corrected/adjusted p-value, the original p-value was divided by the number of comparisons we made on the outcome.

RESULTS

Baseline background characteristics of participants

During the six years follow-up period, 1 136 patients with a schizophrenia spectrum disorder, 1 045 unaffected siblings and 583 healthy controls participated. As presented in Table 1, patients were more likely to be male, young and unmarried, and had poor premorbid IQ, premorbid adjustment, functioning and quality of life compared to siblings and controls. Siblings demographic characteristics, functioning and quality of life laid between patients and controls. In general, more than 80% of participants were Caucasians.

Principal components of cognition

Based on the Kaiser criterion and visual inspection of scree plots (Figure S1), the PCA provided three components for each group of samples as well as all combined samples. For nearly all cognitive tests, as shown in Table S1, the component loading was > 0.71 indicating that the correlation between the observed cognitive tests and the extracted components was excellent. In addition, all three principal components accounted for 72.7%, 70%, 68.1% and 72.4% of the cognitive variance in patients, unaffected siblings, controls and all combined samples, respectively.

Cognitive subtypes and trajectories

As illustrated in figure 1, we found five subgroups of patients, labeled as severely impaired (2.7%), moderately impaired (15.8%), mildly impaired (35.7%), normal (33.6%), and high performing (12.3%) group. Except the moderate group, which had a flat/stable trajectory, all groups had a quadratic trajectory (Table 3). In healthy siblings and controls, we identified four trajectory subgroups denoted as moderately impaired (2.5% and 5.7%), mildly impaired (24.5% and 25.8%), normal (42.0% and 47.3%) and high performing (31.1% and 21.3%) group. Moderately impaired sibling group had flat/stable trajectory, mild and high groups had linear trajectory and the normal group had a quadratic trajectory (Table S4). In controls, mild and moderate groups had flat/stable trajectory, whereas normal and high groups linear trajectory (Table S5). In all combined samples, we identified six trajectory subgroups (ranging from very severely impaired to high performance) of individuals with a variable trajectory of general cognitive function over the six years. Severe and very severe groups (7%) had flat/stable trajectory, mild group had linear trajectory and moderate, normal and high groups had a quadratic trajectory (Table S6). The average group posterior probability (AvePP) was 0.73 to 0.96 in all trajectory modeling analyses, indicating that our model had good accuracy of classification of individuals into the respective trajectory groups. A substantial difference in the variability of cognitive function was observed across patient, sibling and control samples. In general, the variability of siblings' cognitive function over time laid between patients and controls. Details on the model fit indices and parameter estimates are presented in supplementary Table S2-S6. Of interest, most individuals belong to similar trajectory groups when analyzed within their corresponding samples (patient, siblings and unrelated controls) and when all samples are pooled together (Table S7).

Comparisons of participants PRSs

The mean PRS_{SCZ} of patients was significantly higher than the mean PRS_{SCZ} of siblings and controls ($p < 0.001$ at $P_T = 0.05$). Similarly, the mean PRS_{SCZ} of siblings was significantly higher than the mean PRS_{SCZ} of controls ($p < 0.001$ at $P_T = 0.05$) (Figure 2, Table S8).

Table 1: Comparisons of participants background characteristics.

| Variable | Participants Controls (C) (n=583) | Siblings (S) (n=1,045) | Patients (P) (n=1,136) | Overall group difference | Pair-wise comparisons |
|---|-----------------------------------|------------------------|------------------------|----------------------------------|-----------------------|
| Background characteristics | | | | | |
| Age, mean(SE) | 30.60 (0.39) | 27.70 (0.28) | 27.40 (0.27) | F= 24.0, p<0.001 | P<C; S<C |
| Gender, male n(%) | 269 (46.10) | 477 (45.60) | 857 (75.40) | X ² = 240.97, p<0.001 | |
| Ethnicity, Caucasian n(%) | 523 (92.10) | 871 (83.60) | 871 (79.10) | X ² = 45.84, p<0.001 | |
| Years of education, mean(SE) | 15.6(0.16) | 13.5(0.12) | 12.4(0.12) | F= 64.5, p<0.001 | P<S & C; S<C |
| Marital status n(%) | | | | | |
| Not married | 319(57.60) | 589(57.50) | 960(87.80) | X ² = 303.17, p<0.001 | |
| Married/Living together | 219(39.50) | 412(40.20) | 104(9.50) | | |
| Other (divorced and widowed) | 16(2.90) | 24(2.30) | 30(2.70) | | |
| Premorbid IQ, mean(SE) | 109.61 (0.69) | 102.58 (0.52) | 94.81 (0.49) | F= 190.93, p<0.001 | P<S & C; S<C |
| Premorbid adjustment, mean(SE) | 1.13 (0.03) | 1.11 (0.02) | 1.98 (0.02) | F= 448.7, p<0.001 | P>S; P>C |
| Age onset illness, mean(SE) | - | - | 231 (0.23) | - | - |
| Duration of illness, mean(SE) | - | - | 4.98(4.46) | - | - |
| Psychotic episode, mean(SE) | - | - | 1.72(1.17) | - | - |
| Use of antipsychotics ^a n(%) | - | - | - | - | - |
| Not currently using | - | - | 38(5.22) | - | - |
| Currently using | - | - | 574(78.85) | - | - |
| Unknown if currently using | - | - | 116(15.93) | - | - |
| Schizotypy, mean(SE) | | | | | |
| Positive | 0.30 (0.02) | 0.38 (0.01) | - | F= 59.45, p<0.001 | S>C |
| Negative | 0.23 (0.01) | 0.27 (0.001) | - | F= 49.25, p<0.001 | S>C |

Table 1: Continued.

| Variable | Participants Controls (C) (n=583) | Siblings (S) (n=1,045) | Patients (P) (n=1,136) | Overall group difference | Pair-wise comparisons |
|--|-----------------------------------|------------------------|------------------------|--------------------------|-----------------------|
| Psychotic symptoms, mean(SE) | | | | | |
| Positive | - | - | 13.90 (0.19) | - | - |
| Negative | - | - | 15.00 (0.19) | - | - |
| Expressive deficit | - | - | 10.72 (0.14) | - | - |
| Social amotivation | - | - | 6.15 (0.09) | - | - |
| Psychotic-like experiences, mean(SE) | | | | | |
| Positive symptoms frequency | 0.19 (0.01) | 0.20 (0.01) | 0.67 (0.01) | F= 571.97, p<0.001 | P>S & C |
| Positive symptoms distress | 0.43(0.03) | 0.46(0.02) | 1.25(0.02) | F=511.36, p<0.001 | P>S & C |
| Negative symptoms frequency | 0.49 (0.02) | 0.54(0.01) | 1.02(0.01) | F= 401.58, p<0.001 | P>S & C |
| Negative symptoms distress | 0.67(0.02) | 0.68(0.02) | 1.25(0.02) | F= 305.75, p<0.001 | P>S & C |
| Depressive symptoms frequency | 0.58(0.02) | 0.62(0.02) | 1.00(0.02) | F= 217.76, p<0.001 | P>S & C |
| Depressive symptoms distress | 0.88(0.03) | 0.92(0.02) | 1.44(0.02) | F= 212.19, p<0.001 | P>S & C |
| General neurocognition | 0.62(0.07) | 0.13(0.05) | -1.18(0.05) | F= 312.77, p<0.001 | P<S & C; S<C |
| Functioning and quality of life LSM(SE) | | | | | |
| Occupational functioning [†] | 8.93(0.13) | 8.92(0.10) | 5.91(0.10) | F=292.98, p<0.0001 | P<S & C |
| Social functioning [†] | 124.04(0.36) | 122.32(0.28) | 112.51(0.27) | F=474.62, p<0.0001 | P<S & C; S<C |
| Health-related quality of life | 4.07(0.02) | 3.97(0.02) | 3.40(0.02) | F=492.16, p<0.0001 | P<S & C; S<C |

α = Used from the second wave, SE=Standard Error

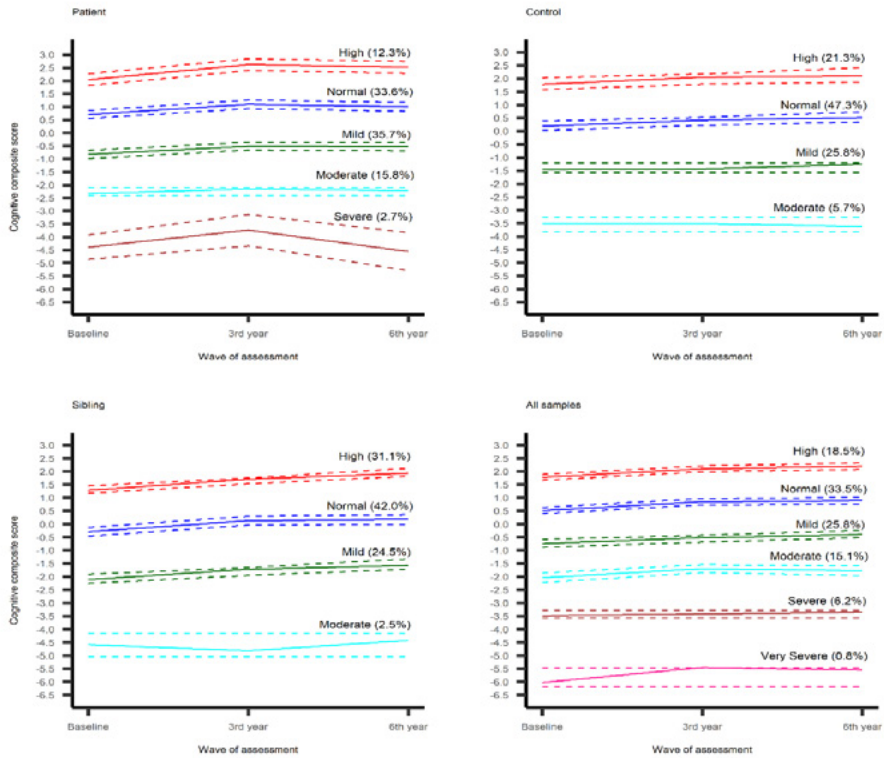


Figure 1: Cognitive subtypes and trajectories.

Association between PRS_{SCZ} and baseline cognitive function

In patients, PRS_{SCZ} was significantly correlated with low digit symbol substitution ($r = -0.11$, $p < 0.01$ at $P_T = 0.1$), calculation ($r = -0.09$, $p < 0.05$ at $P_T = 0.1$), block design ($r = -0.13$, $p < 0.01$ at $P_T = 0.1$), information ($r = -0.08$, $p < 0.05$ at $P_T = 0.1$), general cognitive function ($r = -0.12$, $p < 0.01$ at $P_T = 0.1$) and high CPT-variability ($r = 0.08$, $p < 0.05$ at $P_T = 0.1$) score, but not with immediate recall, delayed recall and CPT-performance index score. In siblings, PRS_{SCZ} was also significantly correlated with low digit symbol substitution ($r = -0.07$, $p < 0.05$ at $P_T = 0.1$), calculation ($r = -0.15$, $p < 0.01$ at $P_T = 0.1$), block design ($r = -0.15$, $p < 0.01$ at $P_T = 0.1$) and information ($r = -0.08$, $p < 0.05$ at $P_T = 0.1$) score, but not with immediate recall, delayed recall, CPT-performance index, CPT-variability and general cognitive function score. Likewise, in controls, PRS_{SCZ} was significantly correlated with low digit symbol substitution ($r = -0.12$, $p < 0.05$ at

$P_T=5 \times 10^{-8}$), calculation ($r=-0.11$, $p < 0.05$ at $P_T=0.1$) and immediate recall ($r=-0.12$, $p < 0.05$ at $P_T=0.05$) score, but not with information, block design, delayed recall, CPT-performance index, CPT-variability and general cognitive function score. After combining all samples, PRS_{SCZ} was significantly correlated with low digit symbol substitution ($r=-0.17$, $p < 0.01$ at $P_T=0.1$), calculation ($r=-0.16$, $p < 0.01$ at $P_T=0.1$), block design ($r=-0.16$, $p < 0.01$ at $P_T=0.1$), information ($r=-0.10$, $p < 0.01$ at $P_T=0.1$), immediate recall ($r=-0.08$, $p < 0.01$ at $P_T=0.1$), delayed recall ($r=-0.07$, $p < 0.01$ at $P_T=5 \times 10^{-8}$), CPT-performance index ($r=-0.06$, $p < 0.05$ at $P_T=0.1$), general cognitive function ($r=-0.18$, $p < 0.01$ at $P_T=0.1$) and high CPT-variability ($r=0.12$, $p < 0.01$ at $P_T=0.1$) score.

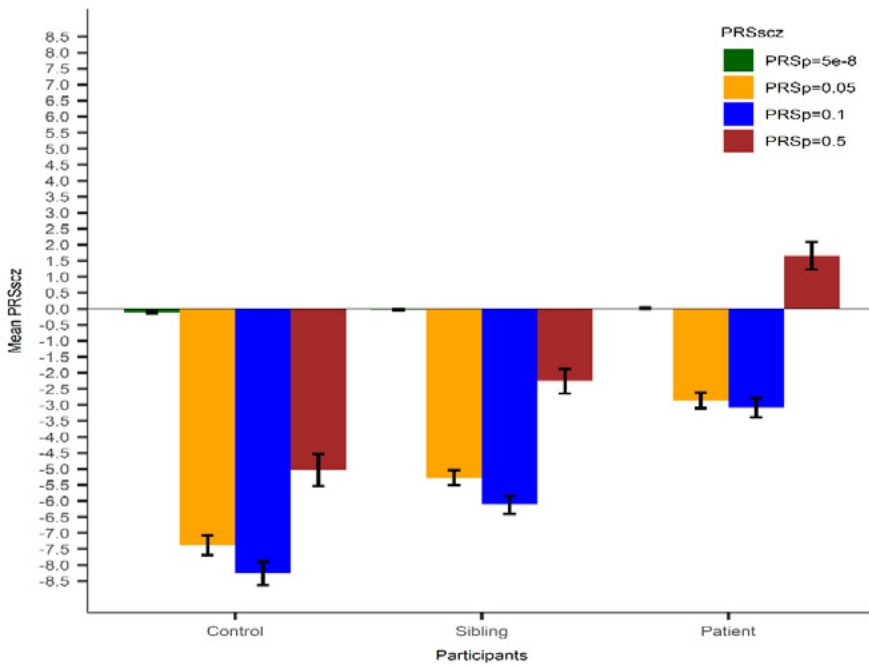


Figure 2: Comparisons of the mean of PRS_{SCZ} across participants.

Association between PRS_{SCZ} and six years cognitive trajectories

In all four PRSs, as presented in Figure 3, individuals with poor cognitive trajectories had high mean PRS_{SCZ} that support the results from the bivariate correlation analyses at baseline.

5

In the univariable proportional and non-proportional odds regression model, PRS_{SCZ} significantly predicted the trajectory of general cognition in patients (at $P_T = 0.05, 0.1$ and 0.5) and all combined samples (at all p-value thresholds). For a 1SD increase in PRS_{SCZ} at $P_T = 0.05$, based on the proportional odds model, the odds of patients being in the impaired group (severely and moderately impaired) was 1.33 (95%CI = 1.08 – 1.65, $p = 0.007$) times the odds of patients in the high performing group (mildly impaired, normal and high). In the combined analysis of all samples of nonproportional odds model, for a 1SD increase in PRS_{SCZ} at $P_T = 0.05$, the odds of individuals being in the impaired group (severely and very severely impaired group) was 1.72 (OR = 1.72, 95%CI = 1.42 – 2.08, $p < 0.0001$) times the odds of individuals in the high performing group (moderately impaired to high performers). Details on comparisons have been presented in Table 2.

In the multivariable regression model, the significant association between PRS_{SCZ} and cognitive subtypes was persisted only in the combined analyses of all samples. While adjusting for non-genetic covariables in the nonproportional odds model, the odds of individuals being in the impaired group (severely and very severely impaired group) was 1.72 (OR = 1.72, 95%CI = 1.18 – 2.50, $p = 0.004$) times the odds of individuals in the high performing group (moderately impaired to high performers) for a 1SD increase in PRS_{SCZ} at $P_T = 0.1$ (Table 3).

DISCUSSION

We found five general cognitive subtypes with variable trajectories in patients (ranging from severely impaired to high functioning), whereas four subtypes were identified in siblings and healthy controls (ranging from moderately impaired to high functioning). Combining all sample groups, we found six cognitive subtypes ranging from very severely impaired to high functioning. Interestingly, all identified cognitive trajectories showed nominal stability, there were small learning effects over time and cognitive degeneration was not observed in the patient sample. PRS_{SCZ} significantly associated with low cognitive function at baseline in all samples. Of note, PRS_{SCZ} is also associated with six-year cognitive trajectories in patients and all combined samples. After Bonferroni correction and adjustment for non-genetic factors, only the results in all combined samples remained significant. In patients, we found five general cognitive subtypes with substantial variability of trajectories during the six years.

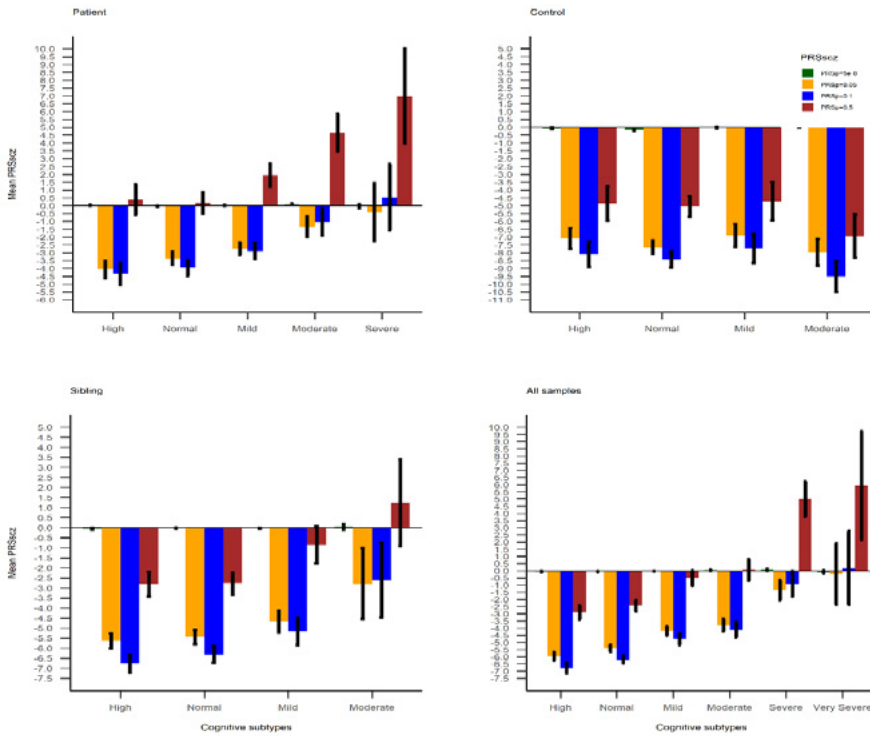


Figure 3: The mean of PRS_{Sz} of cognitive subtypes.

A cross-sectional study (61) identified five clusters of cognitive functioning in patients with schizophrenia, but they did not evaluate the trajectory over time. In agreement with previous cross-sectional studies (41,62-65), we found four cognitive trajectory groups in unaffected siblings. Needless to mention, the present findings inpatients and siblings were similar to our previous study (20) even though we followed a slightly different method to generate the composite score of cognition. In the current study, PCA was used to constructed composite scores as a measure of general cognition. We further identified four cognitive trajectory groups in healthy controls. Moreover, we found six cognitive subtypes for the first time in all combined samples irrespective of their diagnostic status. To date, only one study (66) identified three subtypes by combining controls and patients. Overall, we distinguished more subgroups with homogenous cognitive profiles, as compared to most of the previous studies, which may be because we included large samples with a longer follow-up period and used advanced data-driven statistical methods.

We observed both normally functioning (31.1% resp. 21.3%) and severely impaired (5.7% resp. 2.7%) groups in patients as well as in healthy individuals. This supports the previous finding that patients with schizophrenia may sometimes exhibit even better cognitive performance than their matched healthy controls (67). In addition, the presence of severe cognitive impairment in healthy controls shows that there is great variability in cognitive functioning. It also suggests that cognitive heterogeneity may not be fully explained by disease-related factors (as seen in patients) or shared genes and environmental exposures (as seen in siblings) because non-related healthy subjects are independent of disease-related and shared environmental factors. In patients, controls and all combined samples, 15.8%, 31.5% and 7% had stable cognitive trajectory during the six years. Subjects (independent from their phenotypes) stay at the same state of cognitive performance over six years. When they start in low levels of cognition, they stay low, and the same is true for other subtypes as well. Taking this into account, we have shown in the past and in this study that our primary hypothesis is confirmed and people can be categorized to certain subgroups with homogenous stable cognitive functioning. This supports the notion that cognitive function can be a suitable endophenotype in schizophrenia (16,68). It is conceivable that each cognitive subgroup may have a partly different risk-profile for schizophrenia, predict the incidence of schizophrenia at a different level and associated at a different level with disease genetic and none-genetic risk factors.

In our study, high PRS_{scz} was significantly associated with low baseline IQ (in patients, siblings and controls), verbal learning and memory (in siblings and controls), attention (only in patients) and general cognitive function (only in patients) though unadjusted for covariables. Similarly, the domain-specific as well as general cognitive deficits in combined analyses of all samples cross-sectionally and longitudinally (also after adjustment for covariables) associated with high PRS_{scz} at least for one p-value threshold. A cross-sectional study by the B-SNIP Consortium found that high PRS_{scz} was significantly associated with lower general cognitive performance among healthy individuals, but not in people with psychosis (13). Likewise, a meta-analysis found a negative significant correlation between PRS_{scz} and global cognition in the general healthy population, while no significant correlation was seen in patients with schizophrenia (69). Reports from a number of independent consortia, such as the B-SNIP Consortium (70), COGS (10-12), COGENT (14) and PGC, CHARGE and UK Biobank cohort (15) revealed multiple common genetic variants between

schizophrenia and various endophenotypes of cognition of IQ, executive function, inhibition, episodic memory, attention, learning, vigilance, working memory, verbal-numerical reasoning, reaction time and general cognitive ability in both non-clinical and clinical samples. Altogether, accumulating evidence persists for an existing common genetic susceptibility between schizophrenia and cognitive impairment, yielding the higher genetic susceptibility to schizophrenia, the worse cognitive impairment is observed.

Inconsistencies between studies have been observed even though most studies showed common genetic susceptibility for schizophrenia and cognitive deficits. This difference might be due to a difference in the severity of schizophrenia across cohorts, cognitive assessment tools and their mode of administration, selection of samples (e.g. age difference) and statistical modeling (e.g. adjustment for confounding factors). We refer to one important aspect that, unlike other studies, our study tackled the confounding effect of heterogeneity by focusing on longitudinal trajectories consisting of people with homogeneous cognitive profiles that identified using data-driven methods. This approach empowers the likelihood of finding genetic associations in our study. In contrast, previous studies lumped samples from various cohorts together, which increasing sample diversity, heterogeneity of severity of illness and use of different cognitive battery tests (69). These factors may cause sample dilution and non-significant or weak genetic associations, as a result. We also combined association analysis with a data-driven subtyping approach. The previous studies investigated the link between PRS_{SCZ} and neurocognition using a traditional method (i.e. correlation or regression analysis) and interpreted for all study participants. However, significant associations may come from a specific subgroup of individuals who are the carrier of high-risk SNPs. Additionally, we applied a principal component analysis to obtain a composite score as a measure of general cognition based on the individualized weights of each cognitive test. Constructing a composite score reduces type 1 error by reducing the number of measures to a more manageable level. Furthermore, using composite score improves signal detection by being more sensitive to disease state and treatment effects, more highly correlated with putative biomarkers such as polygenic risk score, and being better at predicting disease progression (71). Taken together, our study identified a larger number of stable cognitive subtypes that strongly predicted by PRS_{SCZ} at least in all combined samples.

Table 2: Univariable random-effects model on the association between PRS_{scz} and cognitive subtypes.

| Sample groups | Proportional-odds model | | Nonproportional-odds model* | | | | | | |
|--------------------------|-------------------------|---------|-----------------------------|-----------------|----------|-----------------|------------|-----------------|----------|
| | OR(95%CI) | P-value | 2-4 vs 1 | 3-4 vs 1-2 | 4 vs 1-3 | 2-5 vs 1 | 3-5 vs 1-2 | 4-5 vs 1-3 | 5 vs 1-4 |
| Controls | | | | | | | | | |
| PRS _{PT-5x10-8} | 1.16(0.93,1.44) | 0.17 | 0.93(0.70,1.22) | 1.36(1.04,1.77) | 0.02 | 1.50(0.89,2.52) | 0.13 | | |
| PRS _{PT-0.05} | 0.99(0.80,1.23) | 0.96 | 0.92(0.68,1.23) | 1.07(0.83,1.37) | 0.61 | 0.90(0.56,1.46) | 0.68 | | |
| PRS _{PT-0.1} | 0.99(0.80,1.23) | 0.93 | 0.95(0.71,1.28) | 1.04(0.81,1.33) | 0.77 | 0.84(0.51,1.38) | 0.49 | | |
| PRS _{PT-0.5} | 0.96(0.77,1.20) | 0.73 | 0.97(0.72,1.30) | 0.98(0.76,1.26) | 0.87 | 0.81(0.48,1.34) | 0.41 | | |
| Siblings | | | | | | | | | |
| PRS _{PT-5x10-8} | 1.03(0.86,1.23) | 0.74 | 1.07(0.87,1.32) | 0.98(0.79,1.21) | 0.84 | 1.12(0.63,1.99) | 0.70 | | |
| PRS _{PT-0.05} | 1.17(0.98,1.39) | 0.08 | 1.11(0.90,1.37) | 1.20(0.98,1.48) | 0.08 | 1.41(0.86,2.29) | 0.17 | | |
| PRS _{PT-0.1} | 1.23(1.03,1.46) | 0.02 | 1.17(0.94,1.45) | 1.26(1.02,1.55) | 0.03 | 1.46(0.92,2.32) | 0.11 | | |
| PRS _{PT-0.5} | 1.19(1.01,1.42) | 0.047 | 1.10(0.89,1.36) | 1.27(1.03,1.56) | 0.02 | 1.34(0.85,2.13) | 0.21 | | |
| Patients | | | | | | | | | |
| PRS _{PT-5x10-8} | 1.05(0.90,1.23) | 0.51 | 0.96(0.75,1.23) | 1.05(0.89,1.25) | 0.55 | 1.13(0.91,1.42) | 0.27 | 0.92(0.57,1.49) | 0.74 |
| PRS _{PT-0.05} | 1.26(1.08,1.47) | 0.003 | 1.25(0.97,1.62) | 1.21(1.03,1.44) | 0.02 | 1.33(1.08,1.65) | 0.007 | 1.43(0.88,2.32) | 0.15 |
| PRS _{PT-0.1} | 1.28(1.10,1.49) | 0.001 | 1.20(0.93,1.56) | 1.25(1.06,1.48) | 0.01 | 1.36(1.11,1.68) | 0.004 | 1.51(0.96,2.39) | 0.08 |
| PRS _{PT-0.5} | 1.30(1.11,1.51) | 0.001 | 1.14(0.87,1.49) | 1.27(1.07,1.52) | 0.007 | 1.38(1.12,1.69) | 0.002 | 1.51(0.98,2.34) | 0.06 |

Table 2: Continued.

| Sample groups | Proportional-odds model | | Nonproportional-odds model* | | | | | | | | | |
|--------------------------|-------------------------|---------|-----------------------------|------------|-----------------|------------|-----------------|---------|-----------------|---------|-----------------|------|
| | OR(95%CI) | P-value | 2-6 vs 1 | 3-6 vs 1-2 | 4-6 vs 1-3 | 5-6 vs 1-4 | 6 vs 1-5 | P-value | | | | |
| All samples | | | | | | | | | | | | |
| PRS _{PT-9x10-8} | 1.15(1.04,1.28) | 0.006 | 1.08(0.94,1.25) | 0.29 | 1.15(1.03,1.30) | 0.02 | 1.25(1.09,1.44) | 0.002 | 1.23(0.99,1.52) | 0.06 | 0.93(0.51,1.69) | 0.81 |
| PRS _{PT-9.05} | 1.44(1.29,1.60) | <0.0001 | 1.35(1.16,1.57) | 0.0001 | 1.43(1.27,1.62) | <0.0001 | 1.44(1.26,1.65) | <0.0001 | 1.72(1.42,2.08) | <0.0001 | 1.86(1.02,3.36) | 0.04 |
| PRS _{PT-0.1} | 1.45(1.31,1.62) | <0.0001 | 1.34(1.15,1.56) | 0.0002 | 1.45(1.28,1.63) | <0.0001 | 1.48(1.29,1.69) | <0.0001 | 1.76(1.46,2.13) | <0.0001 | 1.83(1.03,3.25) | 0.04 |
| PRS _{PT-0.5} | 1.42(1.28,1.58) | <0.0001 | 1.29(1.11,1.51) | 0.001 | 1.42(1.26,1.60) | <0.0001 | 1.44(1.26,1.64) | <0.0001 | 1.72(1.45,2.05) | <0.0001 | 1.71(0.99,2.95) | 0.06 |

1=High; 2=Normal; 3=Mild; 4=Moderate; 5=Severe; 6=Very Severe.

*=significant p-value set to 0.017 for controls and siblings (0.05 divided by 3 (ie., number of comparisons)), 0.0125 for patients (0.05/4), and 0.01 for all samples (0.05/5), where 3, 4 and 5 are the number of comparisons made in our model

OR = odds ratio, PRS_{scz} = polygenic risk score of schizophrenia (each PRS separately modeled), PRS_{PT} = polygenic risk score with p-value threshold.

Overall, we have seen a clear pattern of results across samples and cognitive measures that suggests specific cognitive assessments used in the present study may be differentially informative and linked with PRS_{SCZ} (16). For example, all IQ measures were significantly associated with PRS_{SCZ} in patients and siblings. In addition, only CPT-variability (a measure of attention domain) associated with PRS_{SCZ} in patients while immediate recall (a measure of verbal learning and memory domain), and calculation and digit symbol substitution (a measure of IQ domain) associated with PRS_{SCZ} in controls. These results showed that cognitive alteration in schizophrenia is selective as well as general and siblings' degree of proneness to cognitive impairment closer to patients than healthy controls (2,3). Additionally, cognitive variation in patients with schizophrenia and healthy individuals arises at least in part due to shared genetic factors (72-74). In general, cognitive deficits are partly genetically determined in patients as well as in healthy subjects based on bivariate correlation analyses.

Our study may provide a more accurate estimate of trajectories and their genetic susceptibility through minimizing methodological biases in cross-sectional studies. The trajectory modeling approach aims to capture inherent patterns on the course of illness longitudinally by characterizing the subgroups of the population with a similar course of symptoms and examining illness characteristics and their predictors (44). Such characterization is sensitive to between- and within-patient heterogeneity in symptoms (66). The application of data-driven methods on endophenotypes can provide an opportunity for discovering new specific treatment strategies to subgroup of population to optimize clinical recovery, evaluating the effects pharmacological treatment and discovering gene if subtypes with different trajectory are accurately identified (44,75,76). Determining the existence of cognitive heterogeneity in genetically non-related healthy samples could be clinically useful and may help to identify individuals with poor cognitive function due to exposure to certain environmental factors and who need early intervention. Our study may also give insight into the molecular basis of the genetic overlap between schizophrenia and cognitive ability to understand the pathogenesis of cognitive impairment in schizophrenia. Another relevant implication of this study is that the results can be baseline evidence for imaging, molecular, and biological subtyping of cognitive impairment in schizophrenia. In addition, the results of this study could be helpful to optimize precision medicine, and thereby, to perform personalizing assessment, diagnosing a high-risk group of population and promptly treating subgroups of patients with poor outcome.

Table 3: Multivariable random-effects model on the association between PRS_{scz} and cognitive subtypes.

| Sample groups | Proportional-odds model | | Nonproportional-odds model ^a | | | | | | P-value |
|--------------------------------|-------------------------|---------|---|------------|-----------------|----------|-------------------|------|---------|
| | OR(95%CI) | P-value | 2-4 vs 1 | 3-4 vs 1-2 | 4 vs 1-3 | 5 vs 1-4 | 6 vs 1-5 | | |
| Controls^a | | | | | | | | | |
| PRS _{PR-Scz0.8} | 0.91(0.67,1.22) | 0.52 | 0.63(0.41,0.97) | 0.04 | 1.37(0.85,2.20) | 0.20 | 1.80(0.59,5.53) | 0.30 | |
| PRS _{PR-0.05} | 1.04(0.77,1.40) | 0.82 | 1.05(0.68,1.62) | 0.83 | 1.03(0.66,1.60) | 0.91 | 0.25(0.05,1.38) | 0.11 | |
| PRS _{PR-0.1} | 1.01(0.74,1.37) | 0.96 | 1.11(0.71,1.75) | 0.64 | 0.99(0.62,1.58) | 0.96 | 0.45(0.02,1.32) | 0.09 | |
| PRS _{PR-0.5} | 0.99(0.73,1.36) | 0.98 | 1.17(0.75,1.84) | 0.48 | 0.86(0.53,1.39) | 0.53 | 0.17(0.02,1.31) | 0.09 | |
| Siblings^b | | | | | | | | | |
| PRS _{PR-Scz0.8} | 0.96(0.80,1.18) | 0.74 | 1.07(0.88,1.31) | 0.50 | 0.98(0.79,1.21) | 0.82 | 1.16(0.63,2.11) | 0.64 | |
| PRS _{PR-0.05} | 1.09(0.87,1.36) | 0.45 | 1.10(0.90,1.36) | 0.35 | 1.21(0.98,1.48) | 0.07 | 1.49(0.91,2.43) | 0.11 | |
| PRS _{PR-0.1} | 1.18(0.94,1.48) | 0.16 | 1.16(0.94,1.43) | 0.16 | 1.26(1.03,1.55) | 0.02 | 1.53(0.96,2.43) | 0.07 | |
| PRS _{PR-0.5} | 1.20(0.95,1.51) | 0.13 | 1.10(0.89,1.35) | 0.39 | 1.27(1.04,1.56) | 0.02 | 1.41(0.89,2.23) | 0.14 | |
| Patients^c | | | | | | | | | |
| PRS _{PR-Scz0.8} | 0.92(0.72,1.18) | 0.51 | 0.88(0.59,1.33) | 0.55 | 0.99(0.70,1.40) | 0.95 | 0.84(0.52,1.35) | 0.46 | |
| PRS _{PR-0.05} | 1.07(0.81,1.42) | 0.62 | 1.26(0.79,2.02) | 0.33 | 1.02(0.70,1.50) | 0.90 | 1.00(0.59,1.68) | 0.99 | |
| PRS _{PR-0.1} | 1.08(0.81,1.44) | 0.60 | 1.25(0.76,2.06) | 0.38 | 1.04(0.70,1.55) | 0.83 | 0.96(0.56,1.63) | 0.88 | |
| PRS _{PR-0.5} | 1.04(0.77,1.40) | 0.78 | 1.09(0.65,1.82) | 0.74 | 0.96(0.65,1.43) | 0.85 | 1.09(0.64,1.87) | 0.74 | |
| All samples^d | | | | | | | | | |
| PRS _{PR-Scz0.8} | 1.08(0.95,1.22) | 0.24 | 1.07(0.88,1.30) | 0.51 | 1.14(0.97,1.35) | 0.11 | 1.04(0.85,1.27) | 0.70 | |
| | | | | | | | 1.01(0.75,1.37) | 0.93 | |
| | | | | | | | 0.04(0.001,28.74) | 0.35 | |

Table 3: Continued.

| Sample groups | Proportional-odds model | | Nonproportional-odds model ^a | | P-value | OR(95%CI) | P-value | OR(95%CI) | P-value | OR(95%CI) | P-value | OR(95%CI) | P-value |
|------------------------|-------------------------|---------|---|---------|-----------------|-----------|-----------------|-----------|-----------------|-----------|------------------|-----------|---------|
| | OR(95%CI) | P-value | OR(95%CI) | P-value | | | | | | | | | |
| PRS _{PI-0.05} | 1.12(0.98,1.29) | 0.09 | 1.01(0.81,1.26) | 0.93 | 1.19(0.99,1.44) | 0.07 | 1.04(0.84,1.29) | 0.73 | 1.58(1.11,2.26) | 0.01 | 1.77(0.12,25.23) | 0.67 | |
| PRS _{PI-0.1} | 1.13(0.98,1.30) | 0.09 | 0.99(0.79,1.25) | 0.97 | 1.19(0.98,1.44) | 0.08 | 1.04(0.83,1.29) | 0.74 | 1.72(1.18,2.50) | 0.004 | 2.13(0.13,34.03) | 0.59 | |
| PRS _{PI-0.5} | 1.11(0.96,1.28) | 0.15 | 0.95(0.75,1.21) | 0.67 | 1.20(0.98,1.46) | 0.07 | 1.01(0.81,1.26) | 0.90 | 1.93(1.33,2.81) | 0.0006 | 1.93(0.16,22.94) | 0.60 | |

1=High; 2=Normal; 3=Mild; 4=Moderate; 5=Severe; 6=Very Severe.

^a=significant p-value set to 0.017 for controls and siblings (0.05 divided by 3 (i.e., number of comparisons), 0.0125 for patients (0.05/4), and 0.01 for all samples (0.05/5), where 3, 4 and 5 are the number of comparisons made in our model.

OR = odds ratio. PRS_{SZ}⁺ = polygenic risk score of schizophrenia (each PRS separately modeled), PRS_{PI}⁻ = polygenic risk score with p-value threshold.

^a= Adjusted for age, sex, years of education, ethnicity, premorbid IQ, premorbid adjustment, and positive and negative schizotypy.

^b= Adjusted for age, sex, years of education, ethnicity, premorbid IQ, premorbid adjustment, positive schizotypy, and frequency of positive psychotic-like experiences.

^c= Adjusted for age, years of education, ethnicity, premorbid IQ, premorbid adjustment, and positive and negative symptoms.

^d= Adjusted for age, sex, years of education, ethnicity, premorbid IQ, and premorbid adjustment.

Besides the longitudinal design and application of advanced methodological and statistical tools, our sample included matched healthy controls for whom similar cognitive and genetic (PRS) data have been available, and therefore, we can directly evaluate to what extent PRS_{scz} contributes to cognitive differences between those with and without the illness. The use of health controls can also ensure the validity of identified cognitive subtypes in our study. To date, only a couple of studies (66,77) have used data from healthy controls or general population, where none of these studies explored the longitudinal cognitive trajectory. Apart from this, we assessed a broad range of cognitive domains. This study has also some limitations. The sample size is small in comparison to the recommended sample size of 2,000 for the application of PRS-based analysis (78). Nevertheless, we directly compared results between patients, siblings, and controls, and we confirmed our results using data from a longitudinally assessed cognitive measurement. Besides, the summary statistics used to construct PRS obtained from the latest genome-wide association study of PGC. Our study has enough power to examine the association in the current sample (23). General limitations of the PRS analysis also applies to our study, which PRS cannot provide specific associations to a single SNP, elucidate functional mechanisms, or capture the effect of common variants in sex chromosomes, and rare and structural variants.

CONCLUSIONS

We found five cognitive subtypes in patients, four in siblings and controls, and six by combining all samples. Despite the huge variability of trajectories observed across sample groups, stability has been commonly seen in most subtypes. After adjustment for environmental factors, PRS_{scz} significantly predicted poor cognitive types in the all combined samples. Our finding confirmed the presence of genetic overlap between schizophrenia and cognitive function, and can give insight about the mechanisms of cognitive deficits. These findings can be helpful to tackle heterogeneity and initiate person-based treatment instead of symptom-based treatment. Identifying subgroups with high risk polygenic load can be beneficial to perform gene fine-mapping and enrichment analyses (79). Further deep endophenotyping by performing trajectory analysis on specific cognitive domain and investigating the effect of genetic risk factors is required as we observed in general cognition. Moreover, characterizing cognitive subtypes based on polygenic

risk score of cognition, educational attainment, attention deficit hyperactivity disorder (ADHD) and major depressive disorder (MDD) is needed in the present sample and can shed more light on the genetic architecture of subtypes.

Acknowledgment

We are grateful to the patients, their families and healthy subjects for their generosity of time and effort. Furthermore, we would like to thank all research personnel involved in the GROUP project, in particular Joyce van Baaren, Erwin Veermans, Ger Driessen, Truda Driesen and Erna van 't Hag.

References

- (1) Savilla K, Kettler L, Galletly C. Relationships between cognitive deficits, symptoms and quality of life in schizophrenia. *Australian & New Zealand Journal of Psychiatry* 2008;42(6):496-504.
- (2) Shmukler AB, Gurovich IY, Agius M, Zaytseva Y. Long-term trajectories of cognitive deficits in schizophrenia: A critical overview. *Eur Psychiatry* 2015;30(8):1002-1010.
- (3) Petrova N, Dorofeikova M. Cognition in schizophrenia: Selective impairment and factors that influence it. *Eur Psychiat* 2017;41:S193.
- (4) Trampush J, Yang M, Yu J, Knowles E, Davies G, Liewald D, et al. GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Mol Psychiatry* 2017;22(3):336.
- (5) Davies G, Armstrong N, Bis JC, Bressler J, Chouraki V, Giddaluru S, et al. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N= 53 949). *Mol Psychiatry* 2015;20(2):183.
- (6) Ohi K, Sumiyoshi C, Fujino H, Yasuda Y, Yamamori H, Fujimoto M, et al. Genetic Overlap between General Cognitive Function and Schizophrenia: A Review of Cognitive GWASs. *International journal of molecular sciences* 2018;19(12):3822.
- (7) Aukes MF, Alizadeh BZ, Sitskoorn MM, Selten J, Sinke RJ, Kemner C, et al. Finding Suitable Phenotypes for Genetic Studies of Schizophrenia: Heritability and Segregation Analysis. *Biological Psychiatry* 2008 15 July 2008;64(2):128-136.
- (8) Hill SK, Reilly JL, Keefe RS, Gold JM, Bishop JR, Gershon ES, et al. Neuropsychological impairments in schizophrenia and psychotic bipolar disorder: findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) study. *Am J Psychiatry* 2013;170(11):1275-1284.
- (9) Bigdeli TB, Nuechterlein KH, Sugar CA, Subotnik KL, Kubarych T, Neale MC, et al. Evidence of shared familial factors influencing neurocognitive endophenotypes in adult- and childhood-onset schizophrenia. *Psychol Med* 2019:1-8.
- (10) Greenwood TA, Swerdlow NR, Gur RE, Cadenhead KS, Calkins ME, Dobie DJ, et al. Genome-wide linkage analyses of 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. *Am J Psychiatry* 2013;170(5):521-532.
- (11) Greenwood TA, Lazzeroni LC, Calkins ME, Freedman R, Green MF, Gur RE, et al. Genetic assessment of additional endophenotypes from the Consortium on the Genetics of Schizophrenia Family Study. *Schizophr Res* 2016;170(1):30-40.
- (12) Greenwood TA, Lazzeroni LC, Maihofer AX, Swerdlow NR, Calkins ME, Freedman R, et al. Genome-wide association of endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia (COGS) study. *JAMA psychiatry* 2019;76(12):1274-1284.
- (13) Shafee R, Nanda P, Padmanabhan JL, Tandon N, Alliey-Rodriguez N, Kalapurakel S, et al. Polygenic risk for schizophrenia and measured domains of cognition in individuals with psychosis and controls. *Translational psychiatry* 2018;8(1):78.
- (14) Lencz T, Knowles E, Davies G, Guha S, Liewald DC, Starr JM, et al. Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: a report from the Cognitive Genomics consortium (COGENT). *Mol Psychiatry* 2014;19(2):168-174.

- (15) Smeland OB, Frei O, Kauppi K, Hill WD, Li W, Wang Y, et al. Identification of genetic loci jointly influencing schizophrenia risk and the cognitive traits of verbal-numerical reasoning, reaction time, and general cognitive function. *JAMA psychiatry* 2017;74(10):1065-1075.
- (16) Gur RE, Calkins ME, Gur RC, Horan WP, Nuechterlein KH, Seidman LJ, et al. The consortium on the genetics of schizophrenia: neurocognitive endophenotypes. *Schizophr Bull* 2006;33(1):49-68.
- (17) Luperdi SC, Tabarés-Seisdedos R, Livianos L, Vieta E, Cuesta MJ, Balanzá-Martínez V. Neurocognitive endophenotypes in schizophrenia and bipolar disorder: A systematic review of longitudinal family studies. *Schizophrenia Research* 2019 August 2019;210:21-29.
- (18) Aukes MF, Alizadeh BZ, Sitskoorn MM, Kemner C, Ophoff RA, Kahn RS. Genetic Overlap Among Intelligence and Other Candidate Endophenotypes for Schizophrenia. *Biological Psychiatry* 2009 15 March 2009;65(6):527-534.
- (19) Quee PJ, Alizadeh BZ, Aleman A, van den Heuvel E, GROUP Investigators. Cognitive subtypes in non-affected siblings of schizophrenia patients: characteristics and profile congruency with affected family members. *Psychol Med* 2014;44(2):395-405.
- (20) Islam MA, Habtewold T, van Es F, Quee P, van den Heuvel E, Alizadeh B, et al. Long term cognitive trajectories and heterogeneity in patients with schizophrenia and their unaffected siblings. *Acta Psychiatr Scand* 2018.
- (21) Habtewold TD, Rodijk LH, Liemburg EJ, Sidorenkov G, Bruggeman R, Alizadeh BZ. Schizophrenia symptoms are inherently heterogeneous: a systematic review of cluster and group-based studies. *bioRxiv* 2019;599498.
- (22) Kochunov P, Hong LE. Neurodevelopmental and neurodegenerative models of schizophrenia: white matter at the center stage. *Schizophr Bull* 2014;40(4):721-728.
- (23) Korver N, Quee PJ, Boos HB, Simons CJ, de Haan L, GROUP investigators. Genetic Risk and Outcome of Psychosis (GROUP), a multi-site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *Int J Methods Psychiatr Res* 2012;21(3):205-221.
- (24) Andreasen NC, Flaum M, Arndt S. The Comprehensive Assessment of Symptoms and History (CASH): an instrument for assessing diagnosis and psychopathology. *Arch Gen Psychiatry* 1992;49(8):615-623.
- (25) Aboraya A, Tien A, Stevenson J, Crosby K. Schedules for Clinical Assessment in Neuropsychiatry (SCAN): introduction to WV's mental health community. *WV Med J* 1998;94(6):326-328.
- (26) Rijnders CT, Van den Berg J, Hodiament P, Nienhuis F, Furer J, Mulder J, et al. Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Soc Psychiatry Psychiatr Epidemiol* 2000;35(8):348-352.
- (27) Cannon-Spoor HE, Potkin SG, Wyatt RJ. Measurement of premorbid adjustment in chronic schizophrenia. *Schizophr Bull* 1982;8(3):470-484.
- (28) Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull* 1987;13(2):261-276.
- (29) Vollema MG, Ormel J. The reliability of the structured interview for schizotypy-revised. *Schizophr Bull* 2000;26(3):619-629.
- (30) Vollema M, Sitskoorn M, Appels M, Kahn R. Does the Schizotypal Personality Questionnaire reflect the biological-genetic vulnerability to schizophrenia? *Schizophr Res* 2002;54(1-2):39-45.

- (31) Kendler KS, Lieberman JA, Walsh D. The Structured Interview for Schizotypy (SIS): a preliminary report. *Schizophr Bull* 1989;15(4):559-571.
- (32) Brenner K, Schmitz N, Pawliuk N, Fathalli F, Joober R, Ciampi A, et al. Validation of the English and French versions of the Community Assessment of Psychic Experiences (CAPE) with a Montreal community sample. *Schizophr Res* 2007;95(1):86-95.
- (33) Trompenaars FJ, Masthoff ED, Van Heck GL, Hodiamont PP, De Vries J. Content validity, construct validity, and reliability of the WHOQOL-Bref in a population of Dutch adult psychiatric outpatients. *Quality of Life Research* 2005;14(1):151-160.
- (34) Birchwood M, Smith J, Cochrane R, Wetton S, Copestake S. The social functioning scale: the development and validation of a new scale of social adjustment for use in family intervention programmes with schizophrenic patients. *The British Journal of Psychiatry* 1990;157(6):853-859.
- (35) Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JM, et al. The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. *Am J Psychiatry* 2008;165(2):214-220.
- (36) Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry* 2008;165(2):203-213.
- (37) Nuechterlein KH, Barch DM, Gold JM, Goldberg TE, Green MF, Heaton RK. Identification of separable cognitive factors in schizophrenia. *Schizophr Res* 2004;72(1):29-39.
- (38) Strauss GP, Horan WP, Kirkpatrick B, Fischer BA, Keller WR, Miski P, et al. Deconstructing negative symptoms of schizophrenia: Avolition-apathy and diminished expression clusters predict clinical presentation and functional outcome. *J Psychiatr Res* 2013;47(6):783-790.
- (39) Wang Y, Neumann D, Shum DHK, Chan RCK. A cross-validation study of clustering of schizotypy using a non-clinical Chinese sample. *Psychiatry Res* 2012;200(1):55-58.
- (40) Barrantes-Vidal N, Lewandowski KE, Kwapił TR. Psychopathology, social adjustment and personality correlates of schizotypy clusters in a large nonclinical sample. *Schizophr Res* 2010;122(1-3):219-225.
- (41) Geisler D, Walton E, Naylor M, Roessner V, Lim KO, Charles Schulz S, et al. Brain structure and function correlates of cognitive subtypes in schizophrenia. *Psychiatry Res Neuroimaging* 2015;234(1):74-83.
- (42) Chang JS, Kim Y, Kim SH, Hwang S, Kim J, Chung I-, et al. Differences in the internal structure of hallucinatory experiences between clinical and nonclinical populations. *Psychiatry Res* 2015;226(1):204-210.
- (43) Bell MD, Corbera S, Johannesen JK, Fiszdon JM, Wexler BE. Social cognitive impairments and negative symptoms in schizophrenia: Are there subtypes with distinct functional correlates? *Schizophr Bull* 2013;39(1):186-196.
- (44) Abdin E, Chong SA, Vaingankar JA, Peh CX, Poon LY, Rao S, et al. Trajectories of positive, negative and general psychopathology symptoms in first episode psychosis and their relationship with functioning over a 2-year follow-up period. *PLoS one* 2017;12(11):e0187141.
- (45) Jones BL, Nagin DS, Roeder K. A SAS procedure based on mixture models for estimating developmental trajectories. *Sociol Methods Res* 2001;29(3):374-393.

- (46) Haviland AM, Jones BL, Nagin DS. Group-based trajectory modeling extended to account for nonrandom participant attrition. *Sociological Methods & Research* 2011;40(2):367-390.
- (47) Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychol Methods* 1999;4(2):139-157.
- (48) Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. *Annu Rev Clin Psychol* 2010;6:109-138.
- (49) Nagin DS. Group-based trajectory modeling: an overview. *Ann Nutr Metab* 2010;65(2-3):205-210.
- (50) Niyonkuru C, Wagner AK, Ozawa H, Amin K, Goyal A, Fabio A. Group-based trajectory analysis applications for prognostic biomarker model development in severe TBI: a practical example. *J Neurotrauma* 2013;30(11):938-945.
- (51) 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):559-575.
- (52) Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38(8):904.
- (53) McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016;48(10):1279-1283.
- (54) Price AL, Weale ME, Patterson N, Myers SR, Need AC, Shianna KV, et al. Long-range LD can confound genome scans in admixed populations. *Am J Hum Genet* 2008 Jul;83(1):132-5; author reply 135-9.
- (55) Ripke S, Neale BM, Corvin A, Walters JTR, Farh K, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;511(7510):421-7.
- (56) Lewis CM, Vassos E. Prospects for using risk scores in polygenic medicine. *Genome medicine* 2017;9(1):96.
- (57) Hedeker D. Generalized linear mixed models. *Encyclopedia of statistics in behavioral science* 2005.
- (58) Hedeker D. Multilevel models for ordinal and nominal variables. *Handbook of multilevel analysis*: Springer; 2008. p. 237-274.
- (59) Hedeker D. Methods for multilevel ordinal data in prevention research. *Prevention Science* 2015;16(7):997-1006.
- (60) Dong Y, Peng CJ. Principled missing data methods for researchers. *SpringerPlus* 2013;2(1):222.
- (61) Dawes SE, Jeste DV, Palmer BW. Cognitive profiles in persons with chronic schizophrenia. *J Clin Exp Neuropsychol* 2011;33(8):929-936.
- (62) Reser MP, Allott KA, Killackey E, Farhall J, Cotton SM. Exploring cognitive heterogeneity in first-episode psychosis: What cluster analysis can reveal. *Psychiatry Res* 2015;229(3):819-827.
- (63) Rangel A, Muñoz C, Ocampo MV, Quintero C, Escobar M, Botero S, et al. Neurocognitive subtypes of schizophrenia. *Actas Esp Psiquiatr* 2015;43(3):80-90.
- (64) Lewandowski KE, Baker JT, McCarthy JM, Norris LA, Öngür D. Reproducibility of cognitive profiles in psychosis using cluster analysis. *J Int Neuropsychol Soc* 2018;24(4):382-390.

- (65) Lewandowski K, Sperry S, Cohen B, Öngür D. Cognitive variability in psychotic disorders: a cross-diagnostic cluster analysis. *Psychol Med* 2014;44(15):3239-3248.
- (66) Thompson WK, Savla GN, Vahia IV, Depp CA, O'Hara R, Jeste DV, et al. Characterizing trajectories of cognitive functioning in older adults with schizophrenia: Does method matter? *Schizophr Res* 2013;143(1):90-96.
- (67) Bechi M, Spangaro M, Agostoni G, Bosinelli F, Buonocore M, Bianchi L, et al. Intellectual and cognitive profiles in patients affected by schizophrenia. *J Neuropsychol* 2018 Apr 23.
- (68) Gur RE, Nimgaonkar VL, Almasy L, Calkins ME, Ragland JD, Pogue-Geile MF, et al. Neurocognitive endophenotypes in a multiplex multigenerational family study of schizophrenia. *Am J Psychiatry* 2007;164(5):813-819.
- (69) Mallet J, Le Strat Y, Dubertret C, Gorwood P. Polygenic Risk Scores Shed Light on the Relationship between Schizophrenia and Cognitive Functioning: Review and Meta-Analysis. *Journal of Clinical Medicine* 2020;9(2):341.
- (70) Bishop J, Hill S, Mills L, Alliey-Rodriguez N, Reilly J, Shaffee R, et al. M94. Genetic Analyses of Cognitive Performance in Psychotic Disorders From the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) Consortium. *Schizophr Bull* 2017;43(suppl_1):S245-S245.
- (71) Riordan H. Constructing composites to optimise cognitive outcomes. *J Clin Stud* 2017;9(2):40-45.
- (72) Avramopoulos D. Recent Advances in the Genetics of Schizophrenia. *Mol Neuropsychiatry* 2018 Jun;4(1):35-51.
- (73) Richards AL, Pardiñas AF, Frizzati A, Tansey KE, Lynham AJ, Holmans P, et al. The Relationship Between Polygenic Risk Scores and Cognition in Schizophrenia. *Schizophr Bull* 2019.
- (74) Xavier RM, Dungan JR, Keefe RSE, Vorderstrasse A. Polygenic signal for symptom dimensions and cognitive performance in patients with chronic schizophrenia. *Schizophrenia Research: Cognition* 2018 June 2018;12:11-19.
- (75) Austin SF, Mors O, Budtz-Jørgensen E, Secher RG, Hjorthøj CR, Bertelsen M, et al. Long-term trajectories of positive and negative symptoms in first episode psychosis: a 10year follow-up study in the OPUS cohort. *Schizophr Res* 2015;168(1):84-91.
- (76) Reilly JL, Sweeney JA. Generalized and specific neurocognitive deficits in psychotic disorders: utility for evaluating pharmacological treatment effects and as intermediate phenotypes for gene discovery. *Schizophr Bull* 2014;40(3):516-522.
- (77) Lysaker PH, Vohs JL, Tsai J. Negative symptoms and concordant impairments in attention in schizophrenia: Associations with social functioning, hope, self-esteem and internalized stigma. *Schizophr Res* 2009;110(1-3):165-172.
- (78) Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry* 2014;55(10):1068-1087.
- (79) Lin S-, Liu C-, Liu Y-, Shen-Jang Fann C, Hsiao P-, Wu J-, et al. Clustering by neurocognition for fine mapping of the schizophrenia susceptibility loci on chromosome 6p. *Genes Brain Behav* 2009;8(8):785-794.
- (80) Andruff H, Carraro N, Thompson A, Gaudreau P, Louvet B. Latent class growth modelling: a tutorial. *Tutor Quant Methods Psychol* 2009;5(1):11-24.

