Data supplement for Dickinson et al., Distinct Polygenic Score Profiles in Schizophrenia Subgroups With Different Trajectories of Cognitive Development. Am J Psychiatry (doi: 10.1176/appi.ajp.2019.19050527)

CONTENTS

SUPPLEMENTARY METHODS

Cluster analysis

Our hypothesis was that the three premorbid/current IQ schizophrenia subgroups identified in earlier studies would emerge from cluster analyses of WRAT and WAIS IQ. To test this, WRAT and WAIS performance data were analyzed for the full schizophrenia sample using the TwoStep Cluster Analysis procedure in SPSS, version 24 (SPSS, Armonk, NY: IBM Corp.).[1,2] To avoid collinearity between the two indicators (r=.50 in schizophrenia sample), we used their average ([WRAT + WAIS]/2) and difference (WRAT – WAIS) (r=-.05) as input variables (r=-.05). We used the stepwise decrease in log-likelihood as the distance measure for identifying clusters and change in the Bayesian Information Criterion (conceptually similar to scree plot analysis) to determine the number of clusters to retain.[1,2] The two-step procedure performs well with continuous variables and is robust to departures from normality.[2,3] It forms small "preclusters" in an initial step and applies traditional clustering algorithms to the preclusters in a second step. The preclustering step makes the procedure sensitive to data input order effects.[2] To counter these effects, unsupervised clustering was performed 1000 times, with random re-orderings, to determine the optimal number of clusters. After determining that three clusters were optimal (see Results), we performed an additional 50 analyses, specifying three clusters for each solution, again with random re-orderings, to determine the assignment of individuals to subgroups. Fleiss's kappa[4] was calculated as an index of the extent of agreement in cluster assignment across runs. General linear model (GLM), chi-square, and logistic regression analyses were used to compare groups and subgroups on demographic and clinical variables that were external to the clustering, controlling for age, sex and race. Pairwise analyses used Fisher's LSD procedure, which controls for multiple comparisons among three groups.[5]

Neuropsychological Assessment

Cluster analyses were focused on the WRAT and a four-subtest estimate of Full Scale IQ, for consistency with earlier subgrouping work.[6] These two variables were part of a larger, comprehensive neuropsychological battery that was administered to all study participants. To provide a more global index of general cognitive ability going beyond current and premorbid IQ, 25 variables were selected to represent the full battery.[7] The 25 variables were selected as meeting 3 principal criteria: (1) they are variables that have been commonly used in schizophrenia research, representing the range of key domains of performance impairment, (2) they had shown evidence of impaired performance in healthy siblings of patients with schizophrenia, suggesting that they might serve as intermediate cognitive phenotypes related to genetic risk for schizophrenia, and (3) they showed good distributional characteristics across schizophrenia cases, unaffected siblings, and controls.

Scores for the following individual measures were included: Logical Memory I & II, Verbal Paired Associates I, and Visual Reproduction I & II from the Wechsler Memory Scale; Digit Symbol, Arithmetic, Similarities, Picture Completion, Digit Span Forward and Backward, and Letter-Number Sequencing from the Wechsler Adult Intelligence Scale; immediate recall from the California Verbal Learning Test; correct responses for the 1-back, 2-back and 3-back conditions from the N-Back Task; categories completed, correct responses and percent perseverative errors from the Wisconsin Card Sorting Test; parts A and B from the Trail Making Test; letter and category fluency; the reading subtest from the Wide Range Achievement Test; and total correct for the Benton Line Orientation Test.

Scores on the different measures were converted to z-scores using control means and standard deviations and then averaged to yield the “general cognition” composite score.[7] The general cognition score was the simple average of z-scores for the individual measures and was not further standardized or normalized.

Genotyping and PGS calculation and analysis

We used standard procedures to extract DNA and obtained genotypes using Illumina Bead Chips (510K-2.5M SNP chips). Pre-imputation quality control procedures for each SNP chip were performed separately, based on previously reported methods.[8] Prior to imputation, phasing was performed using Shapeit version 2.2[9] and then imputation was performed on each chip separately using IMPUTE2.[10] For the densest chip (Illumina Infinium Omni2.5M BeadChip), imputation was performed using the 1000 Genomes Phase 3 data as a reference panel.[11] For all other chips, imputation was performed using the imputed result of the Omni2.5M chip as a reference panel. SNP concordance rates were 98% for all imputed chips. Individual chips were then combined to yield a final imputed genome for each sample, such that only SNPs with high quality imputation (INFO > 0.9 and Certainty >0.9) on all chips were retained. We derived the first 10 principal components (PCs) of the whole genome data using PLINK version 1.90 (https://www.cog-genomics.org/plink/1.9) for use in further analyses as population stratification covariates.

To assess broad differences in subgroup genetics we used available GWAS summary statistics to construct four sets of PGSs in our sample. SNPs used to calculate PGS were selected through linkage-disequilibrium-based clumping in PLINK, with a cutoff of $r^2 = 0.25$ within a 500 kb window, excluding the MHC region of the genome.[12] For each phenotype, we derived PGS at 10 p-value thresholds (ranging from P_T<5×10E-08 to P_T<1.0).[13] Polygenic risk was estimated as the sum of phenotype-associated alleles identified in the relevant GWAS, weighted by their effect sizes.[12] The information contained in each set of 10 scores was concentrated through principal component analysis.[14] Consistent with previous work,[14] we expected the first principal component (PC1) to account for most of the variance across underlying PGS scores for each phenotype and focused on this variable for genetics analyses. Each of the four principal components analyses yielded 2 interpretable components. The first principal components (PC1s) accounted for between 57.7 and 78.3% of the variance in the

different PGS score sets. Our finding that PC1 for schizophrenia explained 65.6% of the variance in the 10 individual schizophrenia PGSs was similar to the 69% figure obtained in the Bergen et al. analysis.[14] The second components (PC2s) accounted for between 14.2 and 20.9% of the variance in the PGS score sets. In each case variable weights for the PC1 were strongest for the more inclusive p-value thresholds (i.e., thresholds from pT=.01 to pT=1.0). In each case variable weights for the PC2 were strongest for the most restrictive p-value thresholds (pT=1x10E-06 and pT=5x10E-08). We interpreted these results as indicating that there were two main polygenic signals for each phenotype reflected in the respective PC1 and PC2 for each.

Creation of cognitive trajectory subgroups in siblings

Cognitive development trajectory subgroup assignments for the 247 unaffected siblings were carried over to each unaffected sibling from his or her affected sibling, yielding parallel unaffected sibling subgroups. As in the schizophrenia subgroups, separate GLM analyses in the sibling cognitive trajectory subgroups tested whether schizophrenia, cognition, educational attainment, and ADHD PGS differed by group. Additionally, multinomial logistic regression tested whether the four PGSs predicted cognitive trajectory subgroup membership (compared to a reduced model without the PGS scores). All PGS analyses controlled for age, sex and ancestry (i.e., the first 10 genomic principal components).

SUPPLEMENTARY RESULTS

Full sample statistics (without ancestry restrictions)

Table 1 in the main text reflects descriptive statistics for the subsamples and schizophrenia subgroups after the ancestry restriction (Caucasians of European descent). Table S1 includes parallel descriptive statistics for the full sample, without any race/ancestry restriction. Apart from racial composition, the descriptive statistics are quite comparable.

(Because the ancestry restriction, there are no PGSs available for the complete unrestricted samples and these scores are not reflected in Table S1.)

Correlations among PGS

See Table S2 for correlations among PGSs for each diagnostic group. Cognition and educational attainment PGSs were moderately positively correlated in the schizophrenia sample (r=.38, p=8.67E-20) and educational attainment and ADHD PGS were more modestly, negatively correlated (r=-.15, p=6.47E-04). Other PGSs were not correlated in the schizophrenia cases. Controls and siblings showed similar patterns and magnitudes of PGSs correlations. For controls, because of larger sample size, the small, positive correlation of schizophrenia and ADHD PGS reached significance (r=.08, p=.018).

Principal components analyses of four PGS sets

Each of the four principal components analyses yielded 2 interpretable components. The first principal components (PC1s) accounted for between 57.7 and 78.3% of the variance in the different PGS score sets. Our finding that PC1 for schizophrenia explained 65.6% of the variance in the 10 individual schizophrenia PGSs was similar to the 69% figure obtained in the Bergen et al. analysis.[14] The second components (PC2s) accounted for between 14.2 and 20.9% of the variance in the PGS score sets. In each case, the loadings of individual PGS scores for the PC1 were strongest for the more inclusive p-value thresholds (i.e., thresholds from pT=.01 to pT=1.0). In each case, individual PGS score loadings for the PC2 were strongest for the most restrictive p-value thresholds (pT=1x10E-06 and pT=5x10E-08). We interpreted these results as suggesting that there were two main polygenic signals for each phenotype reflected in the respective PC1 and PC2 for each.

Multinomial logistic regression of subgroup assignment on four PGSs

Multinomial logistic regression – focused on the first component from principal components analysis of each of the four sets of ten PGS – confirmed that PGS patterns across the four phenotypes significantly predicted cognitive trajectory subgroup membership

(ΔX^2[8]=43.83, p=6.10E-07, ES=.079). Although separate analyses of the four PGSs showed that each PGS differed significantly across the schizophrenia trajectory subgroups (Table 1), in the simultaneous analysis of the four PGSs, the PGS for cognition was no longer individually significant (ΔX^2[2]=3.56, p=.168). The other three PGSs were individually significant in the four-PGS model (schizophrenia, ΔX^2[2]=13.73, p=.001; educational attainment, ΔX^2[2]=10.24, p=.006; ADHD, ΔX^2[2]=7.28, p=.026). A modified regression model including only PGSs for schizophrenia, educational attainment, and ADHD yielded similar results overall (ΔX^2[6]=40.26, p=4.00E-07, ES=.073) and each of the three PGSs in this model was individually significant in the context of other PGSs and covariates (schizophrenia, ΔX^2[2]=14.34, p=.001; educational attainment, ΔX^2[2]=15.44, p=4.44E-04; ADHD, ΔX^2[2]=7.40, p=.025). The substantial correlation between the educational attainment and cognition PGSs (r=.368, Table S2), as well as the change in the individual statistics for the educational attainment PGS between the four-PGS and the three-PGS regression models, suggest collinearity between educational attainment and cognition PGSs in relation to the schizophrenia cognitive trajectory subgroups. Substantial genetic level association between cognition and educational attainment has been found in large scale investigations.[15,16]

Separately, to respond to a reviewer's inquiry, we conducted a number of *post hoc* analyses to confirm that the four PC1s were good variables to use to summarize and concentrate the polygenic signal for the four phenotypes, while minimizing multiple comparisons. Specifically, we compared the main multinomial logistic regression using the four PC1s with alternative, parallel analyses. The alternative analyses examined (1) the four PGSs (together) at each of the 10 p-value thresholds, (2) the set of four "optimal" individual PGS in terms of variance explained in schizophrenia subgroup (for schizophrenia and education, pT=0.2; for cognition, pT=.01; for ADHD, pT=.05), and (3) the four PC2s (details in Table S3).

The results clarify that our focus on the four PC1s as key predictor variables was effective compared with various alternatives. Results for the PC1s were stronger than results at each of the 10 p-value thresholds. Results for the PC1s were essentially equivalent to the results for set of four optimal individual PGS; however, the PC1s were derived and analyzed on an *a priori* basis, whereas the optimal individual PGS could only be identified through *post hoc* analyses involving additional comparisons. The PC2s were not significantly related to schizophrenia subgroup, reflecting weak or non-significant associations with subgroup at the most restrictive p-value thresholds for each PGS (pT=1x10E-06 and pT=5x10E-08).

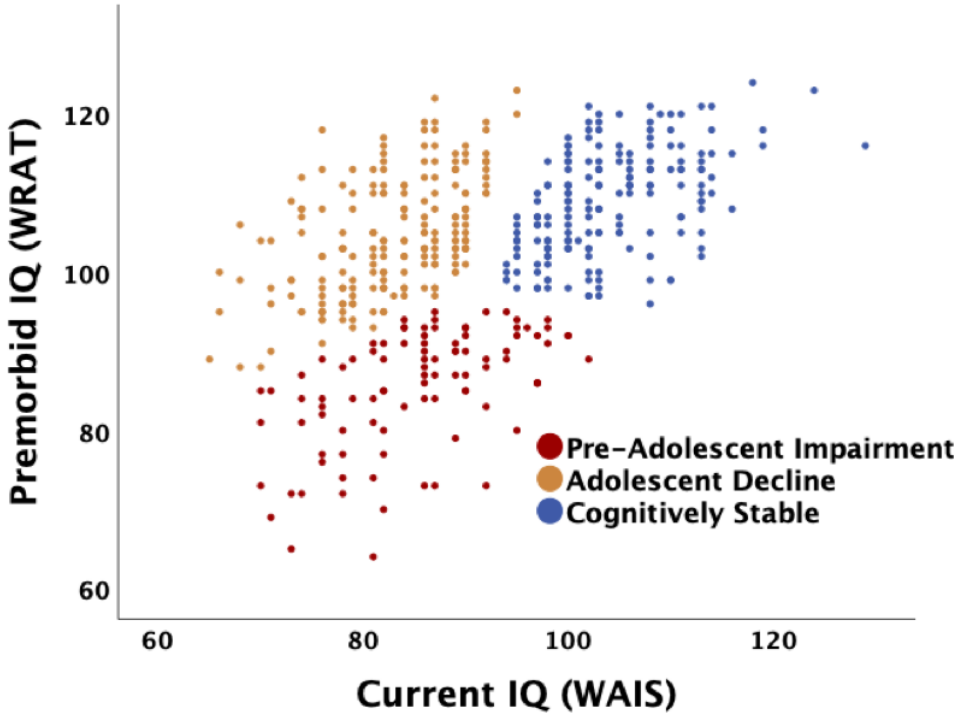
Sensitivity analyses removing those with inconsistent subgroup assignments

In repeated clustering runs to determine subgroup assignments, 86 individuals (11.4%) were less consistently assigned to one specific subgroup than others (i.e., assigned to the same subgroup less than 60% of the time). We tested whether the decision to retain these 86 individuals in main analyses was fundamental to the results of those analyses by repeating the main analyses of schizophrenia subgroups without them. After this exclusion, there were 660 schizophrenia cases across the three clusters for further analyses. Of these, 470 met ancestry restrictions and had genotype information. Table S3 includes a complete set of descriptive statistics (comparable to the main results in Table 1) for the schizophrenia subgroups after this culling (i.e., total N=470). Table S4 includes contrasts of each PGS in each subgroup with controls (comparable to the main results in Table 2). The results for multivariate analyses (reported in the main text, Results, “PGSs in cognitive trajectory subgroups” section) were also consistent. After excluding the 86 participants, as in the unreduced sample, the profile of the four PGSs in multinomial logistic regression predicted cognitive trajectory subgroup membership ($\Delta X^2[8]=33.50$, p=5.0E-05, ES=.068). Thus, across demographic, academic/cognitive, clinical, and functional variables, excluding these 86 individuals from analyses made only modest numerical differences to results, and did not alter our interpretation of the data.

Sibling cognitive trajectory subgroups results

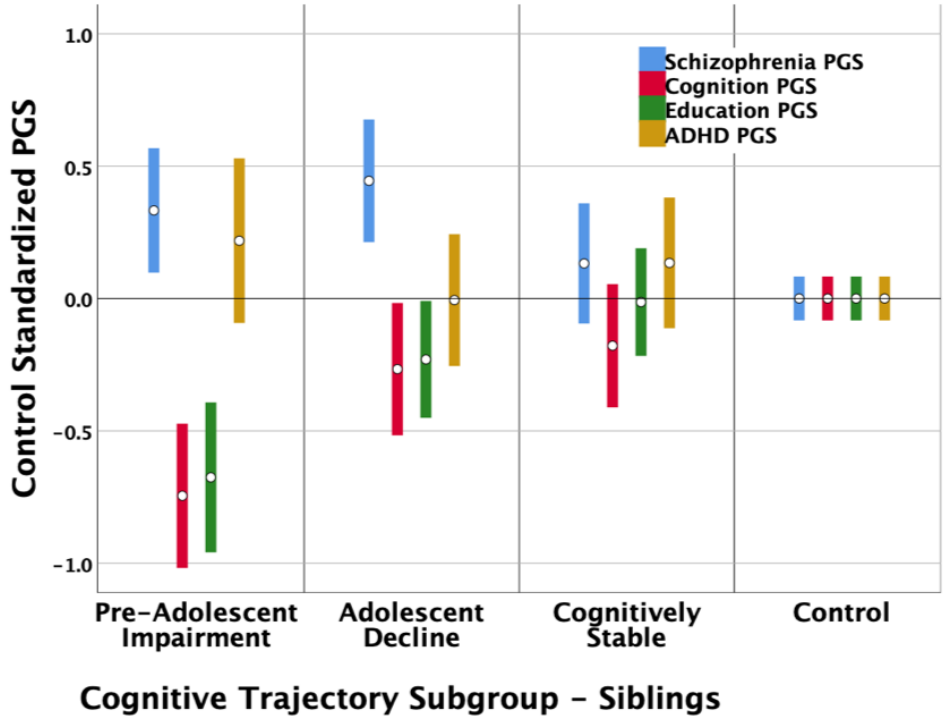
When the cognitive trajectory subgroup assignments were carried over from schizophrenia cases to a subset of 247 of their unaffected siblings, subgroup proportions (19% pre-adolescent impairment, 44% adolescent decline, and 37% cognitively stable) matched those of the schizophrenia cases. Unaffected siblings were more likely to be female, although sex did not differ by sibling subgroup. The most prominent subgroup differences related to academic and cognitive performance (Table S5). Siblings of the cognitively stable schizophrenia cases had higher levels of education than siblings in the other subgroups. The siblings of pre-adolescent impairment cases performed relatively worse on WAIS, WRAT and general cognitive ability measures, while the cognitively stable siblings performed best. Adolescent decline siblings performed at an intermediate level. Adolescent decline siblings did not show the pattern of lower WAIS than WRAT IQ estimates that is so prominent in the adolescent decline schizophrenia cases (compare Table 1 with Table S6).

FIGURE S1. Clustering results in 470 schizophrenia cases as a function of premorbid (WRAT) and current (WAIS) IQ, including only individuals assigned to the same subgroup on >/=60% of clustering runs



The scatterplot shows 470 schizophrenia cases plotted on the basis of WRAT (y-axis) and WAIS (x-axis) scores. Individual cases are labelled by cognitive trajectory subgroup: cognitively stable–blue; pre-adolescent impairment–red; adolescent decline–gold. Comparison of this figure with Figure 1 in the text provides a graphical sense of the effect of eliminating less consistently assigned cases from analyses. As detailed in Supplementary Results and Tables S4 and S5, using this criterion as an exclusion had only minor effects on study results.

FIGURE S2. Polygenic scores (PGSs) by cognitive trajectory subgroup for unaffected siblings (N=247)



Bars represent 95% confidence intervals. Statistical details are in Tables S6 and S7. The figure illustrates the differences in the profiles of PGS for the cognitive trajectory subgroups that were created by carrying the subgroup assignments of individuals in the schizophrenia subgroups to 247 of their unaffected siblings. Thus, this Figure is parallel to Figure 3b, but reflects PGSs for unaffected siblings rather than schizophrenia cases. PGSs were derived in all our samples for schizophrenia (blue), cognition (red), educational attainment (green), and ADHD (gold). It bears emphasis that for schizophrenia and ADHD PGS, higher standardized scores indicate higher disorder risk. For cognition and education PGSs, lower standardized scores predict worse cognitive and academic performance. All PGSs were adjusted to account for age, sex, and population stratification, and then standardized. We used control means and SD's to standardize the PGSs so that controls serve as the reference for differences in PGSs across these unaffected sibling cognitive trajectory subgroups.

TABLE S1. Descriptive statistics for full samples, with no ancestry restriction, by diagnostic group (top) and schizophrenia subgroup (bottom)

Diagnostic Group	Schizophrenia Cases (n=746)		Unaffected Siblings (n=370)		Community Controls (n=1525)		Statistic	df	P-value	Effect Size	Pairwise
	Mean/N	SD/%	Mean/N	SD/%	Mean/N	SD/%					
Demographics:											
Age	33.4	10.0	34.8	10.0	31.8	9.8	F=15.2	2, 2636	2.78E-07	0.011	SC=US>CC
Male	544	72.9%	162	43.8%	663	43.5%	X^2=184.5	2	8.62E-41	0.086	SC>US=SC
Caucasian	599	80.3%	319	86.2%	1074	70.4%	X^2=53.5	2	6.00E-12	0.024	US>SC>CC
Family SES	52.0	12.6	52.4	12.6	50.0	12.8	F=7.0	2,1669	1.00E-03	0.008	SC=US>CC
Functioning:											
Education Years	14.0	2.2	15.9	2.5	16.5	2.5	F=274.8	2, 2613	4.66E-109	0.174	SC<US<CC
Global Functioning	45.1	14.5	84.6	7.4	87.0	4.7	F=5050.0	2, 2526	<.0001	0.801	SC<US<CC
Learning Difficulties	234	31.4%	39	10.6%	332	21.8%	X^2=57.5	2	3.26E-13	0.036	SC>CC>US
Current Employment	218	29.2%	309	83.5%	1202	78.8%	X^2=519.6	2	1.46E-113	0.255	US>CC>SC
Cognition:											
WAIS Full Scale IQ	91.3	11.8	106.0	11.0	106.9	10.5	F=612.7	2, 2598	1.03E-218	0.321	SC<US<CC
WRAT Reading	101.6	11.8	106.2	10.7	107.3	10.1	F=87.0	2, 2603	2.63E-37	0.063	SC<US<CC
General Cognition*	-1.1	0.8	-0.1	0.5	0.0	0.5	F=890.8	2, 2498	9.39E-293	0.416	SC<US<CC
Schizophrenia Subgroup	**Pre-Adolescent Impairment (n=179)**		**Adolescent Decline (n=308)**		**Cognitively Stable (n=259)**		**Statistic**	**df**	**P-value**	**Effect Size**	**Pairwise**
	Mean/N	SD/%	Mean/N	SD/%	Mean/N	SD/%					
Demographics:											
Age	31.7	8.9	31.8	9.6	36.4	10.5	F=19.0	2, 742	8.80E-09	0.049	PI=AD<CS
Male	126	70.4%	233	75.6%	185	71.4%	X^2=1.6	2	ns	-	-
Caucasian	123	68.7%	256	83.1%	220	84.9%	X^2=15.7	2	3.84E-04	0.032	PI<AD=CS
Family SES	45.8	13.0	52.5	12.8	55.3	10.8	F=22.8	2, 474	3.49E-10	0.088	PI<AD<CS
Functioning:											
Education Years	13.1	1.9	13.7	2.0	14.9	2.2	F=36.3	2, 738	8.93E-16	0.09	PI<AD<CS
Global Functioning	44.3	12.5	42.1	13.9	49.4	15.6	F=15.6	2, 700	2.47E-07	0.043	PI=AD<CS
Learning Difficulties	80	44.8%	94	30.6%	59	22.7%	X^2=17.0	2	1.99E-04	0.032	PI>AD>CS
Currently Employed	39	21.6%	85	27.7%	98	37.8%	X^2=10.3	2	0.006	0.019	PI=AD<CS
Cognition:											
WAIS Full Scale IQ	85.3	8.7	84.5	6.7	103.5	6.8	F=464.0	2, 741	2.21E-131	0.556	PI=AD<CS
WRAT Reading	85.4	7.9	105.6	7.6	108.3	6.7	F=562.6	2, 741	2.40E-149	0.603	PI<AD<CS
General Cognition*	-1.5	0.6	-1.4	0.6	-0.5	0.5	F=235.8	2, 713	2.47E-79	0.398	PI<AD<CS
Clinical:											
Duration of illness	10.1	7.9	10.5	8.6	13.6	10.1	F=2.4	2, 621	ns	-	-
On antipsychotics	167	93.2%	285	92.6%	236	91.0%	X^2=0.38	2	ns	-	-
CPZE	651	419	607	394	547	366	F=2.8	2, 572	ns	-	-
PANSS Total (30-210)	61.0	20.6	64.8	23.0	53.6	19.0	F=12.4	2, 565	6.00E-06	0.042	PI=AD>CS
Negative (6-42)	16.2	8.9	17.7	9.4	14.0	8.3	F=7.8	2, 603	4.65E-04	0.025	PI=AD>CS
Positive (4-28)	9.0	5.4	10.2	6.0	8.4	5.0	F=4.6	2, 575	0.01	0.016	PI=CS<AD
Disorganized (3-21)	7.8	3.9	7.5	3.9	5.6	3.8	F=20.9	2, 593	1.77E-09	0.066	PI=AD>CS

Analyses control for age and sex. In addition to age and sex covariates, analyses of polygenic scores controlled for 10 ancestry principal components. For pairwise analyses, significance set at p<.05, after accounting for three comparisons. For continuous dependent variables, 'effect size' refers to partial eta2 for the independent variable of interest from GLM analysis and, for categorical dependent variables, to the difference in Nagelkerke R2 estimates between a covariates-only logistic regression model and a model also including the independent variable of interest. 'SC', schizophrenia; 'US', unaffected sibling; 'CC', community control; 'SES', socio-economic status; 'GAF', Global Assessment of Functioning; 'PI', pre-adolescent impairment; 'AD', adolescent decline; 'CS' cognitively stable; 'PANSS', Positive and Negative Syndrome Scale; 'CPZE' chlorpromazine equivalents; 'na', not applicable; 'ns', not significant.

* "General Cognition' is a composite of 25 cognitive variables based on earlier work. Details are provided in the Supplementary Methods.

TABLE S2. Pearson correlations among polygenic scores (PGSs) by diagnostic group

Diagnostic group	Polygenic Score	Schizophrenia PGS	Cognition PGS	Education PGS	ADHD PGS
Schizophrenia (N=540)	**Schizophrenia PGS**	1			
	Cognition PGS	-0.078	1		
	Education PGS	-0.002	.378**	1	
	ADHD PGS	0.078	-0.055	-.146**	1
Sibling (N=247)	**Schizophrenia PGS**	1			
	Cognition PGS	-0.102	1		
	Education PGS	-0.01	.310**	1	
	ADHD PGS	0.074	-0.101	-.219**	1
Control (N=844)	**Schizophrenia PGS**	1			
	Cognition PGS	-0.055	1		
	Education PGS	0.042	.332**	1	
	ADHD PGS	.081*	-0.063	-.187**	1

** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

TABLE S3. Contrast of statistics for main study regression of schizophrenia subgroup on polygenic scores with alternative analyses at different p-value thresholds

P Threshold for Analyses Including PGSs	ΔX^2 for PGSs Plus Covariates Model Relative to Covariates-Only Model (df=8)	P-value for ΔX^2	Effect Size for ΔX^2
Main Study Analysis:			
PC1 for all PGSs	43.83	6.10E-07	0.079
Alternative Analyses with All Four PGSs at Each pT:			
All PGSs at pT=1.0	25.02	.001	0.046
All PGSs at pT=0.5	22.62	.004	0.042
All PGSs at pT=0.2	40.27	2.85E-06	0.073
All PGSs at pT=0.1	37.05	1.13E-05	0.067
All PGSs at pT=0.05	37.67	8.66E-06	0.069
All PGSs at pT=0.01	34.42	3.41E-05	0.063
All PGSs at pT=0.001	28.59	3.74E-04	0.052
All PGSs at pT=1.0E-04	24.09	.002	0.044
All PGSs at pT=1.0E-06	24.86	.002	0.046
All PGSs at pT=5.0E-08	13.09	ns	0.024
Alternative Analysis Using Optimal pT for Each PGS:			
pT=0.2 for Schizophrenia/ Education, pT=0.01 for Cognition, pT=0.05 for ADHD	44.43	4.71E-07	0.080
Alternative Analysis of PC2:			
PC2 for all PGSs	4.30	ns	0.008

'ΔX^2', the chi-squared difference between the full model with covariates and PGSs and a reduced model with covariates only; 'Effect Size', the difference in Nagelkerke R^2 estimates between the full model with covariates and PGSs and a reduced model with covariates only; 'PC1', score for first component from principal components analysis of polygenic scores at 10 p-value thresholds; 'pT', p-value threshold; 'ns', not significant; 'PC2', score for second component from principal components analysis of polygenic scores at 10 p-value thresholds.

TABLE S4. Descriptive statistics – for reduced samples, including only individuals assigned to the same subgroup on >/=60% of clustering runs

Diagnostic Group	Schizophrenia Cases (n=470)		Unaffected Siblings (n=211)		Community Controls (n=844)		Statistic	df	P-value	Effect Size	Pairwise
	Mean/N	SD/%	Mean/N	SD/%	Mean/N	SD/%					
Demographics:											
Age	34.0	10.2	35.1	10.1	31.1	9.7	F=21.8	2, 1521	4.84E-10	0.028	SC=US>CC
Male	356	75.7%	98	46.4%	390	46.2%	X^2=105.9	2	7.92E-25	0.09	SC>US=CC
Caucasian	470	100.0%	211	100.0%	844	100.0%	na	na	na	-	-
Family SES	52.6	11.8	53.0	12.3	51.6	11.9	F=2.7	2. 1022	ns	-	-
Functioning:											
Education Years	14.1	2.1	15.9	2.5	16.6	2.4	F=193.5	2, 1510	1.59E-75	0.204	SC<US<CC
Global Functioning	45.1	14.1	85.2	6.6	87.8	3.9	F=3345.0	2, 1469	<.0001	0.82	SC<US<CC
Learning Difficulties	142	30.2%	20	9.4%	160	19.0%	X^2=22.3	2	2.00E-06	0.026	SC>CC>US
Current Employment	139	29.5%	180	85.1%	533	79.5%	X^2=315.4	2	1.48E-70	0.261	SC<US=CC
Cognition:											
WAIS Full Scale IQ	91.6	12.0	106.6	10.9	109.3	9.2	F=480.0	2, 1503	6.19E-162	0.39	SC<US<CC
WRAT Reading	102.4	11.2	105.9	10.8	109.4	8.4	F=83.5	2, 1509	4.00E-35	0.1	SC<US<CC
General Cognition	-1.1	0.7	-0.9	0.5	0.13	0.4	F=633.4	2, 1463	7.02E-199	0.464	SC<US<CC
Polygenic Scores:											
Schizophrenia	0.41	0.9	-0.01	0.9	-0.30	1.0	F=77.0	2, 1510	1.47E-32	0.093	SC>US>CC
Cognition	-0.07	1.0	-0.18	1.0	0.13	1.0	F=12.2	2, 1510	6.00E-06	0.016	SC=US<CC
Education	-0.01	1.0	-0.16	0.9	0.06	1.0	F=4.3	2, 1510	0.011	0.006	US<SC=CC
ADHD	0.04	1.1	0.05	1.0	-0.04	1.0	F=1.5	2, 1510	ns	-	-
Schizophrenia Subgroup	**Pre-Adolescent Impairment (n=95)**		**Adolescent Decline (n=193)**		**Cognitively Stable (n=182)**		**Statistic**	**df**	**P-value**	**Effect Size**	**Pairwise**
	Mean/N	SD/%	Mean/N	SD/%	Mean/N	SD/%					
Demographics:											
Age	32.3	8.4	31.9	9.8	37.1	10.6	F=14.2	2, 466	1.00E-06	0.058	PI=AD<CS
Male	70	73.7%	153	79.3%	133	73.1%	X^2=0.01	2	ns	-	-
Caucasian	95	100.0%	193	100.0%	182	100.0%	na	na	na	-	-
Family SES	48.7	10.7	51.4	12.7	55.7	10.6	F=10.3	2, 317	4.50E-05	0.061	PI=AD<CS
Functioning:											
Education Years	13.2	1.8	13.7	1.9	14.9	2.3	F=18.3	2, 463	2.16E-08	0.073	PI<AD<CS
Global Functioning	45.2	11.7	41.5	13.3	49.0	15.1	F=10.8	2, 447	2.60E-05	0.046	PI=CS>AD
Learning Difficulties	44	46.2%	59	30.6%	39	21.2%	X^2=12.7	2	4.17E-04	0.039	PI>AD>CS
Currently Employed	18	19.1%	50	26.0%	71	38.8%	X^2=10.2	2	7.70E-04	0.034	PI=AD<CS
Cognition:											
WAIS Full Scale IQ	85.4	7.9	82.8	6.4	104.0	6.4	F=467.3	2, 465	5.50E-112	0.668	AD<PI<CS
WRAT Reading	85.4	7.4	104.6	7.3	108.9	6.6	F=348.4	2, 465	5.35E-90	0.6	PI<AD<CS
General Cognition	-1.5	0.6	-1.5	0.6	-0.5	0.5	F=226.9	2, 450	7.26E-69	0.502	PI=AD<CS
Clinical:											
Duration of illness	11.0	8.2	11.0	8.9	14.4	10.3	F=1.1	2, 447	ns	-	-
On antipsychotics	94	98.9%	178	92.0%	165	90.4%	X^2=3.9	2	0.047	0.021	PI>AD=CS
CPZE	651	433	613	392	541	357	F=2.8	2, 411	ns	-	-
PANSS Total (30-210)	59.4	19.1	65.4	21.8	54.9	19.5	F=8.7	2, 369	2.00E-04	0.045	AD>PI=CS
Negative (6-42)	15.8	8.7	18.0	8.8	14.3	8.4	F=5.6	2, 392	0.004	0.028	PI=AD>CS
Positive (4-28)	8.6	5.0	10.3	6.0	8.7	6.1	F=3.9	2, 373	0.021	0.02	AD>PI=CS
Disorganized (3-21)	7.9	3.6	7.7	4.0	5.5	3.1	F=18.5	2, 383	2.20E-08	0.088	PI=AD>CS
Polygenic Scores:											
Schizophrenia	0.43	1.0	0.56	0.9	0.22	1.0	F=4.5	2, 455	0.012	0.019	AD>CS
Cognition	-0.28	0.9	-0.15	1.0	0.13	1.0	F=5.3	2, 455	0.005	0.023	PI=AD<CS
Education	-0.35	0.9	-0.03	1.0	0.19	0.9	F=7.0	2, 455	0.001	0.03	PI<AD=CS
ADHD	0.36	1.1	-0.02	1.1	-0.06	1.0	F=4.8	2, 455	0.009	0.021	PI>AD=CS

Analyses control for age and sex. In addition to age and sex covariates, analyses of polygenic scores controlled for 10 ancestry principal components. For pairwise analyses, significance set at p<.05, after accounting for three comparisons. For continuous dependent variables, 'effect size' refers to partial eta2 for the independent variable of interest from GLM analysis and, for categorical dependent variables, to the difference in Nagelkerke R2 estimates between a covariates-only logistic regression model and a model also including the independent variable of interest. 'SC', schizophrenia; 'US', unaffected sibling; 'CC', community control; 'SES', socio-economic status; 'PI', pre-adolescent impairment; 'AD', adolescent decline; 'CS' cognitively stable; 'PANSS', Positive and Negative Syndrome Scale; 'CPZE' chlorpromazine equivalents; 'na', not applicable; 'ns', not significant.

* "General Cognition' is a composite of 25 cognitive variables based on earlier work. Details are provided in the Supplementary Methods.

TABLE S5. GLM results for contrasts of each polygenic score (PGS) in each cognitive trajectory subgroup with control PGS – for the reduced samples, including only individuals assigned to the same subgroup on >/=60% of clustering runs

	Cognitively Stable (n=182)		Community Controls (n=844)		Statistic	df	P-value	Effect Size
Polygenic Scores:	**Mean**	**SD**	**Mean**	**SD**				
Schizophrenia	0.22	1.0	-0.30	1.0	F=42.6	1, 996	1.09E-10	0.040
Cognition	0.13	1.0	0.13	1.0	F=0.1	1, 996	ns	-
Education	0.19	0.9	0.06	1.0	F=3.4	1, 996	0.068***	0.003
ADHD	-0.06	1.0	-0.04	1.0	F=0.2	1, 996	ns	-
	Adolescent Decline (n=193)		**Community Controls (n=844)**		**Statistic**	**df**	**P-value**	**Effect Size**
Polygenic Scores:	**Mean**	**SD**	**Mean**	**SD**				
Schizophrenia	0.56	0.9	-0.30	1.0	F=137.9	1, 1023	5.86E-30	0.119
Cognition	-0.15	1.0	0.13	1.0	F=11.3	1, 1023	8.19E-04	0.011
Education	-0.03	1.0	0.06	1.0	F=0.1	1, 1023	ns	-
ADHD	-0.02	1.1	-0.04	1.0	F=0.5	1, 1023	ns	-
	Pre-Adolescent Impairment (n=95)		**Community Controls (n=844)**		**Statistic**	**df**	**P-value**	**Effect Size**
Polygenic Scores:	**Mean**	**SD**	**Mean**	**SD**				
Schizophrenia	0.43	1.0	-0.30	1.0	F=49.2	1, 925	4.58E-12	0.050
Cognition	-0.28	0.9	0.13	1.0	F=16.2	1, 925	6.00E-05	0.017
Education	-0.35	0.9	0.06	1.0	F=14.0	1, 925	1.91E-04	0.015
ADHD	0.36	1.1	-0.04	1.0	F=15.7	1, 925	7.90E-05	0.017

All analyses control for age, sex, and 10 population stratification principal components. 'ns', not significant. 'Effect size' refers to partial eta^2 from GLM analysis.

TABLE S6. Descriptive statistics for PGS analysis sample of 247 unaffected siblings, by cognitive trajectory subgroup

Sibling Subgroup	Pre-Adolescent Impairment (n=46)		Adolescent Decline (n=110)		Cognitively Stable (n=91)		Statistic	df	P-value	Effect Size	Pairwise
	Mean/N	SD/%	Mean/N	SD/%	Mean/N	SD/%					
Demographics:											
Age	33.6	10.3	34.1	10.7	37.4	9.1	F=3.3	2, 243	3.90E-02	0.026	PI=AD<CS
Male	17	37.5%	56	51.1%	42	45.8%	X^2=0.3	2	ns	-	-
Caucasian	46	100.0%	110	100.0%	91	100.0%	na	na	na	-	-
Family SES	51.1	10.5	52.4	12.9	55.5	11.8	F=2.4	2, 193	ns	-	-
Functioning:											
Education Years	14.9	2.1	15.7	2.1	16.6	2.8	F=5.4	2, 237	0.005	0.043	PI=AD<CS
Global Functioning	83.8	7.1	85.5	6.4	85.7	6.0	F=1.2	2, 231	ns	-	-
Learning Difficulties	5	11.8%	15	13.8%	4	3.9%	X^2=3.3	2	ns	-	-
Currently Employed	33	72.2%	96	87.1%	81	88.9%	X^2=3.1	2	ns	-	-
Cognition:											
WAIS Full Scale IQ	101.2	10.7	105.5	10.2	109.8	10.5	F=6.5	2, 229	0.002	0.054	PI<AD<CS
WRAT Reading	99.2	13.1	107.4	9.1	107.9	9.7	F=9.2	2, 231	5.90E-05	0.081	PI<AD=CS
General Cognition*	-0.3	0.5	-0.2	0.5	0.0	0.5	F=7.7	2, 226	0.001	0.064	PI=AD<CS
Polygenic Scores:											
Schizophrenia	0.00	0.7	0.10	0.9	-0.15	0.9	F=1.5	2, 232	ns	-	-
Cognition	-0.59	0.8	-0.08	1.1	0.00	1.0	F=4.5	2, 232	0.012	0.038	PI<AD=CS
Education	-0.55	0.8	-0.13	1.0	0.08	0.9	F=5.5	2, 232	0.005	0.045	PI=AD<CS
ADHD	0.16	0.9	-0.06	1.0	0.11	1.0	F=0.3	2, 232	ns	-	-

Analyses control for age and sex. In addition to age and sex covariates, analyses of polygenic scores controlled for 10 ancestry principal components. For pairwise analyses, significance set at p<.05, after accounting for three comparisons. For continuous dependent variables, 'effect size' refers to partial eta2 for the independent variable of interest from GLM analysis and, for categorical dependent variables, to the difference in Nagelkerke R2 estimates between a covariates-only logistic regression model and a model also including the independent variable of interest. 'SES', socio-economic status; 'PI', pre-adolescent impairment; 'AD', adolescent decline; 'CS' cognitively stable; 'na', not applicable; 'ns', not significant.
* "General Cognition' is a composite of 25 cognitive variables based on earlier work. Details are provided in the Supplementary Methods.

TABLE S7. GLM results for contrasts of each PGS in each cognitive trajectory subgroup with control PGS for 247 unaffected siblings

	Cognitively Stable (n=91)		Community Controls (n=844)		Statistic	df	P-value	Effect Size
Polygenic Scores:	**Mean**	**SD**	**Mean**	**SD**				
Schizophrenia	-0.15	1.0	-0.30	1.0	F=3.4	1, 921	ns	-
Cognition	0.00	1.0	0.13	1.0	F=1.1	1, 921	ns	-
Education	0.08	0.9	0.06	1.0	F=0.8	1, 921	ns	-
ADHD	-0.11	1.0	-0.04	1.0	F=1.6	1, 921	ns	-
	Adolescent Decline (n=110)		**Community Controls (n=844)**		**Statistic**	**df**	**P-value**	**Effect Size**
Polygenic Scores:	**Mean**	**SD**	**Mean**	**SD**				
Schizophrenia	0.10	0.9	-0.30	1.0	F=20.4	1, 940	7.00E-06	0.021
Cognition	-0.08	1.1	0.13	1.0	F=4.1	1, 940	0.044	0.004
Education	-0.13	1.1	0.06	1.0	F=2.6	1, 940	ns	-
ADHD	-0.06	1.0	-0.04	1.0	F=0.1	1, 940	ns	-
	Pre-Adolescent Impairment (n=46)		**Community Controls (n=844)**		**Statistic**	**df**	**P-value**	**Effect Size**
Polygenic Scores:	**Mean**	**SD**	**Mean**	**SD**				
Schizophrenia	0.00	0.7	-0.30	1.0	F=5.8	1, 876	0.016	0.007
Cognition	-0.59	0.8	0.13	1.0	F=21.1	1, 876	5.00E-06	0.024
Education	-0.55	0.8	0.06	1.0	F=13.9	1, 876	2.01E-04	0.016
ADHD	0.15	1.0	-0.04	1.0	F=1.9	1, 876	ns	-

All analyses control for age, sex, and 10 population stratification principal components. 'ns', not significant. 'Effect size' refers to partial eta^2 from GLM analysis.

SUPPLEMENTARY REFERENCES

1. Chiu T, Fang, D., Chen, J., et al. A robust and scalable clustering algorithm for mixed type attributes in large database environment. Paper presented at: The 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2001.
2. Norusis MJ. *IBM SPSS Statistics 19 Statistical Procedures Companion.* Upper Saddle River, NJ: Prentice Hall; 2011.
3. Bacher J, Wenzig, K., Vogles, M. SPSS Two-Step Cluster: A first evaluation. The RC33 Sixth International Conference on Social Science Methodology; 2004; Amsterdam.
4. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin.* 1971;76:378-382.
5. Meier U. A note on the power of Fisher's least significant difference procedure. *Pharm Stat.* 2006;5(4):253-263.
6. Weickert TW, Goldberg TE, Gold JM, et al. Cognitive impairments in patients with schizophrenia displaying preserved and compromised intellect. *Arch Gen Psychiatry.* 2000;57(9):907-913.
7. Dickinson D, Goldberg TE, Gold JM, et al. Cognitive factor structure and invariance in people with schizophrenia, their unaffected siblings, and controls. *Schizophr Bull.* 2011;37(6):1157-1167.
8. Anderson CA, Pettersson FH, Clarke GM, et al. Data quality control in genetic case-control association studies. *Nat Protoc.* 2010;5(9):1564-1573.
9. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2013;10(1):5-6.
10. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.
11. Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* 2015;526(7571):68-74.
12. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 2013;9(3):e1003348.
13. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* 2014;511(7510):421-427.
14. Bergen SE, Ploner A, Howrigan D, et al. Joint Contributions of Rare Copy Number Variants and Common SNPs to Risk for Schizophrenia. *Am J Psychiatry.* 2018:appiajp201817040467.
15. Savage JE, Jansen PR, Stringer S, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet.* 2018;50(7):912-919.
16. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet.* 2018;50(8):1112-1121.