Deconstructing Cognitive impairment in Psychosis:

A Machine Learning Approach

Robert A. McCutcheon, PhD[1,2,3*]

Richard S. E. Keefe, PhD[4]

Philip M. McGuire, PhD[1,2,5]

Andre Marquand, PhD[6,7]

[1]Department of Psychiatry, University of Oxford, Oxford, UK.

[2]Oxford health NHS Foundation Trust, Oxford health NHS Foundation Trust, Oxford, UK.

[3]Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, London, UK.

[4]Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, US

[5]NIHR Oxford Health Biomedical Research Centre, Oxford, UK.

[6]Donders Institute for Brain, Cognition, and Behavior, Radboud University, Nijmegen, The Netherlands

[7]Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands

*Corresponding author: robert.mccutcheon@psych.ox.ac.uk

Word Count: 2,791

Key Points

Question: What is the relationship between cognitive impairment in psychotic disorders and exposure to a variety of risk factors?

Findings: A significant proportion of the cognitive impairment observed in psychotic disorders may reflect greater exposure to cognitively deleterious risk factors in this group.

Meaning: Cognitive impairments in psychotic disorders do not only reflect disease specific effects. When considering interventions, a diagnosis agnostic, symptom targeted approach may therefore be appropriate.

Abstract

Importance
Cognitive functioning is associated with various factors such as age, sex, education, and childhood adversity and is impaired in people with psychosis. In addition to specific effects of the disorder, cognitive impairments may reflect a greater exposure to general risk factors for poor cognition.

Objective
To determine the extent that impairments in cognition in psychosis reflect risk factor exposures.

Design
We examined the relationship between exposures and cognitive function. Predictive modelling was performed using Extreme Gradient Boosting Regression to train a composite cognitive score prediction model with nested cross-validation. SHapley Additive exPlanations values were used to examine the relationship between exposures and cognitive function.

Setting
Data were obtained from the BSNIP-1 and BSNIP-2 studies comprising six sites.

Participants
3,370 participants were included. 840 healthy controls, 709 with schizophrenia, 541 with schizoaffective disorder, 457 with bipolar 1 with psychosis, and 823 relatives of patients

Exposure
Exposures were chosen based on associations with cognition previously identified: age, sex, race/ethnicity, childhood adversity, education, parental education, parental socioeconomic status, parental age at birth, substance use, antipsychotic dose, and diagnosis.

Outcomes
Cognition was assessed using the Brief Assessment of Cognition in Schizophrenia.

Results
Data from 3370 participants was analysed (44% male, mean age 37.9 years). The model predicted cognitive scores with high accuracy: out of sample correlation between predicted and observed cognitive composite score was r_p=0.72 (SD=0.03). Individuals with schizophrenia(z=-1.4), schizoaffective disorder(z=-1.2), and bipolar 1 with psychosis(z=-0.5) all had significantly worse cognitive composite scores than controls. Factors other than diagnosis and medication accounted for much of this impairment (schizophrenia z=-0.73, schizoaffective z=-0.64, bipolar 1 with psychosis z=-0.13). Diagnosis accounted for a lesser proportion of this deficit (schizophrenia z=-0.29, schizoaffective z=-0.15, bipolar 1 with psychosis z=-0.13), and antipsychotic use accounted for a similar deficit across diagnostic groups(schizophrenia z=-0.37, schizoaffective z=-0.33, bipolar 1 with psychosis z=-0.26).

Conclusion
Transdiagnostic factors account for a meaningful share of the variance in cognitive functioning in psychosis. A significant proportion of the cognitive impairment in psychosis may reflect factors relevant to cognitive functioning in the general population. When considering interventions, a diagnosis agnostic, symptom targeted approach may therefore be appropriate.

Keywords: Cognition, Schizophrenia, Bipolar, Schizoaffective, Cognitive symptoms

Introduction

Psychotic disorders are associated with cognitive impairment.[1,2] Cross-sectional studies show that on average, individuals with schizophrenia have an overall impairment of about 1.5 standard deviations compared to control populations, with lesser impairments in people with schizoaffective disorder or bipolar disorder.[3–5]

Several factors associated with psychotic disorders are also associated with poor performance on cognitive tests in the general population. Some, such as reduced educational attainment,[6] may be sequelae of the disorder. Others, such as childhood socioeconomic deprivation and trauma, may be risk factors for both development of a psychotic disorder,[7] and poor performance on cognitive tests.[8,9] Thus, it is difficult to determine the extent to which patient-control differences in cognitive performance reflect disease-specific effects versus a greater burden of cognition detrimental exposures, which will often be more prevalent in patient groups.

Completely matching patient and control groups on factors influencing cognitive performance is not practically feasible. Furthermore, it may be the relationship between these variables is not simply additive, in which case differences in pattern of co-occurrence could also be important, as opposed to solely the group-level burden. An alternative approach is using nonlinear machine learning algorithms in combination with large datasets to identify mappings between predictors (including both disease status and other variables of interest) and cognitive functioning. Interrogation of this mapping may then reveal the contribution of illness effects compared to other variables.

In this study, we applied a machine learning algorithm to data from 840 controls, 1,707 individuals with a psychotic disorder, and 823 first degree relatives of individuals a psychotic disorder, to quantify the extent to which cognitive impairments in psychosis reflect the influence of diagnosis, antipsychotic medication, and sociodemographic factors. Identifying these relationships is valuable not only for aetiological understanding, but also when considering the extent to which interventions for cognitive impairment may be effective transdiagnostically.

Methods

Participants

Clinical data were obtained from 3,370 participants in the BNSIP-1 (Bipolar-Schizophrenia Network on Intermediate Phenotypes) and BSNIP-2 studies (six sites; Baltimore, Boston, Detroit, Chicago, Dallas, Hartford).[10] Participants comprised healthy controls, individuals with schizophrenia, schizoaffective disorder (bipolar and depressive types), and bipolar 1 disorder with psychosis. In addition, BSNIP-1 included first degree relatives of the patient cohorts. The study protocol was approved by the institutional review board at each local site. Written informed consent was obtained after complete description of the study to the volunteers.

Demographic and clinical variables

We included as predictor variables those available with a priori potential association with cognitive functioning.[1] This consisted of age, sex, race/ethnicity, diagnostic group, antipsychotic dose (olanzapine equivalents), parental factors (age of parents at birth, parental highest educational attainment, parental occupation), substance use, childhood trauma scores (physical, emotional, and sexual abuse/neglect), subject educational attainment, and Wide Range Achievement Test 4 reading score.[15] Given that substance use, educational attainment, and reading score may be a consequence of, as well as an antecedent of a psychotic disorder, all analyses were also performed excluding these variables.

Cognitive testing

Cognition was assessed using the Brief Assessment of Cognition in Schizophrenia (BACS).[16] This comprises 6 tests: verbal memory (Verbal Memory), working memory (Digit Sequencing), motor speed (Token Motor), verbal fluency (Verbal Fluency), attention and processing speed (Symbol Coding), and executive function (Tower of London). For each of the 6 tests, the mean and standard deviation calculated within the control population was used to subsequently obtain z-scores for all participants. A composite cognitive score was then obtained for each participant by calculating the mean z-score across tests and dividing by the standard deviation of the control mean z-scores. Any subjects with composite cognitive scores more than 5 standard

deviations from the mean were excluded. Composite cognitive scores of the extreme 0.5% were then winsorized.

Predictive modelling

To investigate the extent to which various variables were associated with cognitive outcomes a predictive modelling approach was employed. This was performed using nested cross-validation (20 outer folds, 5 inner folds), with all imputation, processing and hyperparameter optimisation performed on the training set without allowing leakage to test set. Mean imputation was used for imputation of missing continuous data, and mode imputation for categorical variables, both performed on a group level basis (i.e. within a diagnostic group) and based only on training set data. Standard scaling was then performed for continuous variables via mean centering and division by the standard deviation (based only on training set data).

We used a supervised learning approach, Extreme Gradient Boosting (XGBoost) Regression, to train a composite cognitive score prediction model.[17] Hyperparameter optimisation was performed via grid search only on the inner fold. Performance was evaluated based on R^2 calculated between predicted and observed cognitive scores in the outer fold. Statistical significance was calculated by comparing this coefficient with a null distribution based on permutation of the cognitive score 1000 times. Linear regression was also performed in order to allow comparison with previous approaches that have used linear models.[18,19]

Characterisation of predictor variable associations

We used Shapley Additive exPlanations (SHAP) to interpret variable importance in the model.[4,5] SHAP values provide a measure of the contribution of each feature to the prediction for each subject, e.g. if an individual has a SHAP value of '-1' for the variable 'age', it means that their age contributes '-1' to the prediction of their cognitive function z-score. The SHAP values were computed for each instance in our dataset, combined within a subject where appropriate (e.g. SHAP values for each childhood adversity subscale were summed). This was performed 100 times (resampling with replacement) to allow for bootstrapped estimates.

We investigated the contribution of predictor variables to both within and between group differences. For example, a variable like age might have major importance in predicting cognitive performance within a diagnostic group, but if all diagnostic groups are of similar age, this may not significantly impact between group differences.

Within group analysis

Within each cohort, we examined the extent to which predictor variables contributed to predicted cognitive score. We looked at the standard deviation of the SHAP values for each variable within each diagnostic group, plotting the standard deviation as a proportion of the sum of all standard deviations for that group.

Between group analysis

We investigated the role of predictor variables in predicting cognitive impairment, i.e. cognitive differences between patient groups and healthy controls. For each diagnostic group we found the mean SHAP value for a given variable, and calculated the difference between that and the variable's mean SHAP value in the healthy control group (and unlike the within-group analysis above did not adjust for the total score). This takes into account both the relationship between variable and outcome, and also imbalances in the distribution of predictor variables between groups. For example, if older age has a negative effect on cognitive performance, and individuals with schizophrenia are younger than the control group, then the mean SHAP value for age for individuals with schizophrenia would be positive, whereas if the schizophrenia group was on average older the group's mean SHAP value would be negative. This analysis illustrates, to what extent group differences in the occurrence of predictor variables contribute to observed differences in cognitive scores from the healthy control cohort.

Investigating Domain Specific Relationships

To investigate whether associations between predictor variables and cognitive measures varied between cognitive domains we refit the model multiple times using individual tests of the BACS as the outcome variable and calculated SHAP values for the predictor variable. For each diagnostic group, we compared each BACS test to every other BACS test, using paired t-tests to examine whether the proportion of the

total SHAP value allocated to each predictor variable set differed between BACS test. The Benjamini-Hochberg procedure was used to correct for the high number of comparisons.

Further characterisation of medication associations

Given that antipsychotic use is diagnosis-dependent, effects on cognition could be erroneously assigned to medication rather than illness. To investigate this, we performed three additional analyses.

First, we addressed the possibility that increased medication dosage might be related to increased symptom severity. We added PANSS positive, negative, and general scores to the predictive model, and examined whether this reduced the magnitude of the SHAP values associated with medication dose. We then examined overall model performance when including and excluding medication dose.

Second, we permuted medication dose *within* diagnostic groups 1000 times, each time assessing relative contribution of medication dosage to cognitive impairment. If the association between medication and cognition primarily relates to confounding by diagnosis then within-group permutation should not diminish the association assigned to medication, as the group level differences will be maintained. If medication effects on cognition are distinct from diagnosis, then permutation should markedly reduce the effect associated with medication.

Third, we fit the predictive model solely on individuals with diagnoses of schizophrenia and schizoaffective disorder to examine the associations between medication and cognition when there was no possibility of confounding by diagnosis.

Where appropriate STROBE reporting guidelines were followed when describing findings.[20]

Results

Participants

The study included 3,370 participants, with 1,926 from BSNIP-1 and 1,444 from BSNIP-2. Clinical and demographic details are in Table 1 (further details in eTable 1).

Cognitive Scores

Compared to controls (mean Z=0.0, SD=1.0), individuals with schizophrenia (mean Z=-1.4, SD=1.2, p<0.001), schizoaffective disorder (mean Z=-1.2, SD=1.1, p=<0.001), bipolar 1 with psychosis (mean Z=-0.5, SD=1.1, p=<0.001), schizophrenia relatives (mean Z=-0.5, SD=1.0, p=<0.001), schizoaffective relatives (mean Z=-0.4, SD=1.1, p=<0.001), and bipolar 1 with psychosis relatives (mean Z=-0.1, SD=0.9, p=0.03) showed lower cognitive scores.

Predictive Modelling

Mean out of sample correlation between predicted (on basis of all predictor variables including diagnostic group) and observed cognitive composite score was r_p=0.72 (SD=0.04), with R^2=0.52 (SD=0.06) (eFigure1). This reduced to r_p=0.63 (SD=0.03) R^2=0.39 (SD=0.04) when subject education level, word reading, and substance use were excluded. Similar accuracy was found using linear models (eFigures 2-6).

Characterisation of Predictor Variable Associations Within Diagnostic Groups

Figure 1 illustrates that diagnosis accounted for minimal variance within groups, as expected given individuals within a group share the same diagnosis. Demographic factors and education (including reading score) accounted for most of the variance within a diagnostic group. When substance use and educational attainment were excluded, parental level factors accounted for most of this variance (Figure 1B). Additional details are shown in eFigures 7-8.

Characterisation of Predictor Variable Associations Between Diagnostic Groups

Figure 1 illustrates relationships between predictor variables and cognitive function within diagnostic groups. The between group results reflect both associations

between predictor variables and cognitive function illustrated above, but also differences between groups in patterns of occurrence of the predictor variables.

Using healthy controls as a reference, reductions in subject education (including reading level) was associated with cognitive score reductions in other cohorts (schizophrenia z=-0.48, schizoaffective z=-0.40, bipolar 1 with psychosis z=-0.09, schizophrenia relative z=-0.22, schizoaffective relative z=-0.20, bipolar 1 with psychosis relative z=-0.06) (Figure 2A). Antipsychotic medication showed similar associations across patient cohorts (schizophrenia z=-0.37, schizoaffective z=-0.33, bipolar 1 with psychosis z=-0.26) and less for relatives (schizophrenia relative z=-0.04, schizoaffective relative z=-0.04, bipolar 1 with psychosis relative z=-0.01). Diagnosis-related impairment was highest for schizophrenia (z=-0.29) and similar for other groups (schizoaffective: -0.15, bipolar 1 with psychosis: -0.13, schizophrenia relative: -0.10, schizoaffective relative: -0.09, bipolar 1 with psychosis relative: -0.06). Excluding education and substance use from the model increased the proportion of impairment associated with other variables (Figure 2B). Further associations are shown in eFigures 9-10.

Investigating Domain Specific Relationships

Relationships between predictor variables and other cognitive tests were similar to that observed with the composite score (see eFigures 11-16), with no statistically significant differences observed (p_{fdr}>0.05 all comparisons).

Medication Associations

Including PANSS scores as predictor variables, PANSS negative scores showed a strong association, reducing SHAP values for diagnosis, but only slightly affecting medication values (eFigure 17). The model showed R^2= 0.52 (SD 0.05) without medication dose, and R^2=0.53 (SD 0.05) including dose.

Permuting medication doses within diagnostic categories reduced the association between medication and cognitive function as quantified by reduction in SHAP values (schizophrenia 82% reduction, schizoaffective 80%, bipolar1 with psychosis 79% reduction, schizophrenia relative 88% reduction, schizoaffective relative 83%

reduction, bipolar1 with psychosis relative 100% reduction) indicating medication's contribution was not primarily due to group-level confounding.

When models were fit within a sample containing solely diagnoses of schizophrenia or schizoaffective disorder, medication maintained a similar association (Figure 3). Linear models assigned less importance to medication (eFigure 5), possibly due to their inability to model a plateau effect, overestimating associations at higher doses and underestimating at lower doses (compare Figure 3 and eFigure 6).

Discussion

Using non-linear machine learning algorithms with a large, deeply phenotyped dataset, we found that cognitive impairments across the psychosis spectrum are associated with sociodemographic and environmental factors, such as educational level, childhood trauma, age, ethnicity/race, parental socioeconomic status, and antipsychotic exposure, rather than solely disorder. The differential prevalence of these factors in patient groups could thereby potentially account for a proportion of the cognitive impairments observed in individuals with psychosis.

We recently highlighted that variability of cognitive performance is greater within psychosis cohorts compared to controls.[21] One interpretation of this is that the impact of illness on cognition varies between individuals. The present analysis, however, suggests the increased heterogeneity in patients may reflect increased heterogeneity in cognition relevant exposures seen in this group.

Previous work has used deviation from predictive models to investigate cognitive function in psychosis,[17,21] and our results align with recent findings showing attenuated cognitive deficits in schizophrenia when adjusting for cannabis use, symptoms, and medication.[25] Prior work, has used linear models, fewer predictive features, smaller sample sizes, and trained models using only control samples. The present study included more predictive features, and included patient, control, and family members in model training, maximizing sample size and allowing estimation of antipsychotic associations and the impact of illness status.

These results suggest that cognitive impairments in psychosis may not be solely due to the illness but also to cognitively detrimental factors more common in individuals with psychotic disorders. This is consistent with the cognitive reserve hypothesis, where certain genetic and environmental exposures reduce cognitive function, which increases vulnerability to psychosis.[22] Another interpretation is that some cognitive impairments reflect confounding factors like male sex, socioeconomic deprivation, and ethnic minority status, which increase the likelihood of both cognitive impairments and psychosis through different pathways. Cross-disorder comparisons suggest that lesser cognitive impairment in bipolar disorders may reflect both differences in illness impact and associated sociodemographic factors (e.g. socioeconomic deprivation and antipsychotic exposure). These interpretations are in contrast with, but not mutually exclusive to findings that deviation from familially expected cognitive performance predicts schizophrenia risk.[24]

The possibility that a diagnosis of a psychotic disorder may not account for the majority of impairment seen in the disorder aligns with findings that polygenic risk for schizophrenia is not consistently associated with cognitive impairments in patients with schizophrenia.[25–27] Disentangling the impact of antipsychotic medication is complex given that psychotic symptoms also contribute to poor cognitive performance.[28,29] Our analyses suggest that associations observed between medication dose and cognition do not solely reflect confounding by diagnosis or symptom severity, but we cannot definitively determine if this reflects a direct causal effect or untested confounding sources. Medication may have deleterious effects while at the same time preventing a relapse of positive symptoms, which itself can have an even greater impact on cognitive function. Careful dosing strategies are needed to balance these competing needs.

Strengths and Limitations

Our results derived from a large, thoroughly phenotyped dataset, examining associations between cognitive function and a wide range of predictors within a nonlinear framework. Our findings are thoroughly cross validated thereby avoiding overfitting. Further work would benefit from even larger and more comprehensively characterised samples. While BSNIP is a research sample, its large size and geographically broad recruitment do mean that results are potentially generalisable.

Most predictor variables, such as ethnicity/race and socioeconomic status, are potentially proxy markers for factors with more direct causal impact. Definitively determining causal contributions, however, is not possible given we cannot randomly assign these exposures, and a more formal causal model is mostly precluded given the complexity of the system. This means that inferences may be potentially confounded.

One strength of using a model such as XGBoost regression is that there is no assumption of a linear relationship, nor that residuals follow a specific distribution. This increases flexibility, meaning it may be more suitable for modelling complex, non-linear relationships.[16,30] For example, the non-linear relationship between antipsychotic dose and cognitive function identified by XGBoost may more plausible than an ever increasing dose deficit relationship identified by linear regression.

In conclusion, while observed cognitive deficits in psychosis are large, these may not solely represent illness effects but also reflect sociodemographic and clinical factors. Many of these factors are not psychosis specific and may be modifiable. These results highlight the potential for future interventions (both psychosocial and pharmacological) to possibly have transdiagnostic impact, and supports a multifaceted approach to treating cognitive impairment.

Acknowledgements

RAMs work is funded by a Wellcome Trust Clinical Research Career Development (224625/Z/21/Z). AFM gratefully acknowledges support from the Wellcome Trust under a Psychosis in Cognition grant (226706/Z/22/Z). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission. The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. RAM had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis

Declaration of Interests

RAM has received speaker/consultancy fees from Karuna, Janssen, Boehringer Ingelheim, and Otsuka, and co-directs a company that designs digital resources to support treatment of mental illness. RSEK has received consulting fees in the past 2 years from Karuna, Biogen, Merck, Boehringer Ingelheim, WCG and received royalties from the BACS and VRFCAT.

References

1. McCutcheon RA, Keefe RSE, McGuire PK. Cognitive impairment in schizophrenia: aetiology, pathophysiology, and treatment. *Mol Psychiatry*. Published online January 23, 2023:1-17.

2. Keefe RSE, Harvey PD. Cognitive impairment in schizophrenia. *Handb Exp Pharmacol*. 2012;(213):11-37.

3. Li W, Zhou FC, Zhang L, et al. Comparison of cognitive dysfunction between schizophrenia and bipolar disorder patients: A meta-analysis of comparative studies. *J Affect Disord*. 2020;274(March):652-661.

4. Catalan A, Radua J, Mccutcheon R, et al. Examining the variability of neurocognitive functioning in individuals at clinical high risk for psychosis : a meta-analysis. 2022; (December 2021). doi:10.1038/s41398-022-01961-7

5. Catalan A, McCutcheon RA, Aymerich C, et al. The magnitude and variability of neurocognitive performance in first-episode psychosis: a systematic review and meta-analysis of longitudinal studies. *Transl Psychiatry*. 2024;14(1):15.

6. Lövdén M, Fratiglioni L, Glymour MM, Lindenberger U, Tucker-Drob EM. Education and Cognitive Functioning Across the Life Span. *Psychol Sci Public Interest*. 2020;21(1):6-41.

7. Radua J, Ramella-Cravaro V, Ioannidis JPA, et al. What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry*. 2018;17(1):49-66.

8. Majer M, Nater UM, Lin JMS, Capuron L, Reeves WC. Association of childhood trauma with cognitive function in healthy adults: A pilot study. *BMC Neurol*. 2010;10. doi:10.1186/1471-2377-10-61

9. Garner CL, Raudenbush SW. Neighborhood Effects on Educational Attainment: A Multilevel Analysis. *Sociol Educ*. 1991;64(4):251.

10. Tamminga CA, Ivleva EI, Keshavan MS, et al. Clinical phenotypes of psychosis in the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP). *Am J Psychiatry*. 2013;170(11):1263-1274.

11. Weiss EM, Kemmler G, Deisenhammer EA, Fleischhacker WW, Delazer M. Sex differences in cognitive functions. *Pers Individ Dif*. 2003;35(4):863-875.

12. Kaplan GA, Turrell G, Lynch JW, Everson SA, Helkala EL, Salonen JT. Childhood socioeconomic position and cognitive function in adulthood. *Int J Epidemiol*. 2001;30(2):256-263.

13. Crean RD, Crane NA, Mason BJ. An evidence based review of acute and long-term effects of cannabis use on executive cognitive functions. *J Addict Med*. 2011;5(1):1-8.

14. Whitson S, O'Donoghue B, Hester R, et al. Cognitive ability and metabolic physical health in first-episode psychosis. *Schizophr Res Cogn*. 2021;24(100194):100194.

15. Keefe RSE, Harvey PD, Goldberg TE, et al. Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS). *Schizophr Res*. 2008;102(1-3):108-115.

16. Chen T, Guestrin C. XGBoost. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM; 2016. doi:10.1145/2939672.2939785

17. Hochberger WC, Combs T, Reilly JL, et al. Deviation from expected cognitive ability across psychotic disorders. *Schizophr Res*. 2018;192:300-307.

18. Lundberg SM, Erion G, Chen H, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell*. 2020;2(1):56-67.

19. Lundberg S, Lee SI. A unified approach to interpreting model predictions. *arXiv [csAI]*. Published online May 22, 2017. http://arxiv.org/abs/1705.07874

20. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370(9596):1453-1457.

21. Keefe RSE, Eesley CE, Poe MP. Defining a cognitive function decrement in schizophrenia. *Biol Psychiatry*. 2005;57(6):688-691.

22. Velthorst E, Mollon J, Murray RM, et al. Cognitive functioning throughout adulthood and illness stages in individuals with psychotic disorders and their unaffected siblings. *Mol Psychiatry*. 2021;26(8):4529-4543.

23. Koenen KC, Moffitt TE, Roberts AL, et al. Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. *Am J Psychiatry*. 2009;166(1):50-57.

24. Kendler KS, Ohlsson H, Mezuk B, Sundquist JO, Sundquist K. Observed cognitive performance and deviation from familial cognitive aptitude at age 16 years and ages 18 to 20 years and risk for schizophrenia and bipolar illness in a Swedish national sample. *JAMA Psychiatry*. 2016;73(5):465.

25. Shafee R, Nanda P, Padmanabhan JL, et al. Polygenic risk for schizophrenia and measured domains of cognition in individuals with psychosis and controls. *Transl Psychiatry*. 2018;8(1):78.

26. Engen MJ, Lyngstad SH, Ueland T, et al. Polygenic scores for schizophrenia and general cognitive ability: associations with six cognitive domains, premorbid intelligence, and cognitive composite score in individuals with a psychotic disorder and in healthy controls. *Transl Psychiatry*. 2020;10(1):416.

27. Legge SE, Cardno AG, Allardyce J, et al. Associations between Schizophrenia Polygenic Liability, Symptom Dimensions, and Cognitive Ability in Schizophrenia. *JAMA Psychiatry*. 2021;78(10):1143-1151.

28. Morgan O, Strassnig MT, Moore RC, et al. Accuracy of immediate self-assessment of neurocognitive test performance: Associations with psychiatric diagnosis and longitudinal psychotic symptoms. *J Psychiatr Res*. 2022;156:594-601.

29. Krishnadas R, Ramanathan S, Wong E, Nayak A, Moore B. Residual negative symptoms differentiate cognitive performance in clinically stable patients with schizophrenia and bipolar disorder. *Schizophr Res Treatment*. 2014;2014:785310.

30. Montgomery DC, Peck EA, Vining GG. *Introduction to Linear Regression Analysis*. 6th ed. John Wiley & Sons; 2021.

31. Fink LA, Bernstein D, Handelsman L, Foote J, Lovejoy M. Initial reliability and validity of the childhood trauma interview: a new multidimensional measure of childhood interpersonal trauma. *Am J Psychiatry*. 1995;152(9):1329-1335.

32. Hollingshead AB. *Four Factor Index of Social Status*.; 1975.

	Healthy Control	Schizophrenia	Schizophrenia Relative	Bipolar 1 w/ Psychosis	Bipolar1 w/ psychosis Relative	Schizoaffective Disorder	Schizoaffective Relative
N	840	709	323	457	291	541	209
Male N(%)	351 (41.8)	457 (64.5)	101 (31.3)	172(37.6)	100 (34.4)	234 (43.3)	64 (30.6)
Female N(%)	489 (58.2)	252 (35.5)	222 (68.7)	285(62.4)	191 (65.6)	307 (56.7)	145 (69.4)
Asian N(%)	63 (7.5)	29 (4.1)	4 (1.2)	12(2.6)	6 (2.1)	10 (1.8)	2 (1.0)
Black or African American N(%)	244 (29.0)	336(47.4)	123 (38.1)	81 17.7)	48 (16.5)	208 (38.4)	70 (33.5)
Hispanic/Latinx N(%)	97(11.5)	64 (9.0)	26 (8.0)	50 (10.9)	19 (6.5)	66 (12.2)	20 (9.6)
White N (%)	403(48.0)	249 (35.1)	164 (50.8)	299 (65.4)	214 (73.5)	223 (41.2)	112 (53.6)
Other[e] N(%)	34(4)	31 (4.4)	6 (1.9)	16 (3.4)	4 (1.4)	35 (6.4)	5 (2.3)
Olanzapine Equivalents (mg)	0.0 (0.6)	15.9 (17.5)	2.3 (10.6)	9.5 (14.1)	0.8 (4.0)	13.3 (16.9)	1.5 (5.7)
Age (years)	35.6 (12.0)	37.7 (12.0)	42.9 (15.0)	35.5 (12.0)	40.4 (16.0)	38.8 (12.0)	39.6 (16.0)
Age of mother at the time of birth (years)	27.4 (6.6)	25.8 (6.7)	N/A	28.0 (6.4)	N/A	25.8 (6.3)	N/A
Age of father at the time of birth (years)	30.6 (7.3)	30.3 (8.5)	N/A	30.4 (7.2)	N/A	29.6 (8.0)	N/A
Father's Occupation[b]	3.5 (1.8)	4.2 (1.9)	3.9 (1.9)	3.4 (1.9)	3.5 (1.8)	4.1 (1.9)	3.8 (1.9)
Mother's Occupation[b]	3.8 (1.7)	4.3 (1.9)	4.6 (1.9)	3.8 (1.8)	4.0 (1.8)	4.2 (1.8)	4.2 (1.9)
Father Educational Scale[b]	3.0 (1.5)	3.4 (1.6)	3.5 (1.6)	3.0 (1.6)	3.2 (1.6)	3.5 (1.7)	3.5 (1.6)
Mother Educational Scale[b]	3.2 (1.5)	3.4 (1.5)	3.7 (1.5)	3.0 (1.4)	3.3 (1.4)	3.5 (1.5)	3.5 (1.5)
Subject Educational Scale[b]	2.5 (1.0)	3.6 (1.2)	3.1 (1.2)	2.8 (1.0)	2.8 (1.3)	3.4 (1.1)	3.1 (1.4)
Emotional Abuse[a]	7.5 (3.6)	11.4 (5.5)	N/A	12.6 (5.8)	N/A	13.6 (6.1)	N/A
Emotional Neglect[a]	8.5 (4.1)	12.3 (5.3)	N/A	12.1 (5.2)	N/A	13.4 (5.3)	N/A
Physical Abuse[a]	6.8 (2.8)	9.2 (4.7)	N/A	8.6 (4.8)	N/A	10.6 (5.6)	N/A
Physical Neglect[a]	6.4 (2.1)	9.2 (4.0)	N/A	8.5 (4.1)	N/A	9.2 (4.0)	N/A
Sexual Abuse[a]	6.2 (3.4)	8.4 (5.5)	N/A	8.9 (6.6)	N/A	11.1 (7.4)	N/A
Age at first use of cannabis (years)	18.5 (4.3)	16.8 (4.3)	N/A	17.6 (4.8)	N/A	16.6 (5.1)	N/A
Use of cannabis after age 18[c]	3.4 (1.6)	4.0 (1.8)	N/A	4.2 (1.6)	N/A	4.0 (1.9)	N/A
Use of cannabis between age 15 and 18[c]	2.1 (1.6)	3.1 (2.1)	N/A	2.8 (2.0)	N/A	3.0 (2.0)	N/A
Use of cannabis before age 15[c]	1.2 (0.8)	1.8 (1.5)	N/A	1.7 (1.4)	N/A	1.9 (1.6)	N/A
Use of other substances before age 18[c]	0.2 (0.3)	0.4 (0.6)	N/A	0.7 (0.8)	N/A	0.5 (0.7)	N/A
Use of other substances after age 18[c]	0.4 (0.7)	0.8 (1.0)	N/A	1.3 (1.2)	N/A	1.1 (1.1)	N/A
Word Reading Score[d]	103.1 (34.1)	93.0 (15.7)	99.8 (53.4)	101.7 (14.3)	102.9 (14.0)	94.3 (14.4)	98.5 (16.0)

Table 1. Participant details

[a] Abuse and neglect scores are from the childhood trauma questionnaire[31]

[b] The Hollingshead Socioeconomic Rating Scale is used to classify educational achievement ranging from 1 (completed graduate degree) to 7 (less than seven years of schooling); and parental occupation ranging from 1(major professional) to 8 (unskilled labouring)[32]

[c] The cannabis/substance use scores were derived from a standardised clinical interview and range from 1(no use) to 6 (over 100 times). The 'other substances' values represent the mean across stimulants, cocaine, opiates, hallucinogens and other substances.

[d] Word reading score is the WRAT-4 standardised score

[e] Other category includes 'Unknown or not reported', 'Hawaiian or Pacific Islander', 'American Indian/Alaska Native'

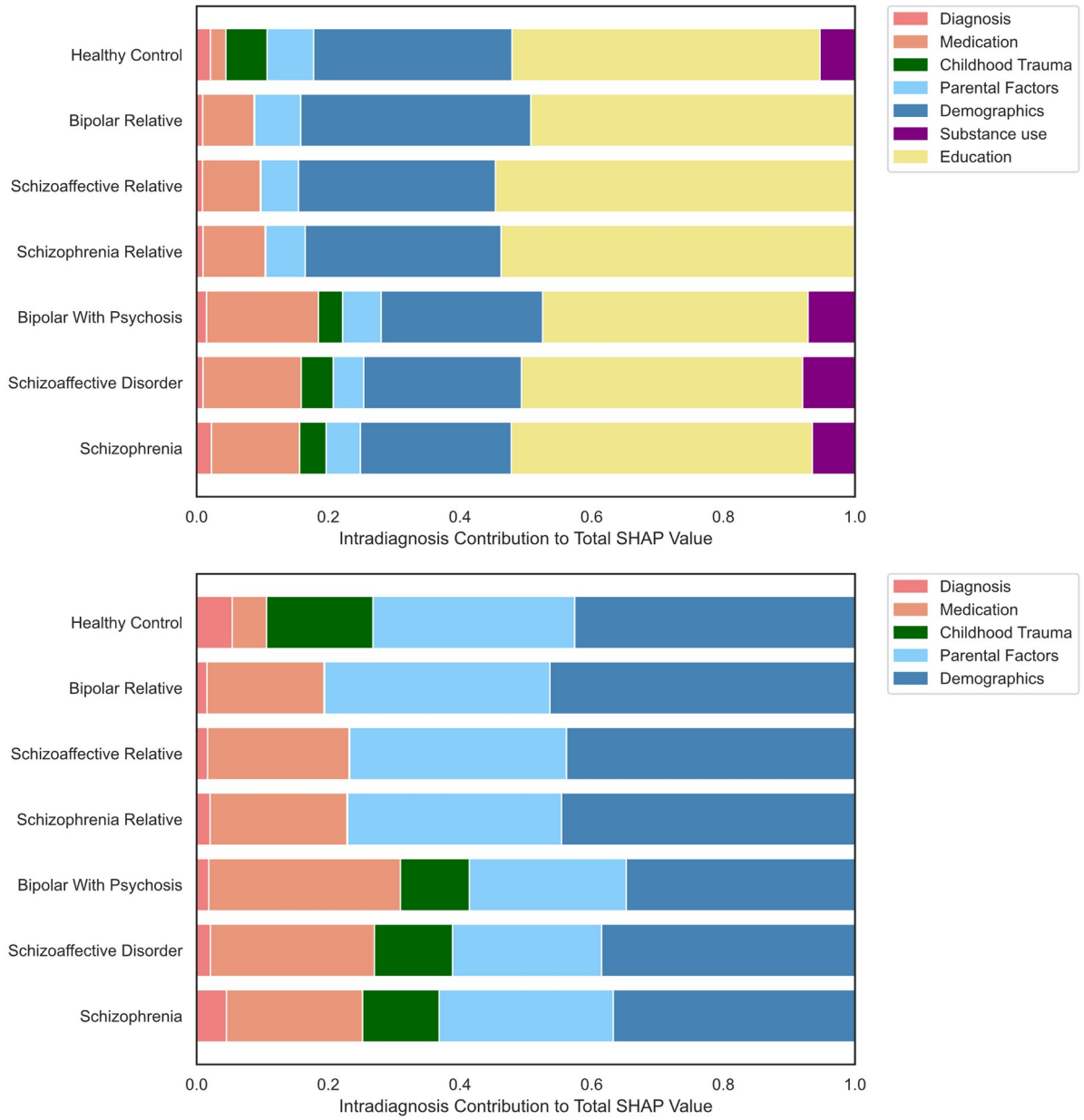

Figure 1 The association of predictor variables and within diagnosis cognitive performance

The width of the bars represents the relative contribution of the variable to the total SHAP value within the diagnostic group (the education bar includes word reading). E.g. The relatively wide bar representing education indicates that education makes a major contribution to the prediction for cognitive performance, within diagnosis. Variables that are relatively constant within a group (e.g. diagnosis) will contribute minimally here although may contribute markedly to between group predictions. The chart represents the contribution of each variable to model predictions i.e. it does not imply that these variables explain 100% of variance in cognitive scores. (A) The chart based on a model including all predictor variables (B) The chart based on a model without including participant substance use, education level, or word reading.

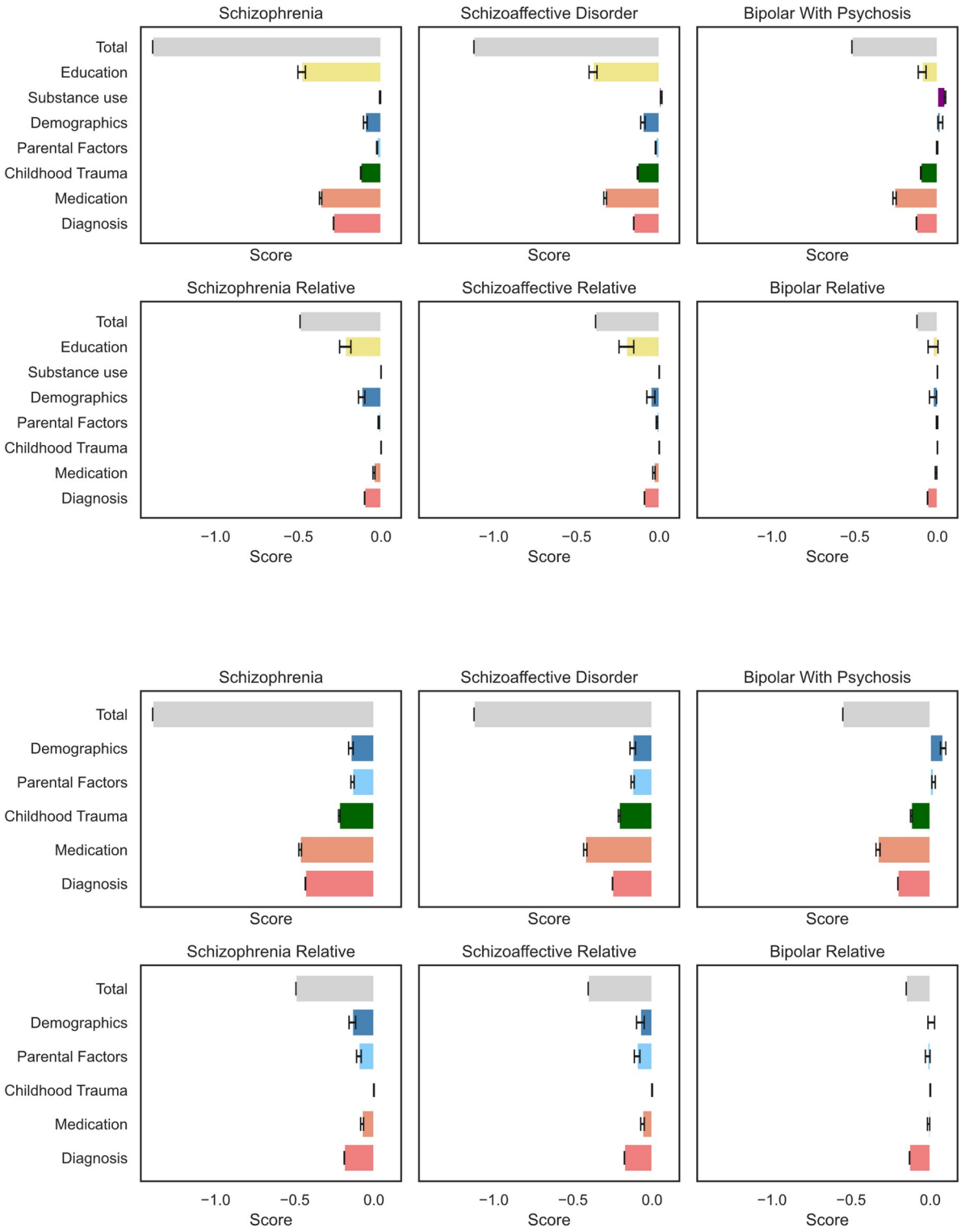

Figure 2 The association of predictor variables and cognitive impairment

The bar width for each variable, for each diagnostic group, represents the extent in terms of z-score to which that variable contributes to cognitive impairment relative to the control cohort. Bars to the right of '0' indicate that in this cohort the variable has a net protective effect. (A) The chart based on a model including all predictor variables (B) The chart based on a model without including participant substance use, education level, or word reading.

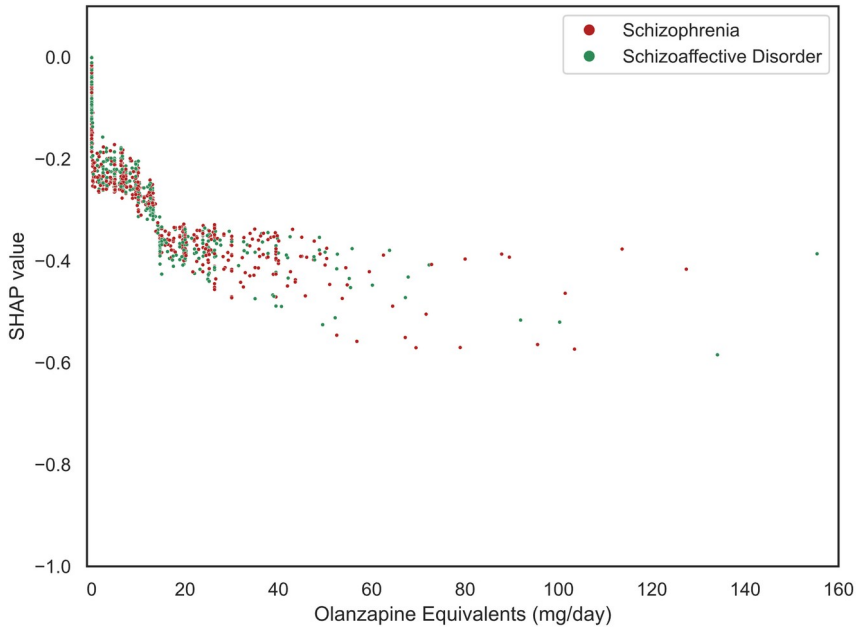

Figure 3 Association between antipsychotic exposure and prediction of cognitive performance

SHAP values associated with medication exposure from models fit to a subsample of individuals with schizophrenia or schizoaffective disorder. Increasing antipsychotic dose is associated with an increasingly negative influence on prediction of cognitive performance.