# Title

Empirically validate cognitive abilities as an RDoC transdiagnostic domain for mental health across neural and genetic units of analysis: a population-based cohort study in adolescents

**Authors:**

Yue Wang, MS[1]
Richard Anney, PhD[2] https://orcid.org/0000-0002-6083-407X
Narun Pat, PhD[1] https://orcid.org/0000-0003-1459-5255

[1]Department of Psychology, University of Otago, Dunedin, New Zealand
[2]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine and Wolfson Centre for Young People's Mental Health, Cardiff University, Cardiff, UK

**Corresponding author:**

Narun Pat, PhD
Also known as Narun Pornpattananangkul, Department of Psychology, University of Otago, William James Building, 275 Leith Walk, Dunedin 9016, New Zealand
Email: narun.pat@otago.ac.nz
Phone: +64 3 470 4629

## Summary

**Background:**

A leading transdiagnostic framework for psychiatry, NIMH Research Domain Criteria (RDoC), posits that cognitive abilities are a major factor (known as functional domain) underlying mental health across various diagnoses. Specifically, RDoC assumes the relationship between cognitive abilities and mental health to be 1) manifested across neural and genetic *units of analysis*, 2) environmentally situated, and 3) reliable. These central assumptions have not been empirically validated. To address this, we used data from the Adolescent Brain Cognitive Development (ABCD) Study, which included 11,876 participants (5,680 females). The study spanned two years and focused on children aged 9-10 and 11-12.

**Methods:**

We applied machine learning to make out-of-sample predictions of cognitive abilities based on measures of mental health (emotional/behavioural problems, at-risk personalities), neuroimaging (45 types of brain MRI), polygenic scores (three definitions), and socio-demographics, lifestyles and developments (44 variables e.g., parental income, screen use). Using linear-mixed-model commonality analyses, we then examined the extent to which the relationship between cognitive abilities and mental health was explained by the other measures.

**Findings:**

Mental health predicted cognitive abilities of unseen children (i.e., those who were not part of the modelling process) at $r$=.39. At baseline, this cognitive-abilities-mental-health relationship was accounted for by neuroimaging (69%), by polygenic scores (18%) and by socio-demographics, lifestyles and developments (70%). Moreover, the variance in the relationship between cognitive abilities and mental health that was captured by socio-demographics, lifestyles and developments was explained by neuroimaging (66%) and polygenic scores (22%). These patterns were consistent across the two-time points.

**Interpretation:**

Consistent with RDoC, the cognitive abilities and mental health relationship was 1) manifested in both neuroimaging and polygenic scores, 2) explained by socio-demographics, lifestyles and developments and 3) reliable across two years. This supports RDoC's view of cognitive abilities as an integrative-functional domain for the aetiology of mental health.



**Funding:**

Health Research Council Funding of New Zealand (21/618) and University of Otago

# Research in Context

### Evidence before this study

It has been over a decade since the initiation of Research Domain Criteria (RDoC) by the US National Institute of Mental Health. The first article outlining its framework (Insel et al., 2010, *AJP*) has been cited over 6,650 times according to Google Scholar as of September 25th, 2023. In one estimate (Pacheco et al., 2022, *JCPP*), 644 RDoC-themed grants were funded between 2008-19 in the US. RDoC has also asserted its impact on psychiatric research outside of the US, as seen in frameworks such as the Psychiatric Ratings using Intermediate Stratified Markers (PRISM) funded by EU-Innovative Medicine Initiative (Kas et al., 2019, *Neuroscience and Biobehavioral Reviews*). Accordingly, RDoC has had a great influence on the landscape of modern psychiatric research. Yet, most RDoC-themed studies have focused on applying the framework (e.g., to improve the aetiology of certain mental illnesses), but few, if any, have tested its central assumptions regarding the properties of its functional domains. These assumptions were largely expert-drivenNY. Here, we focused on empirically validating the properties of one of the six major functional domains, cognitive systems (note we called them cognitive abilities here to map closer to a term used in cognitive literature).

### Added value of this study

During the time when RDoC was first developed, there was no large-scale neuroimaging and genetic data appropriate to validate the framework. The emergence of studies such as the Adolescent Brain Cognitive Development (ABCD) offers not only neuroimaging and genetic data but also high-quality data on phenotypes (e.g., mental health and cognitive abilities) along with socio-demographics and lifestyles which allows us to validate the RDoC framework. Using machine learning, we were able to quantify the extent to which the relationship between cognitive abilities and mental health was 1) manifested in both neuroimaging and polygenic scores, 2) explained by socio-demographics, lifestyles, and developments and 3) reliable across two years. Accordingly, our study validated the role of cognitive abilities as a functional domain according to RDoC.

### Implications of all of the available evidence

We provide empirical evidence to support RDoC's view of cognitive abilities as a transdiagnostic domain for mental health. By doing so, we laid a foundation to achieve the four goals set by RdoC (Pacheco et al., 2022, *JCPP*). First, we have addressed the aetiology of mental health by showing the reliable and predictable transdiagnostic relationship between cognitive abilities and mental health. Second, we also identified neuroimaging and genetic biomarkers that explain this relationship. Third, with our findings on the socio-demographics, lifestyles and developments, we touched on target identification. Finally, our use of machine learning to make out-of-sample predictions on individual children could lead to further development of personalised interventions. Overall, our findings lend credibility to RDoC research focusing on cognitive abilities.

# Introduction

Over a decade ago, the National Institute of Mental Health launched Research Domain Criteria (RDoC) in an effort to improve mental-health aetiology[1]. Conventional diagnostic criteria, such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) and International Classification of Diseases (ICD), classify mental disorders based on consensus among clinician experts who agree on which observable symptoms belong to which disorder[2]. Accordingly, often these diagnostic criteria do not align well with research in neuroscience and genetics and do not integrate the roles of environments, such as socio-demographics and lifestyles[3]. Instead of focusing on specific mental disorders, RDoC focuses on the transdiagnostic relationship between mental health and functional domains, such as cognitive abilities/systems, and more importantly, the extent to which neurobiological and social-environmental factors capture this relationship[4].

Yet, despite being a prominent view with thousands of studies adopting the framework[3], RDoC currently faces the same problem as DSM/ICD – it relies heavily on expert consensus[5]. For example, RDoC's constructs for cognitive abilities were selected based on meetings among experts on cognitive neuroscience and schizophrenia[4,6]. Accordingly, after a decade of work, RDoC is still driven by experts' opinions and in need of empirical evidence to support its framework. Therefore, we aim to empirically investigate if cognitive abilities have the properties of a functional domain as proposed by RDoC.

RDoC stipulates the relationship between cognitive abilities and mental health to be manifested across neurobiological *units of analysis*, from the brain to genes[7]. Recent advances in large-scale data and machine learning[8] allow researchers to build predictive models to estimate cognitive abilities of unseen participants, not part of the modelling process, based on their neuroimaging data (e.g., via different types of brain MRI[9,10]) and genetic risk profiles (e.g., via polygenetic scores[9,11]). Accordingly, these predictive models create neurobiological proxy measures of cognitive abilities, close to what RDoC considers as units of analysis. Thus, RDoC would predict these neurobiological proxy measures to account for the relationship between cognitive abilities and mental health.

RDoC, additionally, assumes the relationship between cognitive abilities and mental health to be affected by environments (https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/developmental-and-environmental-aspects). It is well-accepted that environmental factors play a large role on cognitive abilities[12]. Recent research applied machine learning to compute proxy measures of cognitive abilities in children based on their socio-demographics (e.g., parental income/education), lifestyles (e.g., screen/video game use) along with child developmental adversary (e.g., birth/pregnancy complication)[9,13]. Similar to neurobiological proxy measures, RDoC would predict these proxy measures of cognitive abilities based on socio-demographics, lifestyles and developments to account for the relationship between cognitive abilities and mental health. Moreover, RDoC would further predict these variances explained by socio-demographics, lifestyles and developments to be overlapped with those from the neurobiological proxy measures. Such patterns would provide an integrative, holistic account of the relationship between cognitive abilities and mental health, as claimed by RDoC.

RDoC also emphasises the reliability of measurements[4,14]. The Adolescent Brain Cognition Development (ABCD)[15], an on-going, population-based cohort study, has already provided cognitive, neuroimaging, genetic and social-environmental data from over 11,000 adolescents at ages 9-10 and 11-12 years old. Therefore, the ABCD is suitable for testing whether the cognitive abilities and mental health relationship is reliably manifested in both the neuroimaging and polygenic scores and affected by socio-demographics, across time points.

Here using the ABCD, we operationalised cognitive abilities as a latent variable of behavioural performance during cognitive tasks, covering most of RDoC's cognitive-ability/systems constructs (apart from perception), including attention, working memory, declarative memory, language and cognitive control[7]. We then built predictive, machine-learning models to predict the cognitive abilities of unseen children from their mental health. These predictive models allowed us to quantify the relationship between cognitive abilities and mental health. Following RDoC's holistic view of mental health, covering symptoms, risk factors and family environments[1,4], we included not only children's but also their caretakers' emotional and behavioural problems[16] as well as children's at-risk personalities, including behavioural inhibition/activation[17] and impulsivity[18].

Similar to mental health, we also built predictive models of cognitive abilities based on neuroimaging, polygenic scores, socio-demographics, lifestyles and developments to create different proxy measures of cognitive abilities. For the neuroimaging, we included 45 types of brain MRI, derived from task/resting-state fMRI, structural MRI and diffusion tensor imaging (DTI). For polygenic scores, we used three definitions of cognitive abilities based on previous meta/mega analyses[19–21]. For socio-demographics, lifestyles and developments, we included 44 variables, covering factors, such as parental income/education, screen use and birth/pregnancy complications. Finally, we conducted a series of commonality

analyses[22] to demonstrate the extent to which proxy measures of cognitive abilities based on neuroimaging, polygenic scores and socio-demographics, lifestyles and developments explained the relationship between cognitive abilities and mental health. To examine the reliability, we repeated the analyses across two time points and examined the similarity between the two.

## Methods

### The Adolescent Brain Cognitive Development (ABCD) Study

We used data from Adolescent Brain Cognitive Development (ABCD) Study Curated Annual Release 4.0[23,24], recruited across 21 sites in the US from two time points: the baseline (11,876 children, 5,680 females, 9-10 years old) and the two-year follow-up (11,225 children, 5,353 females), starting on September 1st, 2016. For exclusion criteria, please see Supplementary Methods. The study protocols have been approved by Institutional Review Boards at each site. For ethical details, such as informed consent, confidentiality and communication with participants about assessment results, please see Clark and colleagues[25].

### Measures of Cognitive Abilities

We operationalised individual differences in cognitive abilities across six cognitive tasks as a factor score of a second-order latent variable 'g-factor'. These six cognitive tasks were collected with an iPad in a 70-min session outside of MRI, covering language comprehension and decoding, conflict monitoring and inhibitory control, processing speed, episodic memory and memory recall. For the details about these tasks and our factor analysis, please see Supplementary Methods.

### Measures of Mental Health

Mental health was based on three sets of features: 1) children's emotional and behavioural problems, reflected by Child Behaviour Checklist (CBCL)[16], 2) caretakers' emotional and behavioural problems, reflected by Aseba Adult Self Report (ASR)[16] and 3) children's at-risk personality traits[17,26], reflected by the Urgency, Premeditation, Perseverance, Sensation seeking and Positive urgency (UPPS-P) Impulsive Behaviour Scale[27] and the Behavioural Inhibition System/Behavioural Activation System (BIS/BAS)[28]. For the details about mental health variables, please see Supplementary Methods. For the summary statistics and distributions of these measures, please see Supplementary Tables 1-2.

### Measures of Neuroimaging

Brain-MRI data was organized into 45 sets of neuroimaging features, covering task-fMRI (including contrasts from ENBack, stop signal (SST), and monetary incentive delay (MID) tasks), resting-state fMRI (rs-fMRI), structural MRI (sMRI) and diffusion tensor imaging (DTI). For details about brain-MRI quality control, harmonisation, data organisation, please see Supplementary Methods.

### Measures of Genetic Risk Profiles

Polygenic scores (PGS) were calculated from three large-scale genome-wide association studies (GWAS) on cognitive abilities in adults: n=257,841[20], n=300,486[19] and n=269,867[21]. Because PGS has a lower predictive ability when target samples (i.e., in our case, ABCD children) do not have the same ancestry as those of the discovery GWAS sample[29], we restricted all analyses involving PGS to 5,786 children of European ancestry. For details about PGS quality control and quantification, please see Supplementary Methods.

### Measures of Socio-demographics, lifestyles and developments

Socio-demographics, lifestyles and developments were based on 44 features: a) 14 for child social-demographics (e.g., bilingual use, parental marital status, parental education, parental income, household size, economic insecurities, area deprivation index, lead risk, crime report, neighbourhood safety, school environment, involvement and disengagement), b) five for child social interactions, c) six for parental use of alcohol, tobacco and marijuana before and after pregnancy, d) eight for children's sleep problems, e) four for child's physical activities, f) four for child screen use, and g) three for child developmental adversary (e.g., prematurity, birth complications and pregnancy complications). For references of these measures, please see Supplementary Methods. For the summary statistics and distributions of these measures, please see Supplementary Tables 3-4.

### Predictive modelling

For building predictive, machine-learning models, we implemented a nested leave-one-site-out cross-validation. Specifically, we treated one out of 21 recruitment sites as a test set and the rest as a training set for training predictive

models. We then repeated the model-building process until every recruitment site was a test set once and reported overall predictive performance across all test sites. For predicting cognitive abilities from mental health and socio-demographics, lifestyles and developments, we implemented Partial Least Squares (PLS)[30]. PLS decomposes features into components that not only capture the variance of features but also the variance of the target[30]. PLS has an advantage in dealing with collinear features[31], while maintaining interpretability. For predicting each set of neuroimaging features and polygenic scores, we used Elastic Net[32]. Elastic Net performs relatively well in predicting behaviours from brain MRI, often on par with, if not better than, other more complicated algorithms, while still allowing readily explainable cofficients[33]. To integrate 45 sets of neuroimaging features, we implemented opportunistic stacking[9,34] that combined predicted values from each set of neuroimaging features along with the imputed values of any missing values with Random Forest[35]. Opportunistic stacking allowed us to handle missingness in the neuroimaging data without sacrificing predictive performance[9]. For details about the model-building processes and data-leakage prevention, see Supplementary Methods.

### Commonality analyses

Following the predictive modelling procedure above, we extracted predicted values from different sets of features at each test site and treated them as proxy measures of cognitive abilities[8]. The out-of-sample relationship between observed and proxy measures of cognitive abilities based on certain features reflects variation in cognitive abilities explained by those features. For instance, the relationship between observed and proxy measures of cognitive abilities based on mental health indicates the variation in cognitive abilities that could be explained by mental health. Capitalising on this variation, we, then, used commonality analyses[22] to demonstrate the extent to which other proxy measures captured similar variance of cognitive abilities as mental health.

We applied random-intercept, linear-mixed models[36] to the data from all test sites. In these models, we considered families to be nested within each site, meaning that we allowed each family to have a different intercept within each site and each site to have a different intercept. We treated different proxy measures of cognitive abilities as fixed-effect regressors to explain cognitive abilities. We, then, estimated marginal $R^2$ from the linear-mixed models, which describes the variance explained by all fixed-effects included in the models[37] and multiplied the marginal $R^2$ by 100 to obtain a percentage. By including and excluding each proxy measure in the models, we were able to decompose marginal $R^2$ into unique (i.e., attributed to the variance, uniquely explained by a particular proxy measure) and common (i.e., attributed to the variance, jointly explained by a group of proxy measures) effects[22]. We focused on the common effects a proxy measure based on mental health had with other proxy measures in four sets of commonality analyses, i.e., with 1) neuroimaging, 2) polygenic scores 3) socio-demographics, lifestyles and developments or 4) the combination. Note each of the four sets of commonality analyses used different numbers of participants, depending on the availability of data. For details about the commonality analyses, see Supplementary Methods.

### Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## Results

### Predictive modelling

Predicting cognitive abilities from mental health

Figure 1a and Supplementary Table 5 show the predictive performance of the Partial Least Square (PLS) models, predicting cognitive abilities from mental-health features. These features included 1) children's (CBCL) and 2) caretakers' (ASR) emotional and behavioural problems and 3) children's at-risk personalities. Using these three sets of mental health features separately led to moderate predictive performance, ranging from $r$=.21 to $r$=.31. Combining them into one set of features, "mental health," boosted the performance to around $r$=.4, similarly across the two time points.

Figure 1b shows loadings and the ratio of variance explained by these PLS models. The first PLS component had the highest ratio of variance explained (19.3%-24.3%). Based on its loadings, this component was driven by features such as attention and social problems, rule-breaking and aggressive behaviours, Behavioural-Activation-System drive and the caretaker's personal strength. We found a similar pattern across the two-time points.

Predicting cognitive abilities from neuroimaging

Figures 2a and Supplementary Figures 2-3 and Supplementary Tables 5-7 show the predictive performance of the opportunistic stacking models, predicting cognitive abilities from 45 sets of neuroimaging features. Predictive

performance from each set of neuroimaging features varied considerably, ranging from *r* around 0 (ENBack: Positive vs Neutral Face) to *r* around .4 (ENBack: 2-Back vs 0-Back). Drawing information across 45 sets of neuroimaging features into a stacked model boosted the performance to around *r*=.54, similarly across the two time points. In fact, the stacked model ($R^2$ around .3) accounted for variance in cognitive abilities twice as much as the model based on the best set of neuroimaging features (ENBack: 2-Back vs 0-Back, $R^2$ around .15). Figures 2b,3 and Supplementary Figure 4 show the feature importance of the opportunistic stacking models. Across the two time points, the top contributing sets of neuroimaging features, reflected by SHAP, were ENBack task-fMRI contrasts, rs-fMRI and cortical thickness.

Predicting cognitive abilities from polygenic scores

Figures 2a and Supplementary Tables 5 show the predictive performance of the Elastic Net models, predicting cognitive abilities from three polygenic scores (PGSs) of cognitive abilities. Predictive performance from polygenic scores was *r*=.25 and .24 in baseline and follow-up, respectively. Figures 2c shows the feature importance of these models, suggesting a stronger contribution from Savage and colleagues' PGS[21].

Predicting cognitive abilities from socio-demographics, lifestyles and developments

Figure 4a and Supplementary Table 5 show the predictive performance of the Partial Least Square (PLS) models, predicting cognitive abilities from socio-demographics, lifestyles and developments. Using 44 features covering socio-demographics, lifestyles and developments, we saw predictive performance around *r*=.48, similarly across the two time points. Figure 4b shows loadings and the ratio of variance explained of these PLS models. The first PLS component had the highest ratio of variance explained (around 10%).

Based on its loadings, this first component was a) positively driven by features such as parent's income and education, neighbourhood safety and extracurricular activity and b) negatively driven by features such as area deprivation, having a single parent, screen use, economic insecurities, lack of sleep, mature video games play/movies watch and lead risk.

**Commonality analyses**

Commonality analyses for proxy measures of cognitive abilities based on mental health and neuroimaging

At baseline, having both proxy measures based on mental health and neuroimaging in a linear mixed model explained 33% of the variance in cognitive abilities. And 14.13% of the variance in cognitive abilities was explained by mental health, which included the common effect between the two proxy measures (9.82%) and the unique effect of mental health (4.31%) (see Supplementary Tables 8-9 and Figure 5). This means that neuroimaging accounted for over 69% of the relationship between cognitive abilities and mental health (i.e., (9.8÷14.13)x100). The common effects varied considerably across sets of neuroimaging features, from around 0 to 4.42%, with the highest set being ENBack task fMRI: 2-Back vs 0-Back (see Supplementary Figure 5). The pattern of results was similar across two time points.

Commonality analyses for proxy measures of cognitive abilities based on mental health and polygenic scores

At baseline, having both proxy measures based on mental health and polygenic scores in a linear mixed model explained 11.8% of the variance in cognitive abilities. And 11.4% of the variance in cognitive abilities was explained by mental health, which included the common effect between the two proxy measures (2.06%) and the unique effect of mental health (9.36%) (see Supplementary Tables 10-11 and Figure 5). This means that polygenic scores accounted for over 18% of the relationship between cognitive abilities and mental health (i.e., (2.06÷11.4)x100). The pattern of results was similar across two time points.

Commonality analyses for proxy measures of cognitive abilities based on mental health and socio-demographics, lifestyles and developments

At baseline, having both proxy measures based on mental health and socio-demographics, lifestyles and developments in a linear mixed model explained 27.1% of the variance in cognitive abilities. And 14.13% of the variance in cognitive abilities was explained by mental health, which included the common effect between the two proxy measures (9.92%) and the unique effect of mental health (4.21%) (see Supplementary Tables 12-13 and Figure 5). This means that socio-demographics, lifestyles and developments accounted for over 70% of the relationship between cognitive abilities and mental health (i.e., (9.92÷14.13)x100). The pattern of results was similar across two time points.

Commonality analyses for proxy measures of cognitive abilities based on mental health, neuroimaging, polygenic scores and socio-demographics, lifestyles and developments

At baseline, having all four proxy measures based on mental health, neuroimaging, polygenic scores and socio-demographics, lifestyles and developments in a linear mixed model explained 28.9% of the variance in cognitive abilities.

And 1.59% of the variance in cognitive abilities was the common effect among the four proxy measures. Additionally, out of 11.4% of the variance in cognitive abilities explained by mental health, 9.25% could be explained by any of the other proxy measures. That means the three proxy measures accounted for 81% (i.e., (9.25÷11.4)x100) of the relationship between cognitive abilities and mental health (see Supplementary Tables 14-15 and Figure 5). This also means that, among the variance that socio-demographics, lifestyles and developments accounted for in the relationship between cognitive abilities and mental health, neuroimaging could capture 66% while polygenic scores could capture 28%. The pattern of results was similar across two time points.

## Discussions

Using large-scale data in adolescents collected two years apart, we empirically tested whether cognitive abilities have the properties of a functional domain as proposed by RDoC[7]. We started by quantifying the relationship between cognitive abilities and mental health and found a medium-sized out-of-sample *r* around .4. Consistent with RDoC[4,14], first, this relationship was explained by the neurobiological proxy measures of cognitive abilities based on neuroimaging and polygenic scores (69% and 18%, respectively, at baseline). Second, this relationship between cognitive abilities and mental health was also explained by proxy measures of cognitive abilities based on socio-demographics, lifestyles and developments (70% at baseline). The variance due to socio-demographics, lifestyles and developments was largely accounted for by the proxy measures based on neuroimaging and polygenic scores (66% and 28%, respectively at baseline). Third, this pattern of results was reliable across the two-time points. Accordingly, overall, our results support RDoC's view of cognitive abilities as an integrative, functional domain for mental health.

Our predictive modelling revealed a medium-sized predictive relationship between cognitive abilities and mental health. This aligns with recent meta-analyses linking cognitive abilities and mental disorders[38]. We also found that including, not just children's, but also caretakers' emotional and behavioural problems[16] and children's at-risk personalities[27,39] as mental health features improved the strength of the relationship. This solidifies RDoC's holistic view of mental health as a broad concept, spanning from pathologies to risk factors and incorporating individual and family environments[1,4]. Examining the PLS loading of our predictive models revealed that the relationship was driven by different aspects of mental health, from thought and externalising symptoms to motivation and caretakers' personal strength. Altogether, we confirmed the transdiagnostic nature of cognitive abilities.

Our predictive modelling also created proxy measures of cognitive abilities for two RDoC *units of analysis*[7]: at the neural, reflected by neuroimaging, and genetic, reflected by polygenic scores, levels. For neuroimaging, combining 45 sets of neuroimaging features via opportunistic stacking led to a relatively high effect size (out-of-sample $r$=.54 at baseline), as compared to using any single set, consistent with previous research[33,34,40]. The superior performance from ENBack: 2-Back vs 0-Back was also consistent with work done in adults from the Human Connectome Project[33,41]. For the polygenic scores, using polygenic scores led to a weaker effect size (out-of-sample $r$=.25 at baseline). However, this predictive strength is still in the ballpark of previous work in children[11]. More importantly, these neurobiological proxy measures of cognitive abilities were able to capture the majority of the relationship between cognitive abilities and mental health, consistent with RDoC.

Similarly, our predictive modelling created proxy measures of cognitive abilities from socio-demographics, lifestyles and developments. In line with previous work[9,13], we could predict unseen children's cognitive abilities based on their socio-demographics, lifestyles and developments at a medium-to-high out-of-sample $r$=.48 (at baseline). This prediction was driven more strongly by socio-demographics (e.g., parent's income and education, neighbourhood safety, area deprivation, single parenting), somewhat weaker by lifestyles (e.g., extracurricular activities, sleep, screen time, video gaming, mature movie watching and parental monitoring) and much weaker by developments (e.g., pregnancy complications). More relevant to RDoC, proxy measures based on socio-demographics, lifestyles and developments captured a large proportion of the relationship between cognitive abilities and mental health. Furthermore, as predicted by RDoC, this variance captured by socio-demographics, lifestyles and developments overlapped largely with that of the neurobiological proxy measures. This reiterates RDoC's main tenet that understanding the neurobiology of a functional domain, such as cognitive abilities, could lead us to understand how social environments influence mental health[1,4]. More importantly, all of the aforementioned results regarding neuroimaging, polygenic scores, socio-demographics, lifestyles and developments were reliable across two years during a sensitive period for adolescents.

This study has some limitations, mainly due to the availability of the data provided by ABCD[15]. For mental health, even though we relied on four measures (CBCL[16], ASR[16], UPSS-P[27] and BIS/BAS[28]), we may still miss important mental-health variability, such as learning disability. Similarly, for cognitive abilities, while the six cognitive tasks[42,43] covered most of RDoC cognitive abilities/systems constructs, we still missed variability in perception[7]. For neuroimaging, while we used comprehensive multimodal MRI from 45 sets of features, three fMRI tasks included were not chosen based on RDoC cognitive abilities/systems constructs[15]. It is possible to obtain higher predictive performance based on other fMRI tasks. For genetics, we relied on polygenic scores, built to predict adults' cognitive abilities using cognitive tasks different from ABCD's[19–21], which might not translate well to children's cognitive abilities. Lastly, we also relied on 44 variables of socio-demographics, lifestyles and developments included in the study, which might miss some variables relevant to cognitive abilities (e.g., nutrition). ABCD[15] is still ongoing, and future data might be able to address some of these limitations.

For a decade, RDoC has been driven by experts' opinions[5]. At the start, empirical data that could validate this framework were scarce. The recent revolution in neurobiological big data, such as ABCD[15] and various genome-wide-study consortiums[19–21], has provided us with high-quality mental health and cognitive phenotypes, multimodal neuroimaging, genomics and social-environmental data, across time points. These findings empirically support RDoC's perspective that cognitive abilities are integral to mental health.

## Contributors

**Yue Wang**: Data curation, Formal Analysis, Methodology, Software, Visualization, Writing – original draft, **Richard Anney:** Data curation, Software, Writing – review & editing, **Narun Pat:** Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing

## Declaration of Interests

The authors declare that there is no conflict of interest.

## Data Sharing

We used publicly available ABCD 4.0 data provided by the ABCD study (https://abcdstudy.org), held in the NIMH Data Archive (https://nda.nih.gov/abcd/). We uploaded the R analysis script and detailed outputs here: https://github.com/HAM-lab-Otago-University/Commonality-analysis-ABCD4.0

## Acknowledgements

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093 and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/scientists/workgroups/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. We thank the ABCD team. The authors wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research. New Zealand's national facilities are provided by NeSI and funded jointly by NeSI's collaborator institutions and through the Ministry of Business, Innovation & Employment's Research Infrastructure programme. URL https://www.nesi.org.nz. Yue Wang and Narun Pat were supported by Health Research Council Funding (21/618) and by University of Otago.

# Figures

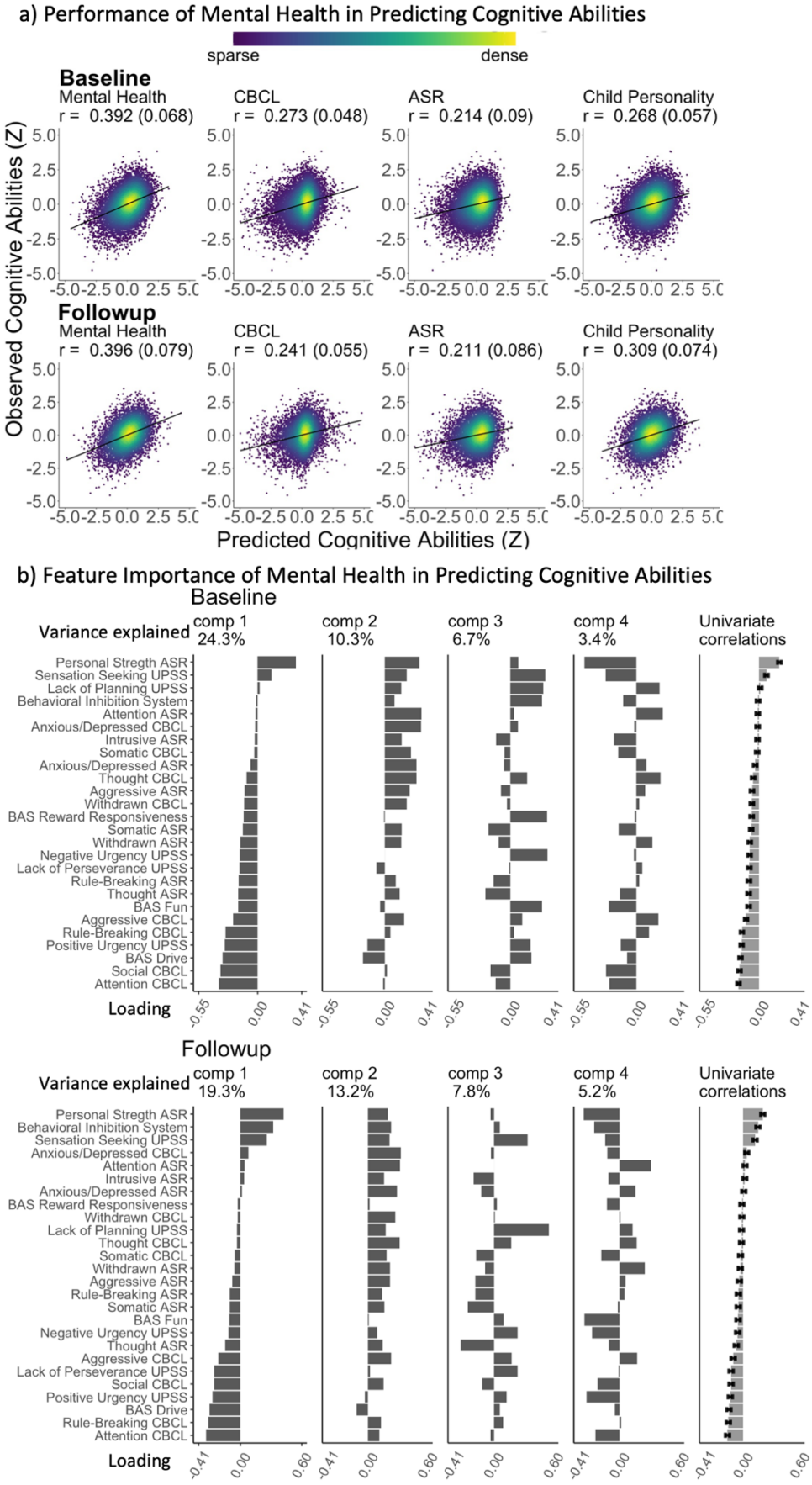

a) Performance of Mental Health in Predicting Cognitive Abilities
sparse
dense
Baseline
Mental Health
r = 0.392 (0.068)
CBCL
r = 0.273 (0.048)
ASR
r = 0.214 (0.09)
Child Personality
r = 0.268 (0.057)
Followup
Mental Health
r = 0.396 (0.079)
CBCL
r = 0.241 (0.055)
ASR
r = 0.211 (0.086)
Child Personality
r = 0.309 (0.074)
Observed Cognitive Abilities (Z)
Predicted Cognitive Abilities (Z)
b) Feature Importance of Mental Health in Predicting Cognitive Abilities
Baseline
Variance explained
comp 1 24.3%
comp 2 10.3%
comp 3 6.7%
comp 4 3.4%
Univariate correlations
Personal Stregth ASR
Sensation Seeking UPSS
Lack of Planning UPSS
Behavioral Inhibition System
Attention ASR
Anxious/Depressed CBCL
Intrusive ASR
Somatic CBCL
Anxious/Depressed ASR
Thought CBCL
Aggressive ASR
Withdrawn CBCL
BAS Reward Responsiveness
Somatic ASR
Withdrawn ASR
Negative Urgency UPSS
Lack of Perseverance UPSS
Rule-Breaking ASR
Thought ASR
BAS Fun
Aggressive CBCL
Rule-Breaking CBCL
Positive Urgency UPSS
BAS Drive
Social CBCL
Attention CBCL
Loading
Followup
Variance explained
comp 1 19.3%
comp 2 13.2%
comp 3 7.8%
comp 4 5.2%
Univariate correlations
Personal Stregth ASR
Behavioral Inhibition System
Sensation Seeking UPSS
Anxious/Depressed CBCL
Attention ASR
Intrusive ASR
Anxious/Depressed ASR
BAS Reward Responsiveness
Withdrawn CBCL
Lack of Planning UPSS
Thought CBCL
Somatic CBCL
Withdrawn ASR
Aggressive ASR
Rule-Breaking ASR
Somatic ASR
BAS Fun
Negative Urgency UPSS
Thought ASR
Aggressive CBCL
Lack of Perseverance UPSS
Social CBCL
Positive Urgency UPSS
BAS Drive
Rule-Breaking CBCL
Attention CBCL
Loading

**Figure 1. Predictive models, predicting cognitive abilities from mental-health features via Partial Least Square (PLS).** a) predictive performance of the models, indicated by scatter plots between observed vs predicted cognitive abilities based on mental health. All data points are from test sets. *r* is the average Pearson's *r* across 21 test sites, and a value in the parenthesis is the standard deviation of Pearson's *r* across sites. UPPS-P Impulsive and Behaviour Scale and the Behavioural Inhibition System/Behavioural Activation System (BIS/BAS) were used for child personality traits, conceptualised as risk factors for mental issues. Mental health includes features from CBCL, ASR and child personality. b) Feature importance of mental health, predicting cognitive abilities. The features were ordered based on the loading of the first PLS component. Univariate correlations were Pearson's *r* between each mental-health feature and cognitive abilities. Error bars reflect 95%CIs of the correlations. CBCL = Child Behavioural Checklist, reflecting children's emotional and behavioural problems; ASR = Aseba Adult Self Report, reflecting caretakers' emotional and behavioural problems; UPPS-P = Urgency, Premeditation, Perseverance, Sensation seeking and Positive urgency Impulsive Behaviour Scale; BAS = Behavioural Activation System.

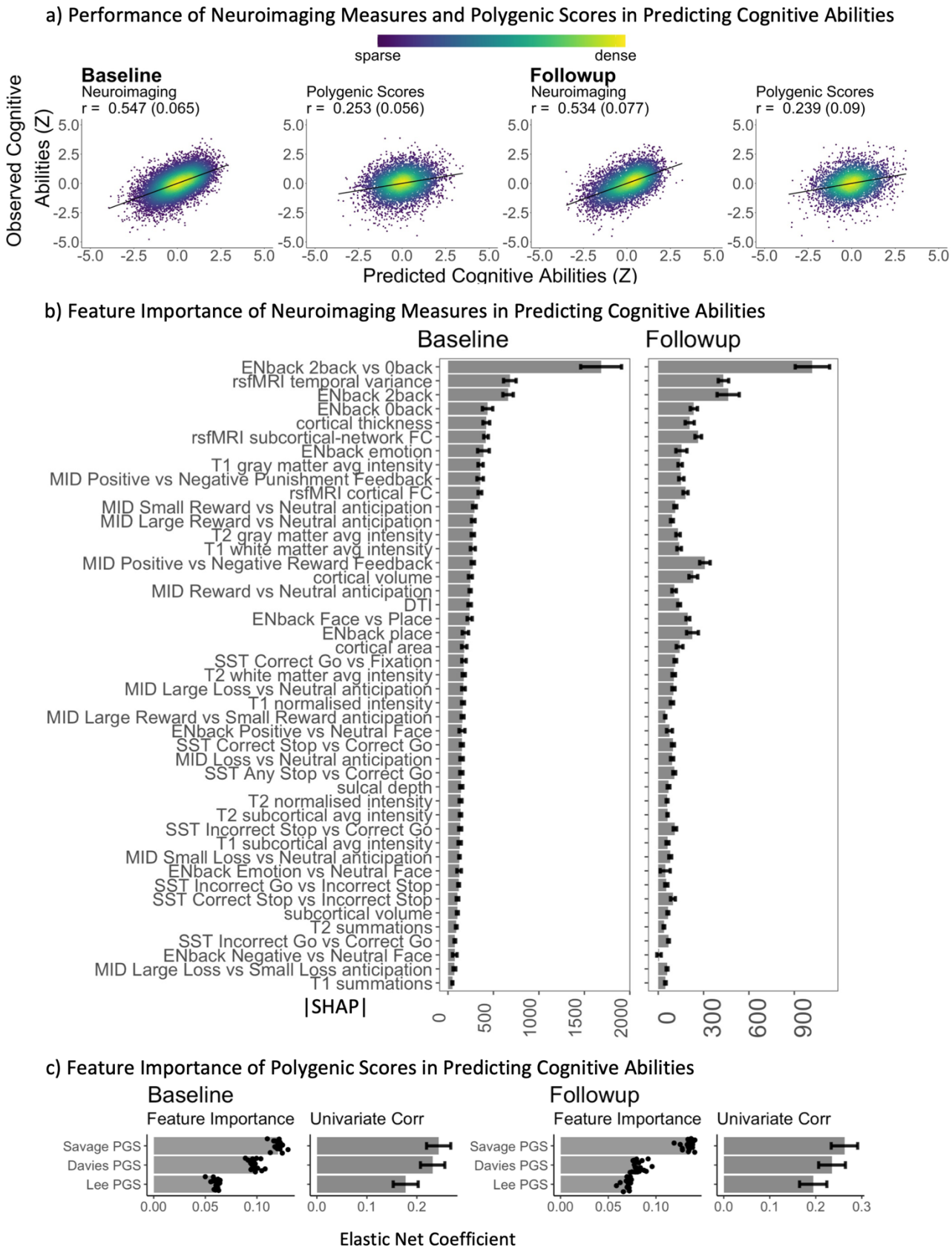



**Figure 2. Predictive models predicting cognitive abilities from neuroimaging via opportunistic stacking and polygenic scores via Elastic Net.** a) Scatter plots between observed vs predicted cognitive abilities based on neuroimaging and polygenic scores. a) All data points are from test sets. *r* is the average Pearson's *r* across 21 test sites, and a value in the parenthesis is the standard deviation of Pearson's *r* across sites. b) Feature importance of the stacking

layer of neuroimaging, predicting cognitive abilities via Random Forest. For the stacking layer of neuroimaging, the feature importance was based on the absolute value of SHAP, averaged across test sites. A higher absolute value of SHAP indicates a higher contribution to the prediction. Error bars reflect standard deviations across sites. c) Feature importance of polygenic scores, predicting cognitive abilities via Elastic Net. For polygenic scores, the feature importance was based on the Elastic Net coefficients, averaged across test sites. We also plotted Pearson's correlations between each polygenic score and cognitive abilities, computed from the full data. Error bars reflect 95%CIs of these correlations.

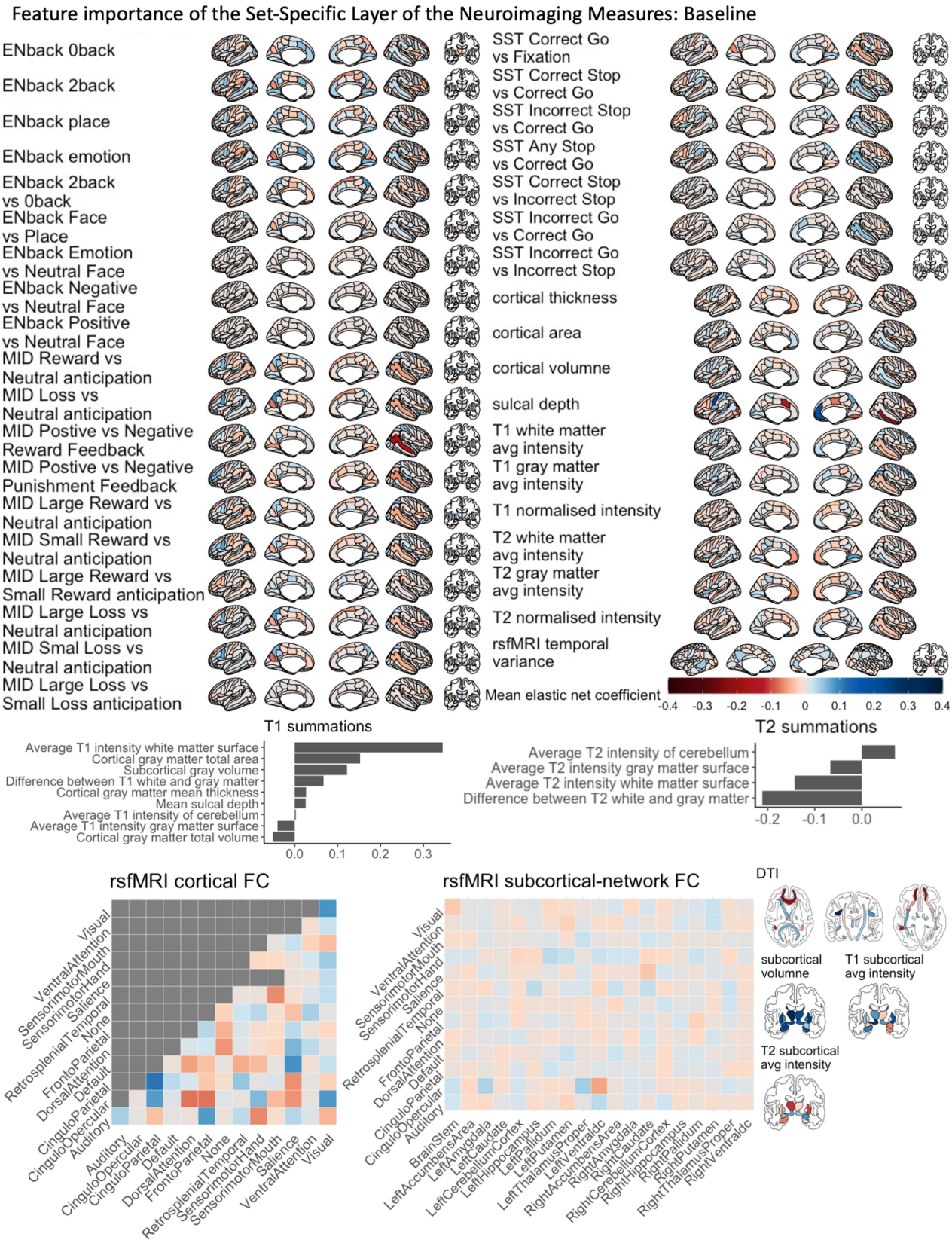
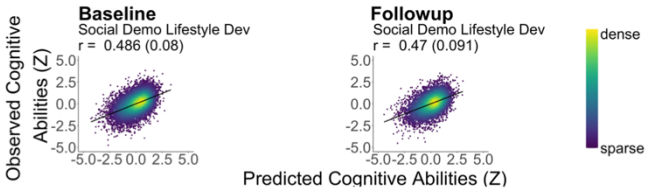
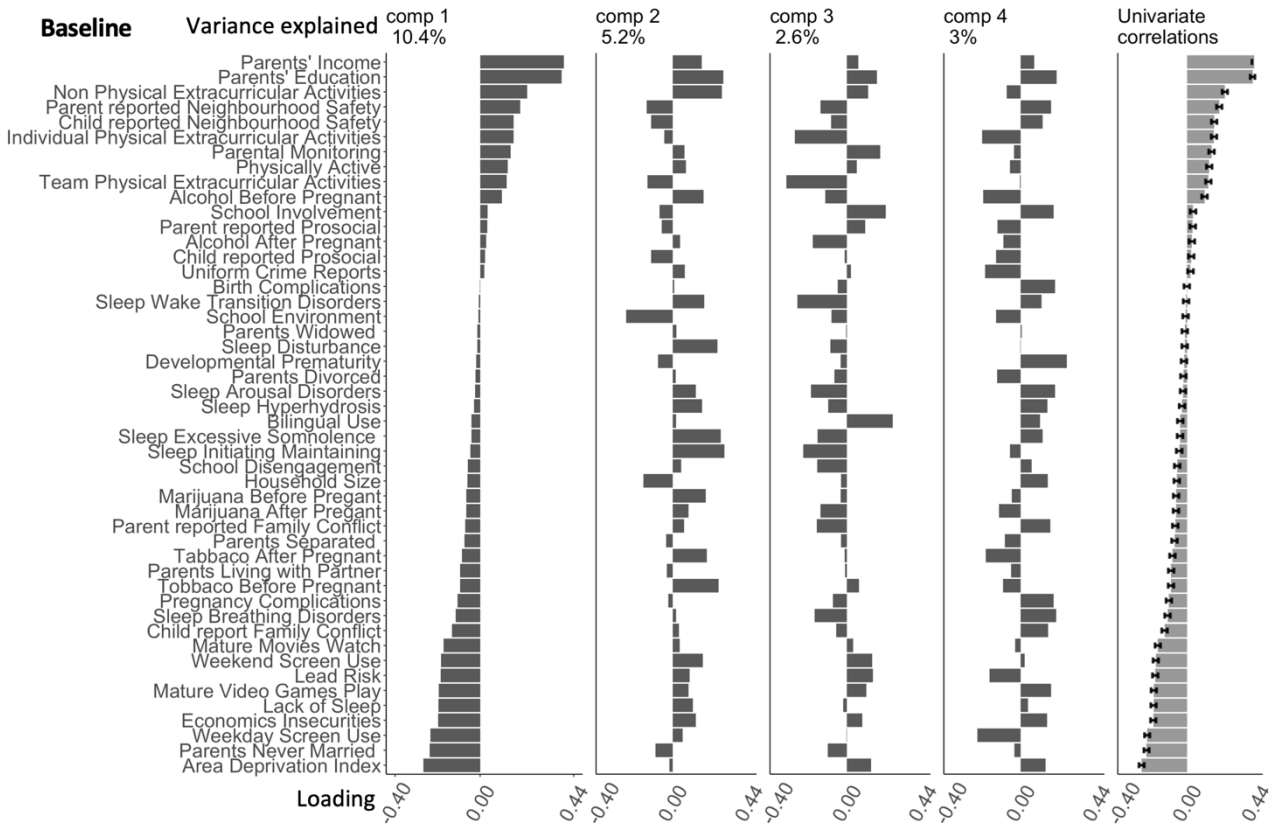



**Figure 3. Feature importance of each set of neuroimaging features, predicting cognitive abilities in the baseline data.** The feature importance was based on the Elastic Net coefficients, averaged across test sites. MID = Monetary Incentive Delay task; SST = Stop Signal Task; DTI = Diffusion Tensor Imaging; FC = functional connectivity.

a) Performance of Socio-Demographics, Lifestyles and Developments In Predicting Cognitive Abilities




b) Feature Importance of Socio-Demographics, Lifestyles and Developments In Predicting Cognitive Abilities



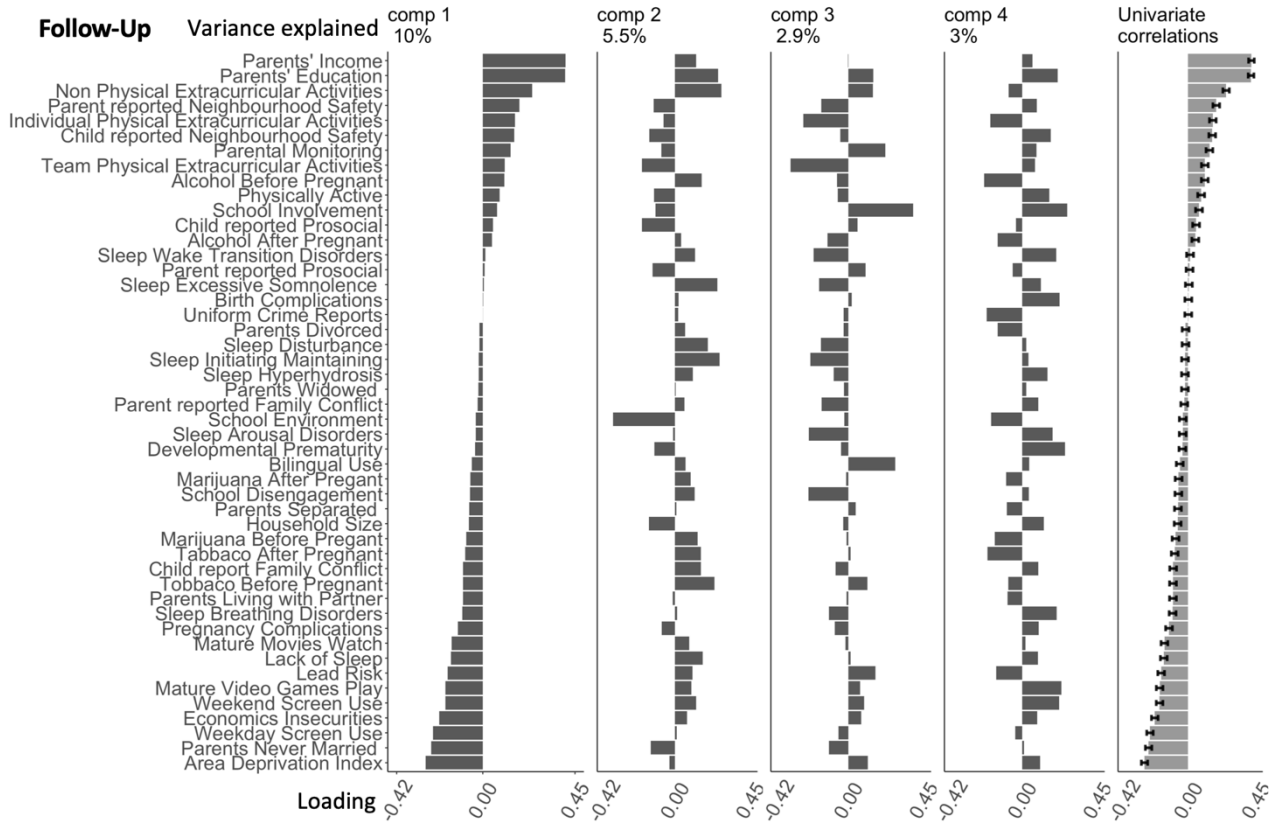



**Figure 4. Predictive models, predicting cognitive abilities from socio-demographics, lifestyles and developments via Partial Least Square (PLS).** a) Scatter plots between observed vs predicted cognitive abilities based on socio-demographics, lifestyles and developments. All data points are from test sets. *r* is the average Pearson's *r* across 21 test sites, and a value in the parenthesis is the standard deviation of Pearson's *r* across sites. b) Feature importance of socio-demographics, lifestyles and developments, predicting cognitive abilities via Partial Least Square. The features were ordered based on the loading of the first component. Univariate correlations were Pearson's correlation between each feature and cognitive abilities. Error bars reflect 95%CIs of the correlations.

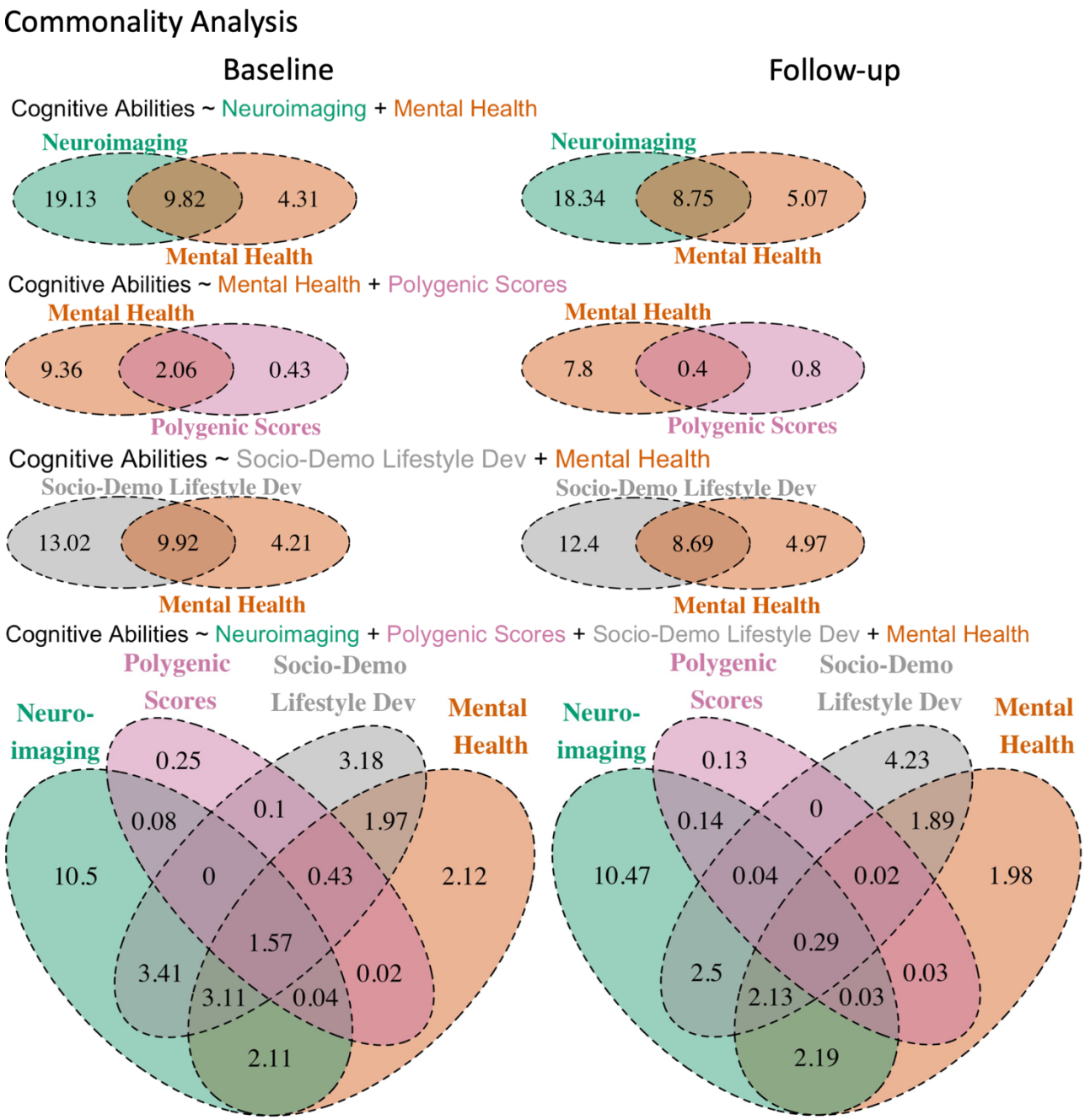



**Figure 5. Venn diagrams showing common and unique effects of proxy measures of cognitive abilities based on mental health, neuroimaging, polygenic scores and/or socio-demographics, lifestyles and developments in explaining cognitive abilities across test sites.** We computed the common and unique effects in % based on the marginal $R^2$ of four sets of linear-mixed models.

## References


1 Insel T, Cuthbert B, Garvey M, *et al.* Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *AJP* 2010; **167**: 748–51.

2 Kendler KS, Solomon M. Expert consensus v. evidence-based approaches in the revision of the DSM. *Psychological Medicine* 2016; **46**: 2255–62.

3 Pacheco J, Garvey MA, Sarampote CS, Cohen ED, Murphy ER, Friedman-Hill SR. Annual Research Review: The contributions of the RDoC research framework on understanding the neurodevelopmental origins, progression and treatment of mental illnesses. *Journal of Child Psychology and Psychiatry* 2022. DOI:10.1111/jcpp.13543.

4 Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013; **11**: 126.

5 Sanislow CA, Pine DS, Quinn KJ, *et al.* Developing constructs for psychopathology research: Research domain criteria. *Journal of Abnormal Psychology* 2010; **119**: 631–9.

6 Carter CS, Barch DM. Cognitive Neuroscience-Based Approaches to Measuring and Improving Treatment Effects on Cognition in Schizophrenia: The CNTRICS Initiative. *Schizophrenia Bulletin* 2007; **33**: 1131–7.

7 Morris SE, Cuthbert BN. Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin Neurosci* 2012; **14**: 29–37.

8 Dadi K, Varoquaux G, Houenou J, Bzdok D, Thirion B, Engemann D. Population modeling with machine learning can enhance measures of mental health. *GigaScience* 2021; **10**: giab071.

9 Pat N, Wang Y, Anney R, Riglin L, Thapar A, Stringaris A. Longitudinally stable, brain-based predictive models mediate the relationships between childhood cognition and socio-demographic, psychological and genetic factors. *Human Brain Mapping* 2022; **43**: 5520–42.

10 Vieira BH, Pamplona GSP, Fachinello K, Silva AK, Foss MP, Salmon CEG. On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. *Intelligence* 2022; **93**: 101654.

11 Allegrini AG, Selzam S, Rimfeld K, von Stumm S, Pingault JB, Plomin R. Genomic prediction of cognitive traits in childhood and adolescence. *Mol Psychiatry* 2019; **24**: 819–27.

12 Flynn JR. What is intelligence? beyond the Flynn effect, 1st expanded pbk. ed. Cambridge, UK ; New York: Cambridge University Press, 2009.

13 Kirlic N, Colaizzi JM, Cosgrove KT, *et al.* Extracurricular Activities, Screen Media Activity, and Sleep May Be Modifiable Factors Related to Children's Cognitive Functioning: Evidence From the ABCD Study®. *Child Development* 2021. DOI:10.1111/cdev.13578.

14 Morris SE, Sanislow CA, Pacheco J, Vaidyanathan U, Gordon JA, Cuthbert BN. Revisiting the seven pillars of RDoC. *BMC Medicine* 2022; **20**: 220.

15 Casey BJ, Cannonier T, Conley MI, *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience* 2018; **32**: 43–54.

16 Achenbach TM, Ivanova MY, Rescorla LA. Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry* 2017; **79**: 4–18.

17 Johnson SL, Turner RJ, Iwata N. BIS/BAS Levels and Psychiatric Disorder: An Epidemiological Study. *Journal of Psychopathology and Behavioral Assessment* 2003; **25**: 25–36.

18 Carver CS, Johnson SL. Impulsive reactivity to emotion and vulnerability to psychopathology. *American Psychologist* 2018; **73**: 1067–78.

19 Davies G, Lam M, Harris SE, *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* 2018; **9**: 2098.

20 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. *Nature Genetics* 2018; **50**: 1112–21.

21 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**: 912–9.

22 Nimon K, Lewis M, Kane R, Haynes RM. An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods* 2008; **40**: 457–66.

23 Casey BJ, Cannonier T, Conley MI, *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci* 2018; **32**: 43–54.

24 Yang R, Jernigan, Terry. Adolescent Brain Cognitive Development Study (ABCD) - Annual Release 4.0. DOI:10.15154/1523041.

25 Clark DB, Fisher CB, Bookheimer S, *et al.* Biomedical ethics and clinical oversight in multisite observational neuroimaging studies with children and adolescents: The ABCD experience. *Dev Cogn Neurosci* 2018; **32**: 143–54.

26 Whiteside SP, Lynam DR. Understanding the role of impulsivity and externalizing psychopathology in alcohol abuse: Application of the UPPS Impulsive Behavior Scale. *Experimental and Clinical Psychopharmacology* 2003; **11**: 210–7.

27 Zapolski TCB, Stairs AM, Settles RF, Combs JL, Smith GT. The Measurement of Dispositions to Rash Action in Children. *Assessment* 2010; **17**: 116–25.

28 Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology* 1994; **67**: 319–33.

29 Duncan L, Shen H, Gelaye B, *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun* 2019; **10**: 3328.

30 Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 2001; **58**: 109–30.

31 Dormann CF, Elith J, Bacher S, *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 2013; **36**: 27–46.

32 Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 2005; **67**: 301–20.

33 Tetereva A, Li J, Deng JD, Stringaris A, Pat N. Capturing brain-cognition relationship: Integrating task-based fMRI across tasks markedly boosts prediction and test-retest reliability. *NeuroImage* 2022; **263**: 119588.

34 Engemann DA, Kozynets O, Sabbagh D, *et al.* Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *Elife* 2020; **9**: e54055.

35 Breiman L. Random Forests. *Machine Learning* 2001; **45**: 5–32.

36 Raudenbush SW, Bryk AS. Hierarchical linear models: applications and data analysis methods, 2nd ed. Thousand Oaks: Sage Publications, 2002.

37 Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 2013; **4**: 133–42.

38 Abramovitch A, Short T, Schweiger A. The C Factor: Cognitive dysfunction as a transdiagnostic dimension in psychopathology. *Clinical Psychology Review* 2021; **86**: 102007.

39 Carver CS, White TL. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology* 1994; **67**: 319–33.

40 Rasero J, Sentis AI, Yeh F-C, Verstynen T. Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLOS Computational Biology* 2021; **17**: e1008347.

41 Sripada C, Angstadt M, Rutherford S, Taxali A, Shedden K. Toward a “treadmill test” for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping* 2020; **41**: 3186–97.

42 Luciana M, Bjork JM, Nagel BJ, *et al.* Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev Cogn Neurosci* 2018; **32**: 67–79.

43 Thompson WK, Barch DM, Bjork JM, *et al.* The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Dev Cogn Neurosci* 2019; **36**: 100606.