Title

The relationship between cognitive abilities and mental health as represented by cognitive abilities at the neural and genetic levels of analysis

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

Keywords: Research Domain Criteria (RDoC), Adolescent Brain Cognitive Development (ABCD), cognitive abilities, mental health, neuroimaging, polygenic scores, machine learning, environment

Abstract

Cognitive abilities are often linked to mental health across various disorders, a pattern observed even in childhood. However, the extent to which this relationship is represented by different neurobiological units of analysis, such as multimodal neuroimaging and polygenic scores (PGS), remains unclear.

Using large-scale data from the Adolescent Brain Cognitive Development (ABCD) Study, we first quantified the relationship between cognitive abilities and mental health by applying multivariate models to predict cognitive abilities from mental health in children aged 9-10, finding an out-of-sample r=.36 . We then applied similar multivariate models to predict cognitive abilities from multimodal neuroimaging, polygenic scores (PGS) and environmental factors. Multimodal neuroimaging was based on 45 types of brain MRI (e.g., task fMRI contrasts, resting-state fMRI, structural MRI, and diffusion tensor imaging). Among these MRI types, the fMRI contrast, 2-Back vs. 0-Back, from the ENBack task provided the highest predictive performance (r=.4). Combining information across all 45 types of brain MRI led to the predictive performance of r=.54. The PGS, based on previous genome-wide association studies on cognitive abilities, achieved a predictive performance of r=.25. Environmental factors, including socio-demographics (e.g., parent's income and education), lifestyles (e.g., extracurricular activities, sleep) and developmental adverse events (e.g., parental use of alcohol/tobacco, pregnancy complications), led to a predictive performance of r=.49.

In a series of separate commonality analyses, we found that the relationship between cognitive abilities and mental health was primarily represented by multimodal neuroimaging (66%) and, to a lesser extent, by PGS (21%). Additionally, environmental factors accounted for 63% of the variance in the relationship between cognitive abilities and mental health. The multimodal neuroimaging and PGS then explained 58% and 21% of the variance due to environmental factors, respectively. Notably, these patterns remained stable over two years.

Our findings underscore the significance of neurobiological units of analysis for cognitive abilities, as measured by multimodal neuroimaging and PGS, in understanding both a) the relationship between cognitive abilities and mental health and b) the variance in this relationship shared with environmental factors.

Introduction

Cognitive abilities across various domains, such as attention, working memory, declarative memory, verbal fluency, and cognitive control, are often altered in several psychiatric disorders (Millan et al., 2012). This is evident in recent meta-analyses of case-control studies involving patients with mood and anxiety disorders, obsessive-compulsive disorder, posttraumatic stress disorder, and attention-deficit/hyperactivity disorder (ADHD), among others (Abramovitch et al., 2021; East-Richard et al., 2020). Beyond typical case-control studies, the association between cognitive abilities and mental health is also observed when mental health varies from normal to abnormal in normative samples (Morris et al., 2022). For instance, our study (Pat, Riglin, et al., 2022) found an association between cognitive abilities and mental health in a relatively large, non-referred sample of 9-10-year-old children from the Adolescent Brain Cognitive Development (ABCD) study (Casey et al., 2018). In this study, we measured cognitive abilities using behavioural performance across cognitive tasks (Luciana et al., 2018) while measuring mental health using a broad range of emotional and behavioural problems (Achenbach et al., 2017). Thus, cognitive abilities are frequently considered crucial for understanding mental health issues throughout life, beginning in childhood (Abramovitch et al., 2021; Hankin et al., 2016; Morris & Cuthbert, 2012).

According to the National Institute of Mental Health's Research Domain Criteria (RDoC) framework (Insel et al., 2010), cognitive abilities should be investigated not only behaviourally but also neurobiologically, from the brain to genes. It remains unclear to what extent the relationship between cognitive abilities and mental health is represented in part by different neurobiological units of analysis -- such as neural and genetic levels measured by multimodal neuroimaging and polygenic scores (PGS). To fully comprehend the role of neurobiology in the relationship between cognitive abilities and mental health, we must also consider how these neurobiological units capture variations due to environmental factors, such as socio-demographics, lifestyles, and childhood developmental adverse events (Morris et al., 2022). Our study investigated the extent to which a) environmental factors explain the relationship between cognitive abilities and mental health, and b) cognitive abilities at the neural and genetic levels capture these associations due to environmental factors. Specifically, we conducted these investigations in a large normative group of children from the ABCD study (Casey et al., 2018). We chose to examine children because, while their emotional and behavioural problems might not meet full diagnostic criteria (Kessler et al., 2007), issues at a young age often forecast adult psychopathology (Reef et al., 2010; Roza et al., 2003). Moreover, the associations among different emotional and behavioural problems in children reflect transdiagnostic dimensions of psychopathology (Michelini et al., 2019; Pat, Riglin, et al., 2022), making children an appropriate population to study the transdiagnostic aetiology of mental health, especially within a framework that emphasises normative variation from normal to abnormal, such as the RDoC (Morris et al., 2022).

Recently, several neuroscientists have developed predictive models using neuroimaging data from brain magnetic resonance imaging (MRI) of various modalities in the so-called Brain-Wide Association Studies (BWAS) (Marek et al., 2022; Sui et al., 2020). BWAS aims to create models from MRI data that can accurately predict behavioural phenotypes in participants not included in the model-building process (Dadi et al., 2021). In one of the most extensive BWAS benchmarks to date, Marek and colleagues (2022) concluded, "More robust BWAS effects were detected for functional MRI (versus structural), cognitive tests (versus mental health questionnaires) and multivariate methods (versus univariate)." This benchmark has significant implications for using neuroimaging as a neural unit of analysis for cognitive abilities. First, while current BWAS may not be robust enough to predict mental health directly, it is more suitable for predicting cognitive abilities (see Zhi et al., 2024 for a similar conclusion). This aligns with the Research Domain Criteria (RDoC) framework, which emphasises neurobiological units of analysis for *functional domains*, such as cognitive abilities, rather than mental health itself (Cuthbert & Insel, 2013). RDoC's functional domains capture basic human functioning and include cognitive abilities along with negative/positive valence, arousal and regulation, and social and sensory processes (Morris et al., 2022). Accordingly, the current study conducted BWAS to capture cognitive abilities rather than mental health.

The second implication of Marek and colleagues' (2022) benchmark is the support it provides for using multivariate algorithms, which draw MRI information simultaneously across regions/voxels, over massively univariate algorithms that draw data from one area/voxel at a time. Similar to Marek and colleagues' (2022) study, which focused on resting-state functional MRI (rs-fMRI), our recent study on task-fMRI also found that multivariate algorithms performed superiorly, up to several folds, in predicting cognitive abilities compared to massively univariate algorithms (Pat et al., 2023). The third implication is that the performance of neuroimaging in predicting cognitive abilities depends on MRI modalities. Previous research has used brain MRI data of different modalities to predict cognitive abilities (Vieira et al., 2022). For instance, many studies have used rs-fMRI, which reflects functional connectivity between regions during rest

(Dubois et al., 2018; Keller et al., 2023; Rasero et al., 2021; Sripada et al., 2020, 2021). Others have utilised structural MRI (sMRI), which reflects anatomical morphology based on thickness, area, and volume in cortical/subcortical areas, and diffusion tensor imaging (DTI), which reflects diffusion distribution within white matter tracts (Mihalik et al., 2019; Rasero et al., 2021). While less common, task-fMRI, which reflects blood-oxygen-level-dependent (BOLD) activity relevant to each task condition, shows relatively good predictive performance, especially from specific contrasts, such as the 2-Back vs 0-Back from the N-Back working-memory task (Barch et al., 2013) nor(Makowski et al., 2024; Pat et al., 2023; Pat, Wang, et al., 2022; Sripada et al., 2020; Tetereva et al., 2022; Zhao et al., 2023). A recent meta-analysis estimated the performance of multivariate methods in predicting cognitive abilities from MRI of different modalities at around an out-of-sample *r* of 0.42 (Vieira et al., 2022). However, we and others found that this predictive performance could be further boosted by drawing information across different MRI modalities, rather than relying on only one modality (Pat, Wang, et al., 2022; Rasero et al., 2021; Tetereva et al., 2022; Tetereva & Pat, 2024). Therefore, the current study used opportunistic stacking (Engemann et al., 2020; Pat, Wang, et al., 2022). This multivariate modelling technique allowed us to combine information across MRI modalities with the added benefit of handling missing values. With opportunistic stacking, we created a 'proxy' measure of cognitive abilities (i.e., predicted value from the model) at the neural unit of analysis using multimodal neuroimaging.

Geneticists, like neuroscientists, have conducted Genome-Wide Association Studies (GWAS) to explore the links between single nucleotide polymorphisms (SNPs) and various behavioural phenotypes (Bogdan et al., 2018). Similar to BWAS, GWAS can develop predictive models from genetic profiles, resulting in polygenic scores (PGS) that predict behavioural phenotypes in participants not included in the model-building process (Choi et al., 2020). Several large-scale GWAS on cognitive abilities have been conducted, with some studies involving over 250,000 participants (Davies et al., 2018; Lee et al., 2018; Savage et al., 2018). Recently, researchers have used these large-scale GWAS to compute PGS for cognitive abilities and applied these scores to predict cognitive abilities in children (Allegrini et al., 2019; Pat, Wang, et al., 2022). For example, Allegrini et al. (2019) found that PGS based on Savage et al.'s (2018) GWAS accounted for approximately 5.3% of the variance in cognitive abilities among 12-year-old children. The current study adopted this approach with children of a similar age in the ABCD study, creating a proxy measure of cognitive abilities at the genetic unit of analysis using PGS.

Environmental factors, broadly defined, significantly influence cognitive abilities (Duyme et al., 1999; Pietschnig & Voracek, 2015). A classic example is the Flynn Effect (Flynn, 1984, 2009; Rundquist, 1936; Williams, 2013), which describes the observed rise in cognitive abilities, as measured by various cognitive tasks, across generations in the general population over time, particularly in high-income countries during the 20th century (Pietschnig & Voracek, 2015; Trahan et al., 2014; Wongupparaj et al., 2017). Experts attribute the Flynn Effect to environmental factors such as improved living standards and better education (Baker et al., 2015; Rindermann et al., 2017). Recently, researchers have used multivariate algorithms to create proxy measures of cognitive abilities in children based on environmental factors, similar to approaches used in neuroimaging and polygenic scores (PGS) (Kirlic et al., 2021; Pat, Wang, et al., 2022). These environmental factors often include socio-demographic variables (e.g., parental income/education, area deprivation index, parental marital status), lifestyle factors (e.g., screen/video game use, extracurricular activities), and developmental adverse events (e.g., parental use of alcohol/tobacco before and after pregnancy, birth complications). Studies, including ours,(Kirlic et al., 2021; Pat, Wang, et al., 2022) have applied multivariate algorithms to predict cognitive abilities from various environmental factors in the ABCD study (Casey et al., 2018). In these predictive models, parental income/education, area deprivation index, and extracurricular activities are particularly important predictors of cognitive abilities (Kirlic et al., 2021; Pat, Wang, et al., 2022). Following this approach, the current study created another proxy measure of cognitive abilities based on socio-demographics, lifestyles, and developmental adverse events.

In this study, inspired by RDoC (Insel et al., 2010), we a) focused on cognitive abilities as a functional domain, b) created predictive models to capture the continuous individual variation (as opposed to distinct categories) in cognitive abilities, c) computed two neurobiological units of analysis of cognitive abilities: multimodal neuroimaging and PGS, and d) investigated the potential contributions of environmental factors. To operationalise cognitive abilities, we estimated a latent variable representing behavioural performance across various cognitive tasks, commonly referred to as general cognitive ability or the g-factor (Deary, 2012). The g-factor was computed from various cognitive tasks pertinent to RDoC constructs, including attention, working memory, declarative memory, language, and cognitive control. However, using the g-factor to operationalise cognitive abilities caused this study to diverge from the original conceptualisation of RDoC, which emphasises studying separate constructs within cognitive abilities (Morris et al., 2022; Morris & Cuthbert, 2012). Recent studies suggest that including a general factor, such as the g-factor, in the model, rather than treating each

construct separately, improved model fit (Beam et al., 2021; Quah et al., 2025). The g-factor in children is also longitudinally stable and can forecast future health outcomes (Calvin et al., 2017; Deary et al., 2013). Notably, our previous research found that neuroimaging predicts the g-factor more accurately than predicting performance from separate individual cognitive tasks (Pat et al., 2023). Accordingly, we decided to conduct predictive models on the g-factor while keeping the RDoC's holistic, neurobiological, and basic-functioning characteristics.

Using the ABCD study (Casey et al., 2018), we first developed predictive models to estimate the cognitive abilities of unseen children based on their mental health. These models enabled us to quantify the relationship between cognitive abilities and mental health, thereby creating a proxy measure of cognitive abilities derived from mental health data. The mental health variables included children's emotional and behavioural problems (Achenbach et al., 2017) and temperaments, such as behavioural inhibition/activation (Carver & White, 1994) and impulsivity (Zapolski et al., 2010). These temperaments are linked to externalising and internalising aspects of mental health and are associated with disorders like depression, anxiety, and substance use (Carver & Johnson, 2018; S. L. Johnson et al., 2003). Next, we built predictive models of cognitive abilities using neuroimaging, polygenic scores (PGS), and socio-demographic, lifestyle, and developmental adverse event data, resulting in various proxy measures of cognitive abilities. For neuroimaging, we included 45 types of brain MRI data from task-fMRI, rs-fMRI, sMRI, and DTI. For PGS, we used three definitions of cognitive abilities based on previous large-scale GWAS (Davies et al., 2018; Lee et al., 2018; Savage et al., 2018). For socio-demographic, lifestyle, and developmental adverse events, we included 44 features, covering variables such as parental income/education, screen use, and birth/pregnancy complications. Finally, we conducted a series of commonality analyses (Nimon et al., 2008) using these proxy measures of cognitive abilities to address three specific questions. First, we examined the extent to which the relationship between cognitive abilities and mental health was represented in part by cognitive abilities at the neural and genetic levels, as measured by multimodal neuroimaging and PGS, respectively. Second, we assessed the extent to which this relationship was partly explained by environmental factors, as measured by socio-demographic, lifestyle, and developmental adverse events. Third, we tested whether the two neurobiological units of analysis for cognitive abilities, measured by multimodal neuroimaging and PGS, could account for the variance due to environmental factors. To ensure the stability of our results, we repeated the analyses at two time points (ages 9-10 and 11-12).

Results

Predictive modelling

Predicting cognitive abilities from mental health

Figure 1a and Supplementary Table 1 illustrate the predictive performance of the Partial Least Square (PLS) models in predicting cognitive abilities from mental health features. These features included: 1) emotional and behavioral problems assessed by the Child Behaviour Checklist (CBCL) (Achenbach et al., 2017), and 2) children's temperaments assessed by the Behavioural Inhibition System/Behavioural Activation System (BIS/BAS) (Carver & White, 1994) and the Urgency, Premeditation, Perseverance, Sensation seeking, and Positive urgency (UPPS-P) impulsive behaviour scale (Zapolski et al., 2010). Using these two sets of mental health features separately resulted in moderate predictive performance, with correlation coefficients ranging from $r = .24$ to $r = .31$. Combining them into a single set of features, termed "mental health," improved the performance to approximately $r = .36$, consistent across the two time points.

Figure 1b illustrates the loadings and the proportion of variance in cognitive abilities explained by each Partial Least Squares (PLS) components. The first PLS component accounted for the highest proportion of variance, ranging from 22.3% to 25.7%. This component was primarily influenced by factors such as attention and social problems, rule-breaking and aggressive behaviours and Behavioural Activation System drive. A similar pattern was observed across both time points.

Predicting cognitive abilities from neuroimaging

Figures 2a, Supplementary Figures 1-2, and Supplementary Tables 1-3 illustrate the predictive performance of the opportunistic stacking models in predicting cognitive abilities from 45 sets of neuroimaging features. The predictive performance of each set of neuroimaging features varied significantly, with correlation coefficients ranging from approximately 0 (ENBack: Negative vs. Neutral Face) to around .4 (ENBack: 2-Back vs. 0-Back). Combining information from all 45 sets of neuroimaging features into a stacked model improved the performance to approximately $r = .54$, consistent across both time points. The stacked model ($R^2 \approx .29$) explained almost twice as much variance in cognitive

abilities as the model based on the best single set of neuroimaging features (ENBack: 2-Back vs. 0-Back, $R^2 \approx .15$). Figures 2b, 3, and Supplementary Figures 3-13 highlight the feature importance of the opportunistic stacking models. Across both time points, the top contributing neuroimaging features, as indicated by SHAP values, were ENBack task-fMRI contrasts, rs-fMRI, and cortical thickness.

Predicting cognitive abilities from polygenic scores

Figures 2a and Supplementary Tables 1 illustrate the predictive performance of the Elastic Net models in predicting cognitive abilities using three polygenic scores (PGSs). The predictive accuracy of these PGSs was $r = .25$ at baseline and $r = .25$ at follow-up. Figure 2c highlights the feature importance within these models, indicating a stronger contribution from the PGS based on Savage's et al. (2018) GWAS.

Predicting cognitive abilities from socio-demographics, lifestyles and developmental adverse events

Figure 4a and Supplementary Table 1 illustrate the predictive performance of the Partial Least Square (PLS) models in predicting cognitive abilities from socio-demographics, lifestyles, and developmental adverse events. Using 44 features covering these areas, the predictive performance was around $r = .49$, consistent across the two time points. Figure 4b shows the loadings and the proportion of variance explained by these PLS models. The first PLS component accounted for the highest proportion of variance (around 10%).

Based on its loadings, this first component was: a) Positively influenced by features such as parental income and education, neighbourhood safety, and extracurricular activities, b) Negatively influenced by features such as area deprivation, having a single parent, screen use, economic insecurities, lack of sleep, playing mature video games, watching mature movies, and lead exposure.

Commonality analyses

We separately conducted the four sets of 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 27% of the variance in cognitive abilities. Specifically, 9.8% of the variance in cognitive abilities was explained by mental health, which included the common effect between the two proxy measures (6.48%) and the unique effect of mental health (3.32%) (see Supplementary Tables 4-5 and Figure 5). This indicates that 66% of the relationship between cognitive abilities and mental health, i.e., $(6.48 \div 9.8) \times 100$, was shared with neuroimaging. The common effects varied considerably across different sets of neuroimaging features, ranging from approximately 0.08% to 2.78%, with the highest being the ENBack task fMRI: 2-Back vs. 0-Back (see Supplementary Figure 14). The pattern of results was consistent across both time points.

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

At baseline, having both proxy measures based on mental health and PGSs in a linear mixed model explained 11.8% of the variance in cognitive abilities. Specifically, 9.21% of the variance in cognitive abilities was explained by mental health, which included the common effect between the two proxy measures (1.93%) and the unique effect of mental health (7.28%) (see Supplementary Tables 6-7 and Figure 5). This indicates that 21% of the relationship between cognitive abilities and mental health, i.e., $(1.93 \div 9.21) \times 100$, was shared with PGSs. The pattern of results was consistent across both time points.

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

At baseline, having both proxy measures based on mental health and socio-demographics, lifestyles, and developmental adverse events in a linear mixed model explained 24.9% of the variance in cognitive abilities. Specifically, 9.75% of the variance in cognitive abilities was explained by mental health, which included the common effect between the two proxy measures (6.12%) and the unique effect of mental health (3.63%) (see Supplementary Tables 8-9 and Figure 5). This indicates that over 63% of the relationship between cognitive abilities and mental health, i.e., $(6.12 \div 9.75) \times 100$, was shared with socio-demographics, lifestyles, and developmental adverse events. The pattern of results was consistent across both time points.

Commonality analyses for proxy measures of cognitive abilities based on mental health, neuroimaging, PGSs and socio-demographics, lifestyles and developmental adverse events

At baseline, having all four proxy measures based on mental health, neuroimaging, PGSs, and socio-demographics, lifestyles, and developmental adverse events in a linear mixed model explained 24.2% of the variance in cognitive abilities. Of the 8.97% of the variance in cognitive abilities explained by mental health, 7.05% represented common effects with the other proxy measures. This indicates that 79%, i.e., (7.05 ÷ 8.97) x 100, of the relationship between cognitive abilities and mental health was shared with the three other proxy measures (see Supplementary Tables 10-11 and Figure 5). Additionally, among the variance that socio-demographics, lifestyles, and developmental adverse events accounted for in the relationship between cognitive abilities and mental health, neuroimaging could capture 58%, while PGSs could capture 21%. The pattern of results was consistent across both time points.

Discussions

We aim to understand the extent to which the relationship between cognitive abilities and mental health is represented in part by cognitive abilities at the neural and genetic levels of analysis. We began by quantifying the relationship between cognitive abilities and mental health, finding a medium-sized out-of-sample correlation of approximately r = .36. This relationship was shared with neuroimaging (66% at baseline) and PGS (21% at baseline), based on two separate sets of commonality analyses. This suggests the significant roles of these two neurobiological units of analysis in shaping the relationship between cognitive abilities and mental health (Morris & Cuthbert, 2012). We also found that the relationship between cognitive abilities and mental health was partly shared with environmental factors, as measured by socio-demographics, lifestyles, and developmental adverse events (63% at baseline). In another set of commonality analysis, this variance due to socio-demographics, lifestyles, and developmental adverse events was explained by neuroimaging and PGS at 58% and 21%, respectively, at baseline. Accordingly, the neurobiological units of analysis for cognitive abilities captured the environmental factors, consistent with RDoC's viewpoint (Morris et al., 2022). Notably, this pattern of results remained stable over two years in early adolescence.

Our predictive modelling revealed a medium-sized predictive relationship between cognitive abilities and mental health. This finding aligns with recent meta-analyses of case-control studies that link cognitive abilities and mental disorders across various psychiatric conditions (Abramovitch et al., 2021; East-Richard et al., 2020). Unlike previous studies, we estimated the predictive, out-of-sample relationship between cognitive abilities and mental disorders in a large normative sample of children. Although our predictive models, like other cross-sectional models, cannot determine the directionality of the effects, the strength of the relationship between cognitive abilities and mental health estimated here should be more robust than when calculated using the same sample as the model itself, known as in-sample prediction/association (Marek et al., 2022; Yarkoni & Westfall, 2017). Examining the PLS loadings of our predictive models revealed that the relationship was driven by various aspects of mental health, including thought and externalising symptoms, as well as motivation. This suggests that there are multiple pathways—encompassing a broad range of emotional and behavioural problems and temperaments—through which cognitive abilities and mental health are linked.

Our predictive modelling created proxy measures of cognitive abilities based on two neurobiological units of analysis: neuroimaging and polygenic scores (PGS) (Morris & Cuthbert, 2012). For neuroimaging, inspired by recent BWAS benchmarks (Engemann et al., 2020; Marek et al., 2022), we used a multivariate modelling technique called opportunistic stacking, which integrates information across various MRI features and modalities. Combining 45 sets of neuroimaging features resulted in relatively high predictive performance (out-of-sample *r* = .54 at baseline), compared to using any single set. This finding aligns with previous research that pooled multiple neuroimaging modalities (Engemann et al., 2020; Rasero et al., 2021; Tetereva et al., 2022). This level of predictive performance is numerically higher than that found in a recent meta-analysis, which mainly included studies using only one set of neuroimaging features, with an *r* of 0.42 (Vieira et al., 2022). Moreover, this performance level in predicting cognitive abilities is nearly the same as our previous attempt using a similar stacking technique to integrate MRI modalities in young adult samples from the Human Connectome Project (HCP) (Van Essen et al., 2013), which achieved an out-of-sample *r* = .57 (Tetereva et al., 2022). Similarly, in the current study, the top contributing set of neuroimaging features, the 2-Back vs. 0-Back task fMRI, was consistent with previous studies using the HCP (Sripada et al., 2020; Tetereva et al., 2022). Altogether, this demonstrates the robustness of our proxy measure of cognitive abilities based on multimodal neuroimaging. In addition to predictive performance, opportunistic stacking offers the added benefit of handling missing values (Engemann et al., 2020; Pat, Wang, et al., 2022), allowing us to retain data from 10,754 participants who completed the cognitive tasks at baseline and has at least one set of neuroimaging features. Consequently, with opportunistic stacking, we were more likely to retain MRI data from participants with higher fMRI noise, such as those with socioeconomic disadvantages (Cosgrove et al., 2022). More importantly, we demonstrated that the proxy measure based on multimodal neuroimaging explained the majority of the variance in the relationship between cognitive abilities and mental health, underscoring its significant role as a neurobiological unit of analysis for cognitive abilities (Morris & Cuthbert, 2012).

For PGS, we created a proxy measure based on three large-scale GWAS on cognitive abilities (Davies et al., 2018; Lee et al., 2018; Savage et al., 2018). Using PGS resulted in a numerically weaker predictive performance (out-of-sample r = .25 at baseline) compared to multimodal neuroimaging. However, this predictive strength is still comparable to previous research. For instance, Allegrini and colleagues (2019) used a different cohort of children and found R^2 = .053 when applying PGS based on Savage and colleagues' (2018) to predict the cognitive abilities of 12-year-old children. Given that PGS based on Savage and colleagues' (2018) also drove the prediction in the current study, as seen in its feature

importance, this similar level of predictive performance between Allegrini and colleagues (2019) and our study suggests consistency in the predictive performance of PGS. Despite this level of performance, PGS was able to explain some variance (21% at baseline) in the relationship between cognitive abilities and mental health, indicating some capacity of PGS as a neurobiological unit of analysis for cognitive abilities.

There are multiple potential reasons why PGS performed much poorer than multimodal neuroimaging. Firstly, unlike genes, the brain changes throughout development and lifespan (Bethlehem et al., 2022), and so do cognitive abilities (Hartshorne & Germine, 2015). This dynamic nature might make multimodal neuroimaging a better tool for tracing cognitive abilities. Secondly, there might be a mismatch in the age of participants between the original GWAS (Davies et al., 2018; Lee et al., 2018; Savage et al., 2018) and the current study. While the original GWAS conducted meta-analyses pooling data from participants aged 5 to 102, these studies might draw more heavily from older cohorts with large participant numbers, such as the UK Biobank (Sudlow et al., 2015). Allegrini and colleagues (2019) also demonstrated that PGS performs better in predicting cognitive abilities in older children (aged 16) compared to younger ones (aged 12). Therefore, a more child-specific PGS might be needed to explain more variance in children. Thirdly, the PGS used here included only common SNPs and not rare variants. Recent studies using whole-genome sequence data have found that rare variants contribute to the heritability of complex traits, such as height and body mass index (Wainschtein et al., 2022). Given that cognitive abilities are also complex traits, future studies might need to examine if including rare variants can improve the predictive performance of PGS.

Similarly, our predictive modelling created proxy measures of cognitive abilities for environmental factors based on socio-demographics, lifestyles, and developmental adverse events. In line with previous work (Kirlic et al., 2021; Pat, Wang, et al., 2022), we could predict unseen children's cognitive abilities based on their socio-demographics, lifestyles, and developmental adverse events with a medium-to-high out-of-sample $r = .49$ (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 developmental adverse events (e.g., pregnancy complications). Importantly, proxy measures based on socio-demographics, lifestyles, and developmental adverse events captured a large proportion of the relationship between cognitive abilities and mental health. Furthermore, this variance captured by socio-demographics, lifestyles, and developmental adverse events overlapped mainly with the neurobiological proxy measures. This reiterates RDoC's central tenet that understanding the neurobiology of a functional domain, such as cognitive abilities, could help us understand the extent to which environments influence mental health (Cuthbert & Insel, 2013; Insel et al., 2010). More importantly, all the results regarding neuroimaging, PGS, and socio-demographics, lifestyles, and developmental adverse events were reliable across two years during a sensitive period for adolescents.

This study has several limitations that might affect its generalisability. Firstly, the range of mental health variables was not exhaustive. While we covered various emotional and behavioural problems (Achenbach et al., 2017) and temperaments, including behavioural inhibition/activation (Carver & White, 1994) and impulsivity (Zapolski et al., 2010), we may still miss other critical mental health variables, such as psychotic-like experiences, eating disorder symptoms, and mania. Similarly, our ABCD samples were young and community-based, likely limiting the severity of their psychopathological issues (Kessler et al., 2007). Future work needs to test if the results found here are generalisable to adults and participants with stronger severity. Next, for cognitive abilities, while the six cognitive tasks (Luciana et al., 2018; Thompson et al., 2019) covered most of the RDoC cognitive abilities/systems constructs, we still missed variability in some domains, such as perception (Morris & Cuthbert, 2012). Additionally, several children (3,274) did not complete all six cognitive tasks at follow-up, which might create a discrepancy between baseline and follow-up samples. However, the differences in social demographics, lifestyles and developmental adverse events between participants who provided cognitive scores in the follow up were minimal (Cohen's d ranging from 0.007 to 0.092, see Supplementary Table 18). Moreover, given that we found a similar pattern of predictive performance across the two time points, we believe excluding the children who did not complete the cognitive tasks at follow up should not alter our conclusions.

Furthermore, while we used comprehensive multimodal MRI from 45 sets of features for neuroimaging, three fMRI tasks were not chosen based on their relevance to cognitive abilities (Casey et al., 2018). It is possible to obtain higher predictive performance based on other fMRI tasks. For all analyses involving PGS, we limited our participants to children of European ancestry due to the lack of summary statistics from well-powered GWAS for cognitive abilities in non-European participants. This prevented us from fully leveraging the diverse samples in the ABCD study (Garavan et al., 2018). Future GWAS work with more diverse samples is needed to ensure equity and fairness in developing

neurobiological units of analysis for cognitive abilities. Lastly, we relied on 44 variables of socio-demographics, lifestyles, and developmental adverse events included in the study, which might have missed some variables relevant to cognitive abilities (e.g., nutrition). The ABCD study (Casey et al., 2018) is ongoing, and future data might address some of these limitations.

Overall, aligning with the RDoC perspective (Morris & Cuthbert, 2012), our findings support the use of neurobiological units of analysis for cognitive abilities, as assessed through multimodal neuroimaging and Polygenic Scores (PGS). These measures explain (a) the relationship between cognitive abilities and mental health and (b) the variance in this cognitive-ability-and-mental-health relationship attributable to environmental factors. Our results emphasise the importance of considering both neurobiology and environmental factors, such as socio-demographics, lifestyles, and adverse childhood events, to gain a comprehensive understanding of the aetiology of mental health (Insel et al., 2010; Morris et al., 2022).

Methods and Materials

The Adolescent Brain Cognitive Development (ABCD) Study

We used data from the Adolescent Brain Cognitive Development (ABCD) Study Curated Annual Release 5.1 (DOI:10.15154/z563-zd24) from two time points. The baseline included data from 11,868 children (5,677 females and 3 others, aged 9-10 years), while the two-year follow-up included data from the same children two years later (10,908 children, 5,181 females and 3 others). Although the ABCD collected data from 22 sites across the United States, we excluded data from Site 22 since this site only provided data from 35 children at baseline and none at follow-up (Garavan et al., 2018). We also excluded 69 children based on the Snellen Vision Screener (Luciana et al., 2018; Snellen, 1862). These children either could not read any line on the chart, could only read the largest line, or could read up to the fourth line clearly but had difficulty reading stimuli on an iPad used for administering cognitive tasks (explained below). We listed the number of participants following each inclusion and exclusion criteria for each variable in Supplementary Figure 15 and Supplementary Tables 12-13. Institutional Review Boards at each site approved the study protocols. Please see Clark and colleagues (2018) for ethical details, such as informed consent and confidentiality.

Measures: Cognitive abilities

Cognitive abilities were assessed using six cognitive tasks collected with an iPad during a 70-minute session outside of MRI at baseline and two-year follow-up (Luciana et al., 2018; Thompson et al., 2019). The first task was Picture Vocabulary, which measured language comprehension (Gershon et al., 2014). The second task was Oral Reading Recognition, which measured language decoding (Bleck et al., 2013). The third task was Flanker, which measured conflict monitoring and inhibitory control (Eriksen & Eriksen, 1974). The fourth task was Pattern Comparison Processing, which measured the speed of processing patterns (Carlozzi et al., 2013). The fifth task was Picture Sequence Memory, which measured episodic memory (Bauer et al., 2013). The sixth task was Rey-Auditory Verbal Learning, which measured memory recall after distraction and a short delay (Daniel & Wahlstrom, 2014). Rey-Auditory Verbal Learning was sourced from Pearson Assessment, while the other five cognitive tasks were from the NIH Toolbox (Bleck et al., 2013; Luciana et al., 2018). The ABCD study administered the Dimensional Change Card Sort and List Sorting Working Memory tasks from the NIH Toolbox (Bleck et al., 2013) only at baseline, not at the two-year follow-up (see DOI: 10.15154/z563-zd24). Consequently, these two tasks were not analysed in the current study. Additionally, 3,274 children at follow-up did not complete some of these tasks and were therefore excluded from the follow-up data analysis.

We operationalised individual differences in cognitive abilities across the six cognitive tasks as a factor score of a latent variable, the 'g-factor'. To estimate this factor score, we fit the standardised performance of the six cognitive tasks to a second-order confirmatory factor analysis (CFA) of a 'g-factor' model, similar to previous work (Ang et al., 2020; Pat, Riglin, et al., 2022; Pat, Wang, et al., 2022; Thompson et al., 2019). In this CFA, we treated the g-factor as the second-order latent variable that underpinned three first-order latent variables, each with two manifest variables: 1) 'language,' underlying Picture Vocabulary and Oral Reading Recognition, 2) 'mental flexibility,' underlying Flanker and Pattern Comparison Processing, and 3) 'memory recall,' underlying Picture Sequence Memory and Rey-Auditory Verbal Learning.

We fixed the variance of the latent factors to one and applied the Maximum Likelihood with Robust standard errors (MLR) approach with Huber-White standard errors and scaled test statistics. To provide information about the internal consistency of the g-factor, we calculated OmegaL2 (Jorgensen et al., 2022). We used the *lavaan* (Rosseel, 2012) (version 0.6-15), *semTools* (Jorgensen et al., 2022), and *semPlots* (Epskamp, 2015) packages for this CFA of cognitive abilities.

We found the second-order 'g-factor' model to fit cognitive abilities well across the six cognitive tasks. This is evidenced by several indices if we apply the model to the whole baseline data: scaled and robust CFI (.994), TLI (.986), RMSEA (.031, 90% CI [.024-.037]), robust SRMR (.013), and OmegaL2 (.78). See Supplementary Figure 16 for the standardised weights of this CFA model. This enabled us to use the factor score of the latent variable 'g-factor' as the target for our predictive models.

Measures: Mental health

Mental health was assessed using two sets of features. The first set involved parental reports of children's emotional and behavioural problems, as measured by the Child Behaviour Checklist (CBCL) (Achenbach et al., 2017).We used eight summary scores: anxious/depressed, withdrawn, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviours, and aggressive behaviours. For CBCL, caretakers rated each item as 0 = not true (as far as you know), 1 = somewhat or sometimes true, and 2 = very true or often true. The third set assessed children's temperaments, conceptualised as risk factors for mental issues (S. L. Johnson et al., 2003; Whiteside & Lynam, 2003), using the Urgency, Premeditation, Perseverance, Sensation Seeking, and Positive Urgency (UPPS-P) Impulsive Behaviour Scale (Zapolski et al., 2010) and the Behavioural Inhibition System/Behavioural Activation System (BIS/BAS) (Carver & White, 1994). We used nine summary scores: negative urgency, lack of planning, sensation seeking, positive urgency, lack of perseverance, BIS, BAS reward responsiveness, BAS drive, and BAS fun. Supplementary Tables 14-15 provide summary statistics, histograms, and missing values for measures of mental health. They also include the actual variable names listed in the data dictionary and their calculations.

Measures: Neuroimaging

Neuroimaging data were based on the tabulated brain-MRI data pre-processed by the ABCD. We organized the brain-MRI data into 45 sets of neuroimaging features, covering task-fMRI (including ENBack, stop signal (SST), and monetary incentive delay (MID) tasks), resting-state fMRI, structural MRI, and diffusion tensor imaging (DTI). The ABCD provided details on MRI acquisition and image processing elsewhere (Hagler et al., 2019; Yang & Jernigan, Terry, n.d.).

The ABCD study provided recommended exclusion criteria for neuroimaging data based on automated and manual quality control (Yang & Jernigan, Terry, n.d.). Specifically, the study created an exclusion flag for each neuroimaging feature (with the prefix 'imgincl' in the 'abcd_imgincl01' table) based on criteria involving image quality, MR neurological screening, behavioural performance, and the number of repetition times (TRs), among others. We removed the entire set of neuroimaging features from each participant if any of its features were flagged or missing. We also detected outliers with over three interquartile ranges from the nearest quartile for each neuroimaging feature. We excluded a particular set of neuroimaging features from each participant when this set had outliers over 5% of the total number of its neuroimaging features. For instance, for the 2-Back vs 0-Back contrast from the ENBack task-fMRI, we had 167 features (i.e., brain regions) based on the brain parcellation atlas used by the ABCD. If (a) one of the 167 features had an exclusion flag, (b) a participant had a vision problem, (c) any of the 167 features was missing, (d) at least nine features (i.e., over 5%) were outliers, then we would remove this 2-Back vs 0-Back contrast from a particular participant but still keep other sets of neuroimaging features that did not meet these criteria (see Supplementary Supplementary Table 12-13 for the number of participants after each exclusion criterion for each set of neuroimaging features).

We standardised each neuroimaging feature across participants and harmonized variation across MRI scanners using ComBat (Fortin et al., 2017; W. E. Johnson et al., 2007; Nielson et al., 2018). Note that under predictive modelling, we discuss strategies we implemented to avoid data leakage and to model the data with missing values using the opportunistic stacking technique (Engemann et al., 2020; Pat, Wang, et al., 2022).

<u>Sets of Neuroimaging Features 1-26: task-fMRI</u>

We used unthresholded generalised-linear model (GLM) contrasts, averaged across two runs (Bolt et al., 2017; Pat et al., 2023; Pat, Wang, et al., 2022) for task-fMRI sets of features. These contrasts were embedded in the brain parcels based on the FreeSurfer's atlases (Dale et al., 1999): 148 cortical-surface Destrieux parcels (Destrieux et al., 2010) and subcortical-volumetric 19 ASEG parcels (Fischl et al., 2002), resulting in 167 features in each task-fMRI set.

Sets of Neuroimaging Features 1-9: ENBack task-fMRI

The "ENBack" or emotional n-back task was designed to elicit fMRI activity related to working memory to neutral and emotional stimuli (Barch et al., 2013). Depending on the block, the children were asked whether an image matched the image shown two trials earlier (2-Back) or at the beginning (0-Back). In this task version, the images shown included emotional faces and places. Thus, in addition to working memory, the task also allowed us to extract fMRI activity related to emotion processing and facial processing. We used the following contrasts as nine separate sets of neuroimaging features for ENBack task-fMRI: 2-Back vs 0-Back, Face vs Place, Emotion vs Neutral Face, Positive vs Neutral Face, Negative vs Neutral Face, 2-Back, 0-Back, Emotion and Place.

Sets of Neuroimaging Features 10-19: Monetary Incentive Delay (MID) task-fMRI

The MID task was designed to elicit fMRI activity related to reward processing (Knutson et al., 2000). In this task, children responded to a stimulus shown on a screen. If they responded before the stimulus disappeared, they could either win $5 (Large Reward), win $0.2 (Small Reward), lose $5 (Large Loss), lose $0.2 (Small Loss), or not win or lose any money (Neutral), depending on the conditions. At the end of each trial, they were shown feedback on whether they won money (Positive Reward Feedback), did not win money (Negative Reward Feedback), avoided losing money (Positive Punishment Feedback), or lost money (Negative Punishment Feedback). We used the following contrasts as ten separate sets of neuroimaging features for MID task-fMRI: Large Reward vs Small Reward anticipation, Small Reward vs Neutral anticipation, Large Reward vs Neutral anticipation, Large Loss vs Small Loss anticipation, Small Loss vs Neutral anticipation, Large Loss vs Neutral anticipation, Loss vs Neutral anticipation, Reward vs Neutral anticipation, Positive vs Negative Reward Feedback, and Positive vs Negative Punishment Feedback.

Sets of Neuroimaging Features 20-26: Stop-Signal Task (SST) task-fMRI

The SST was designed to elicit fMRI activity related to inhibitory control (Whelan et al., 2012). Children were asked to withhold or interrupt their motor response to a 'Go' stimulus whenever they saw a 'Stop' signal. We used two additional quality-control exclusion criteria for the SST task: *tfmri_sst_beh_glitchflag* and *tfmri_sst_beh_violatorflag*, which notified glitches as recommended (Bissett et al., 2021; Garavan et al., 2018). We used the following contrasts as seven separate sets of neuroimaging features for SST task-fMRI: Incorrect Go vs Incorrect Stop, Incorrect Go vs Correct Go, Correct Stop vs Incorrect Stop, Any Stop vs Correct Go, Incorrect Stop vs Correct Go, Correct Stop vs Correct Go, and Correct Go vs Fixation.

Sets of Neuroimaging Features 27-29: Resting-state fMRI (rs-fMRI)

The ABCD study collected rs-fMRI data for 20 minutes while children viewed a crosshair. The study described the pre-processing procedure elsewhere (Hagler et al., 2019). The investigators parcellated the cortical surface into 333 regions and the subcortical volume into 19 regions using Gordon's (Gordon et al., 2016) and ASEG (Fischl et al., 2002) atlases, respectively. They grouped the cortical-surface regions into 13 predefined large-scale cortical networks (Gordon et al., 2016). These large-scale cortical networks included auditory, cingulo-opercular, cingulo-parietal, default-mode, dorsal-attention, frontoparietal, none, retrosplenial-temporal, salience, sensorimotor-hand, sensorimotor-mouth, ventral-attention, and visual networks. Note that the term 'None' refers to regions that did not belong to any network. They then correlated time series from these regions and applied Fisher's z-transformation to the correlations. We included three sets of neuroimaging features for rs-fMRI. The first set was cortical functional connectivity (FC) with 91 features, including the mean values of the correlations between pairs of regions within the same large-scale cortical network and between large-scale cortical networks. The second set was subcortical-network FC with 247 features, including the mean values of the correlations between each of the 19 subcortical regions and the 13 large-scale cortical networks. The third set was temporal variance with 352 features (i.e., 333 cortical and 19 subcortical regions), representing the variance across time calculated for each parcellated region. Temporal variance reflects the magnitude of low-frequency oscillations (Yang & Jernigan, Terry, n.d.).

Sets of Neuroimaging Features 30-44: Structural MRI (sMRI)

The ABCD study collected T1-weighted and T2-weighted 3D sMRI images and quantified them into various measures, mainly through FreeSurfer v7.1.1 (Yang & Jernigan, Terry, n.d.). Similar to task-fMRI, we used 148 cortical-surface Destrieux (Destrieux et al., 2010) and subcortical-volumetric 19 ASEG (Fischl et al., 2002) atlases, resulting in 167 features. We included 15 sets of neuroimaging features for sMRI: cortical thickness, cortical area, cortical volume, sulcal depth, T1 white-matter averaged intensity, T1 grey-matter averaged intensity, T1 normalised intensity, T2 white-matter averaged intensity, T2 grey-matter averaged intensity, T2 normalised intensity, T1 summations, T2 summations, T1 subcortical averaged intensity, T2 subcortical averaged intensity and subcortical volume. Note: see Figure 3 for the neuroimaging features included in T1 and T2 summations and those figures are enlarged in Supplement Figure 8 and 13 for baseline and follow-up respectively.

Sets of Neuroimaging Features 45: Diffusion tensor imaging (DTI)

We included fractional anisotropy (FA) derived from DTI as another set of neuroimaging features. FA characterizes the directionality of diffusion within white matter tracts, which is thought to indicate the density of fiber packing (Alexander et al., 2007). The ABCD study used AtlasTrack (Hagler et al., 2009, 2019) to segment major white matter tracts. These included the corpus callosum, forceps major, forceps minor, cingulate and parahippocampal portions of the cingulum, fornix, inferior fronto-occipital fasciculus, inferior longitudinal fasciculus, pyramidal/corticospinal tract, superior longitudinal fasciculus, temporal lobe portion of the superior longitudinal fasciculus, anterior thalamic radiations, and uncinate. Given that ten tracts were separately labelled for each hemisphere, there were 23 neuroimaging features in this set.

Measures: Polygenic Scores

Genetic profiles were constructed based on polygenic scores (PGS) of cognitive abilities. The ABCD study provides detailed notes on genotyping in another source (Uban et al., 2018). Briefly, the study genotyped saliva and whole blood samples using Smokescreen™ Array. The investigators then quality-controlled the data using calling signals and variant call rates, applied the Ricopili pipeline and imputed the data with TOPMED (see https://topmedimpute.readthedocs.io/). The study also identified problematic plates and data points with a subject-matching issue. We further excluded children with minimal or excessive heterozygosity and excluded Single Nucleotide Polymorphisms (SNPs) based on minor allele frequency (less than 5%) and violations of Hardy–Weinberg equilibrium (P-value less than 1E−10) (details can be found at https://github.com/ricanney/stata).

We calculated PGS using three definitions from three large-scale genome-wide association studies (GWAS) on cognitive abilities: n=300,486 participants aged 16 to 102 (Davies et al., 2018), n=257,84 participants aged 8 to 96 (Lee et al., 2018) and n=269,867 participants aged 5 to 98 (Savage et al., 2018). These GWAS synthesised findings from different cohorts that collected cognitive tasks. Due to the diversity in cognitive tasks used across cohorts, they defined cognitive abilities in unique ways. For instance, Lee and colleagues (2018) utilised principal component analysis to consolidate various cognitive task scores into a single measure within each cohort from the Cognitive Genomics Consortium (COGENT) consortia (Lencz et al., 2014), but only focused on the verbal-numerical reasoning (VNR) test within the UK Biobank cohort (Sudlow et al., 2015). In a similar approach, Davies and colleagues (2018) employed principal component analysis to capture cognitive abilities from different cohorts within both CHARGE consortium data sets (Psaty et al., 2009) and COGENT (Lencz et al., 2014). They also focused on VNR testing within UK Biobank (Sudlow et al., 2015). Similarly, Savage and colleagues (2018) calculated a singular score for cognitive abilities using 'a single sum score, mean score, or factor score' collated from various tasks across thirteen cohort studies alongside logistic regression in one case-control study.

Participants in these GWAS were of European ancestry. 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 (Duncan et al., 2019), we restricted all analyses involving PGS to 5,776 children of European ancestry. These children were within four standard deviations from the mean of the top four principal components (PCs) of the super-population individuals in the 1000 Genomes Project Consortium Phase 3 reference (Auton et al., 2015).

We employed the Pthreshold approach (Choi et al., 2020). In this approach, we defined 'risk' alleles as those associated with cognitive abilities in the three discovery GWASs (Davies et al., 2018; Lee et al., 2018; Savage et al., 2018) at ten different PGS thresholds: 0.5, 0.1, 0.05, 0.01, 0.001, 0.0001, 0.00001, 0.000001, 0.0000001, 0.00000001. We then computed PGS as the Z-scored, weighted mean number of linkage-independent risk alleles in approximate linkage equilibrium derived from imputed autosomal SNPs. We selected the best PGS threshold for each of the three

definitions by choosing the PGS threshold that demonstrated the strongest correlation between its PGS and cognitive abilities in the ABCD (i.e., the g-factor factor score). Refer to the section on predictive modelling below for strategies we implemented to avoid data leakage due to this selection of the PGS threshold and the family structure in the ABCD.

Measures: sociodemographics, lifestyles and developmental adverse events

Environmental factors were based on 44 features, covering socio-demographics, lifestyles, and developmental adverse events. This included (a) 14 features for child social-demographics (Zucker et al., 2018), including bilingual use (Dick et al., 2019), parental marital status, parental education, parental income, household size, economic insecurities, area deprivation index (Kind et al., 2014), lead risk (Frostenson et al., n.d.), crime report (Federal Bureau Of Investigation, 2012), neighbourhood safety (Echeverria et al., 2004), school environment, involvement and disengagement (Stover et al., 2010), (b) five features for child social interactions from Parent Monitoring scale (Chilcoat & Anthony, 1996), Child Report of Behaviour Inventory (Schaefer, 1965), Strength and Difficulties Questionnaire (Goodman et al., 2003) and Moos Family Environment Scale (Moos et al., 1974), (c) eight features from child's sleep problems based on the Sleep Disturbance scale (Bruni et al., 1996), (d) four features for child's physical activities from Youth Risk Behaviour Survey (Centers for Disease Control and Prevention, 2023), (e) four features for child screen use (Bagot et al., 2018), (f) six features for parental use of alcohol, tobacco and marijuana before and after pregnancy from the Developmental History Questionnaire (Kessler et al., 2009; Merikangas et al., 2009), and (g) three features for developmental adverse events from the Developmental History Questionnaire, including prematurity and birth and pregnancy complications (Kessler et al., 2009; Merikangas et al., 2009). Note that we treated developmental adverse events from the Developmental History Questionnaire as environmental factors, as these events are either parental behaviours (e.g., parental use of alcohol, tobacco and marijuana) or parental medical conditions (e.g., pregnancy complications) that affect children. Supplementary Tables 16-17 provide summary statistics, histograms, and missing values for measures of socio-demographics, lifestyles and developmental adverse events. They also include the actual variable names listed in the data dictionary and their calculations.

Predictive modelling

For building predictive multivariate models, we implemented a nested leave-one-site-out cross-validation. Specifically, we treated one out of 21 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 site was a test set once and reported overall predictive performance across all test sites. Within each training set, we applied 10-fold cross-validation to tune the hyperparameters of the predictive models. The nested leave-one-site-out cross-validation allowed us to ensure the generalisability of our predictive models to unseen sites. This is important because different sites involved different MRI machines, experimenters, and participants of other demographics (Garavan et al., 2018). Next, data from children from the same family were collected from the same site. Accordingly, using leave-one-site-out also prevented data leakage due to family structure, which might inflate the predictive performance of the models, particularly those involving polygenic scores. Still, given the different number of participants in each site, one drawback for the nested leave-one-site-out cross-validation is that we ended up with some test sets with fewer participants than others. Accordingly, we provided a supplemental analysis using the classical nested cross-validation, which included ten non-overlapping outer folds, randomly chosen without considering the site information, as test sets and ten inner folds for hyperparameter tuning (see Supplementary Figure 17). Briefly the results of the leave-one-site-out cross-validation and classical nested cross-validation were close to each other, albeit classical nested cross-validation having slightly higher performance.

To demonstrate the stability of the results across two years, we built the predictive models (including hyperparameter tuning) separately for baseline and follow-up data. We separately applied standardisation to the baseline training and test sets for both the target and features to prevent data leakage between training and test sets. To ensure similarity in the data scale across two time points, we used the mean and standard deviation of the baseline training and test sets to standardise the follow-up training and test sets, respectively. For cognitive abilities, which were used as the target for all predictive models, we applied this standardisation strategy both before CFA (i.e., to the behavioural performance of the six cognitive tasks) and after CFA (i.e., to the g-factor factor scores). Moreover, we only estimated the CFA of cognitive abilities- using the baseline training set to ensure that the predictive models of the two time points had the same target. We then applied this estimated CFA model to the baseline test set and follow-up training and test sets. We examined the predictive performance of the models via the relationship between predicted and

observed cognitive abilities, using Pearson's correlation (*r*), coefficient of determination (R^2, calculated using the sum of square definition), mean-absolute error (MAE) and root mean square error (RMSE).

Predicting cognitive abilities from mental health

We developed predictive models to predict cognitive abilities from three sets of mental health features: CBCL and temperaments. We separately modelled each of these two sets and also simultaneously modelled the two sets by concatenating them into one set of features called "mental health". We implemented Partial Least Squares (PLS) (Wold et al., 2001) as a multivariate algorithm for these predictive models. Note that while PLS is sometimes used for reducing the dimensionality of features within a dataset, here we utilised PLS in a predictive framework: we tuned and estimated PLS loadings in each training set and applied the final model to the corresponding test set. PLS decomposes features into components that capture not only the features' variance but also the target's variance (Wold et al., 2001). PLS has an advantage in dealing with collinear features (Dormann et al., 2013), typical for mental health issues (Caspi & Moffitt, 2018).

PLS has one hyperparameter, the number of components. In our grid search, we tested the number of components, ranging from one to the total number of features. We selected the number of components based on the drop in root mean square error (RMSE). We kept increasing the number of components until the component did not reduce 0.1% of the total RMSE. We fit PLS using the *mixOmics* package (Rohart et al., 2017) with the *tidymodels* package as a wrapper (Kuhn & Wickham, 2018/2023).

To understand how PLS made predictions, we examined loadings and the proportion of variance explained. Loadings for each PLS component show how much each feature contributes to each PLS component. The proportion of variance explained shows how much variance each PLS component captures compared to the total variance. We then compared loadings and the proportion of variance explained with the univariate Pearson's correlation between each feature and the target. Note that because we could not guarantee that each training set would result in the same PLS components, we calculated loadings and the proportion of variance explained on the full data without splitting them into training and test sets. It is important to note that the loadings and the proportion of variance explained are for understanding the models, but for assessing the predictive performance and computing a proxy measure of cognitive abilities (i.e., the predicted values), we still relied on the nested leave-one-site-out cross-validation.

Predicting cognitive abilities from neuroimaging

We developed predictive models to predict cognitive abilities from 45 sets of neuroimaging features. To avoid data leakage, we detected the outliers separately in the baseline training, baseline test, follow-up training and follow-up test sets. Similarly, to harmonise neuroimaging features across different sites while avoiding data leakage, we applied ComBat (Fortin et al., 2017; W. E. Johnson et al., 2007; Nielson et al., 2018) to the training set. We then applied ComBat to the test set, using the ComBatted training set as a reference batch.

Unlike PLS used above for predictive models from mental health, we chose to apply opportunistic stacking (Engemann et al., 2020; Pat, Wang, et al., 2022) when building predictive models from neuroimaging. As we showed previously (Pat, Wang, et al., 2022), opportunistic stacking allowed us to handle missingness in the neuroimaging data without sacrificing predictive performance. Missingness in children's MRI data is expected, given high levels of noise (e.g., movement artifact) (Fassbender et al., 2017). For the ABCD, if we applied listwise exclusion using the study's exclusion criteria and outlier detection, we would have to exclude around 68% and 74%, at baseline and followup respectively of the children with MRI data from any set of neuroimaging features flagged (Pat, Wang, et al., 2022) (see Supplementary Figure 18). With opportunistic stacking, we only required each participant to have at least one out of 45 sets of neuroimaging features available. Therefore, we needed to exclude just around 9% and 41%, at baseline and followup respectively, of the children (see Supplementary Figure 18). Our opportunistic stacking method kept 10,754 and 6,412 participants at baseline and follow-up, respectively, while listwise deletion only kept 3,784 and 2,788 participants, respectively. We previously showed that the predictive performance of the models with opportunistic stacking is similar to that with listwise exclusion (Pat, Wang, et al., 2022).

Opportunistic stacking (Engemann et al., 2020; Pat, Wang, et al., 2022) involves two layers of modelling: set-specific and stacking layers. In the set-specific layer, we predicted cognitive abilities separately from each set of neuroimaging features using Elastic Net (Zou & Hastie, 2005). While being a linear and non-interactive algorithm, Elastic Net performs relatively well in predicting behaviours from neuroimaging MRI, often on par with, if not better than, other

non-linear and interactive algorithms, such as support vector machine with non-linear kernel, XGBoost and Random Forest (Pat et al., 2023; Tetereva et al., 2022; Vieira et al., 2022). Moreover, Elastic Net coefficients are readily explainable, enabling us to explain how the models drew information from each neuroimaging feature when making a prediction (Molnar, 2019; Pat et al., 2023).

Elastic Net simultaneously minimises the weighted sum of the features' coefficients. Its loss function can be written as:

$$\mathrm{L}_{\mathrm{enet}}(\widehat{\boldsymbol{\beta}}) = \frac{\sum_{i=1}^{n}(y_i - \boldsymbol{x_i}\widehat{\boldsymbol{\beta}})^2}{2n} + \lambda\left(\frac{1-\alpha}{2}\sum_{j=1}^{m}\widehat{\beta_j^2} + \alpha\sum_{j=1}^{m}\left|\widehat{\beta_j}\right|\right), \quad (1)$$

where $\mathbf{x_i}$ is a row vector of all the features in observation i, and $\widehat{\boldsymbol{\beta}}$ is a column vector of features' coefficient. There are two hyperparameters: (1) the penalty (λ) constraining the magnitude of the coefficients and (2) the mixture (α) deciding whether the model is more of a sum of squared coefficients (known as Ridge) or a sum of absolute values of the coefficients (known as Least Absolute Shrinkage and Selection Operator, LASSO). Using grid search, we chose the pair of penalty and mixture based on the lowest root mean square error (RMSE). The penalty was selected from 20 numbers, ranging from 10^{-10} to 10, equally spaced with the log_{10} scale, and the penalty was selected from 11 numbers, ranging from 0 to 1 on a linear scale.

Training the set-specific layer resulted in the predicted values of cognitive abilities, one from each set of neuroimaging features. The stacking layer, then, took these predicted values across 45 sets of neuroimaging features and treated them as features to predict cognitive abilities, thereby drawing information across (as opposed to within) sets of neuroimaging features. Importantly, we used the same training set across both layers, ensuring no data leakage between training and test sets. Opportunistic stacking dealt with missing values from each set of neuroimaging features by, first, duplicating each feature (i.e., each of 45 predicted values from the set-specific layer) into two features, resulting in 90 features. We then replaced the missing values in the duplicated features with either unrealistically large (1000) or small (-1000) values. Accordingly, we could keep the data as long as at least one set of neuroimaging features had no missing value. Using these duplicated and imputed features, we predicted cognitive abilities from different sets of neuroimaging features using Random Forest (Breiman, 2001). Ultimately, the stacking layer resulted in a predicted value of cognitive abilities based on 45 sets of neuroimaging features.

Random Forest generates several regression trees by bootstrapping observations and including a random subset of features at each split (Breiman, 2001). To make a prediction, Random Forest aggregates predicted values across bootstrapped trees, known as bagging. We used 500 trees and turned two hyperparameters. First, 'mtry' was the number of features selected at each branch. Second, 'min_n' was the minimum number of observations in a node needed for the node to be split further. Using a Latin hypercube grid search of 3,000 numbers (Dupuy et al., 2015; Sacks et al., 1989; Santner et al., n.d.), we chose the pair of mtry, ranging from 1 to 90, and min_n, ranging from 2 to 2,000, based on the lowest root mean square error (RMSE).

To understand how opportunistic stacking made predictions, we plotted Elastic Net coefficients for the set-specific layer and SHapley Additive exPlanations (SHAP) (Lundberg & Lee, 2017) for the stacking layer, averaged across 21 test sites. For the set-specific layer, Elastic Net made a prediction based on the linear summation of its regularised, estimated coefficients, and thus, plotting the coefficient of each neuroimaging feature allowed us to understand the contribution of such feature. For the stacking layer, it is difficult to trace the contribution from each feature from Random Forest directly, given the use of bagging. To overcome this, we computed Shapley values instead (Roth, 1988). Shapley values indicate the weighted differences in a model output when each feature is included versus not included in all possible subsets of features. SHAP (Lundberg & Lee, 2017) is a method to estimate Shapley values efficiently. Thus, SHAP allowed us to visualise the contribution of each set of neuroimaging features to the prediction in the stacking layer. Given that we duplicated the predicted values from each set of neuroimaging features in the stacking layer, we combined the magnitude of SHAP across the duplicates.

We fit Elastic Net and Random Forest using the *glmnet* (Friedman et al., 2010) and ranger (Wright & Ziegler, 2017) packages, respectively, with the *tidymodels* (Kuhn & Wickham, 2018/2023) package as a wrapper. We approximated the Shapley values (Lundberg & Lee, 2017) using the *fastshap* package (Greenwell, 2023). The brain plots were created via the *ggseg*, *ggsegDesterieux, ggsegJHU and ggsegGordon* packages (Mowinckel & Vidal-Piñeiro, 2020).

Predicting cognitive abilities from polygenic scores

We developed predictive models to predict cognitive abilities from polygenic scores, as reflected by PGS of cognitive abilities from three definitions (Davies et al., 2018; Lee et al., 2018; Savage et al., 2018). We first selected the PGS threshold for each of the three definitions that demonstrated the strongest correlation with cognitive abilities within the training set. This left three PGSs as features for our predictive models, one for each definition. To control for population stratification in genetics, we regressed each PGS on four genetic principal components separately for the training and test sets. Later, we treated the residuals of this regression for each PGS as each feature in our predictive models. Similar to the predictive models for the set-specific layer of the neuroimaging features, we used Elastic Net here as an algorithm. Given that the genetic data do not change over time, we used the same genetic features for baseline and follow-up predictive models. We selected participants based on ancestry for predictive models involving polygenic scores, leaving us with a much smaller number of children (n=5,776 vs. n=11,868 in the baseline).

Predicting cognitive abilities from socio-demographics, lifestyles and developmental adverse events

We developed predictive models to predict cognitive abilities from socio-demographics, lifestyles and developmental adverse events, reflected in the 44 features. We implemented partial least squares (PLS) (Wold et al., 2001) as an algorithm similar to the mental health features. To deal with missing values, we applied the following steps separately for baseline training, baseline test, follow-up training and follow-up test sets. We first imputed categorical features using mode and converted them into dummy variables. We then standardised all features and imputed them using K-nearest neighbours with five neighbours. Note that in a particular site, the value in a specific feature was at 0 for all of the observations (e.g., site 3 having a crime report at 0 for all children), making it impossible for us to standardise this feature when using this site as a test set. In this case, we kept the value of this feature at 0 and did not standardise it.

Note that the ABCD study only provided some features in the baseline, but not the follow-up. Accordingly, we treated these baseline features as features in our follow-up predictive models and combined them with the other collected in the follow-up. Supplementary Table 17 listed all of the variables and their calculation.

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 (Dadi et al., 2021). The out-of-sample relationship between observed and proxy measures of cognitive abilities based on specific 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 (Nimon et al., 2008) to demonstrate the extent to which other proxy measures captured similar variance of cognitive abilities as mental health.

First, to control for the influences of biological sex, age at interview and medication information, we residualised those variables from observed cognitive abilities and each proxy measure of cognitive abilities. We defined medication using the su_y_plus table and generated dummy variables based on the medication's functionality, as categorized by the Anatomical Therapeutic Chemical (ATC) Classification System (refer to Supplementary Table 19). We then applied random-intercept, linear-mixed models (Raudenbush & Bryk, 2002) to the data from all test sites, using the *lme4* package (Bates et al., 2015). In these models, we considered families to be nested within each site, which allow different families from each site can have an unique 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 (Nakagawa & Schielzeth, 2013; Vonesh et al., 1996) 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 (Nimon et al., 2008). We focused on the common effects between a proxy measure based on mental health and other proxy measures in four sets of commonality analyses. Note that each of the four sets of commonality analyses used different numbers of participants, depending on the data availability.

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

Here, we included proxy measures of cognitive abilities based on mental health and/or neuroimaging. Specifically, for each proxy measure, we added two regressors in the models: the values centred within each site (denoted cws) and the site average (denoted $savg$). For instance, we applied the following *lme4* syntax for the models with both proxy measures:

$$\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_{mental\ health(cws)} + \beta_2\mathbf{x}_{mental\ health(savg)} + \beta_3\mathbf{x}_{neuroimaging(cws)} + \beta_4\mathbf{x}_{neuroimaging(savg)} + (1|site{:}\,family), \quad (2)$$

We computed unique and common effects (Nimon et al., 2008) as follows:

$$Unique_{mental\ health} = R^2_{mental\ health,neuroimaging} - R^2_{neuroimaging}$$
$$Unique_{neuroimaging} = R^2_{mental\ health,neuroimaging} - R^2_{mental\ health}$$
$$Common_{mental\ health,neuroimaging} = R^2_{mental\ health,neuroimaging} - Unique_{mental\ health} - Unique_{neuroimaging}, \quad (3)$$

where the subscript of R^2 indicates which proxy measures were included in the model.

In addition to using the proxy measures based on neuroimaging from the stacking layer, we also conducted commonality analyses on proxy measures based on neuroimaging from each set of neuroimaging features. This allows us to demonstrate which sets of neuroimaging features showed higher common effects with the proxy measures based on mental health. Note that to include as many participants in the models as possible, we dropped missing values based on the availability of data in each set of neuroimaging features included in the models (i.e., not applying listwise deletion across sets of neuroimaging features).

<u>Commonality analyses for proxy measures of cognitive abilities based on mental health and polygenic scores</u>

Here, we included proxy measures of cognitive abilities based on mental health and/or polygenic scores. Since family members had more similar genetics than non-members, we changed our centring strategy to polygenic scores. With the proxy measure based on polygenic scores, we applied 1) centring on two levels: centring its values within each family first and then within each site (denoted cws, cwf) 2) averaging on two levels: averaging of its values within each family first and then within each site (denoted $savg, favg$). Accordingly, we used the following *lme4* syntax for the models with both proxy measures:

$$\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_{mental\ health(cws)} + \beta_2\mathbf{x}_{mental\ health(savg)} + \beta_3\mathbf{x}_{PGS(cws,cwf)} + \beta_4\mathbf{x}_{PGS(savg,favg)} + (1|site{:}\,family), \quad (4)$$

We computed unique and common effects as follows:

$$Unique_{mental\ health} = R^2_{mental\ health,PGS} - R^2_{PGS}$$
$$Unique_{PGS} = R^2_{mental\ health,PGS} - R^2_{mental\ health}$$
$$Common_{mental\ health,PGS} = R^2_{mental\ health,PGS} - Unique_{mental\ health} - Unique_{PGS}, \quad (5)$$

<u>Commonality analyses for proxy measures of cognitive abilities based on mental health and socio-demographics, lifestyles and developmental adverse events</u>

Here, we included proxy measures of cognitive abilities based on mental health and/or socio-demographics, lifestyles and developmental adverse events. We applied the following *lme4* syntax for the models with both proxy measures:

$$\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_{mental\ health(cws)} + \beta_2\mathbf{x}_{mental\ health(savg)} + \beta_3\mathbf{x}_{soc\ lif\ dev(cws)} + \beta_4\mathbf{x}_{soc\ lif\ dev(savg)} + (1|site{:}\,family), \quad (6)$$

Where *soc lif dev* shorts for socio-demographics, lifestyles and developmental adverse events. We computed unique and common effects (Nimon et al., 2008) as follows:

$$Unique_{mental\ health} = R^2_{mental\ health,soc\ lif\ dev} - R^2_{soc\ lif\ dev}$$
$$Unique_{soc\ lif\ dev} = R^2_{mental\ health,soc\ lif\ dev} - R^2_{mental\ health}$$
$$Common_{mental\ health,soc\ lif\ dev} = R^2_{mental\ health,soc\ lif\ dev} - Unique_{mental\ health} - Unique_{soc\ lif\ dev}, \quad (7)$$

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

Here, we included proxy measures of cognitive abilities based on mental health, neuroimaging, polygenic scores and/or socio-demographics, lifestyles and developmental adverse events. We applied the following *lme4* syntax for the model with all proxy measures included:

$$\mathbf{y} = \beta_0 + \beta_1\mathbf{x}_{mental\ health(cws)} + \beta_2\mathbf{x}_{mental\ health(savg)} + \beta_3\mathbf{x}_{neuroimaging(cws)} + \beta_4\mathbf{x}_{neuroimaging(savg)} + \beta_5\mathbf{x}_{PGS(cws,cwf)} + \beta_6\mathbf{x}_{PGS(savg,favg)} + \beta_7\mathbf{x}_{soc\ lif\ dev(cws)} + \beta_8\mathbf{x}_{soc\ lif\ dev(savg)} + (1|site{:}family), \quad (8)$$

We computed unique and common effects (Nimon et al., 2008, 2017) as follows:

$$\begin{aligned}
Unique_{mh} &= R^2_{mh,b,s,g} - R^2_{b,s,g} \\
Unique_{b} &= R^2_{mh,b,s,g} - R^2_{mh,s,g} \\
Unique_{s} &= R^2_{mh,b,s,g} - R^2_{mh,b,g} \\
Unique_{g} &= R^2_{mh,b,s,g} - R^2_{mh,b,s} \\
Common_{mh,b} &= -R^2_{s,g} + R^2_{mh,s,g} + R^2_{b,s,g} - R^2_{mh,b,s,g} \\
Common_{mh,s} &= -R^2_{b,g} + R^2_{mh,b,g} + R^2_{b,s,g} - R^2_{mh,b,s,g} \\
Common_{mh,g} &= -R^2_{b,s} + R^2_{mh,b,s} + R^2_{b,s,g} - R^2_{mh,b,s,g} \\
Common_{b,s} &= -R^2_{mh,g} + R^2_{mh,b,g} + R^2_{mh,s,g} - R^2_{mh,b,s,g} \\
Common_{b,g} &= -R^2_{mh,s} + R^2_{mh,b,s} + R^2_{mh,s,g} - R^2_{mh,b,s,g} \\
Common_{mh,b,s} &= -R^2_{g} + R^2_{mh,g} + R^2_{b,g} + R^2_{s,g} - R^2_{mh,b,g} - R^2_{mh,s,g} - R^2_{b,s,g} + R^2_{mh,b,s,g} \\
Common_{mh,b,g} &= -R^2_{s} + R^2_{mh,s} + R^2_{b,s} + R^2_{s,g} - R^2_{mh,b,s} - R^2_{mh,s,g} - R^2_{b,s,g} + R^2_{mh,b,s,g} \\
Common_{mh,s,g} &= -R^2_{b} + R^2_{mh,b} + R^2_{b,s} + R^2_{b,g} - R^2_{mh,b,s} - R^2_{mh,b,g} - R^2_{b,s,g} + R^2_{mh,b,s,g} \\
Common_{b,s,g} &= -R^2_{mh} + R^2_{mh,b} + R^2_{mh,s} + R^2_{mh,g} - R^2_{mh,b,s} - R^2_{mh,b,g} - R^2_{mh,s,g} + R^2_{mh,b,s,g} \\
Common_{mh,b,s,g} &= R^2_{mh} + R^2_{b} + R^2_{s} + R^2_{g} - R^2_{mh,b} - R^2_{mh,s} - R^2_{mh,g} - R^2_{b,s} - R^2_{b,g} - R^2_{s,g} + R^2_{mh,b,s} + R^2_{mh,b,g} + \\
&R^2_{mh,s,g} + R^2_{b,s,g} - R^2_{mh,b,s,g}, \quad (9)
\end{aligned}$$

where *mh, b, g,* and *s* denote mental health, brain (i.e., neuroimaging), genetic profile (i.e., polygenic scores) and/or socio-demographics, lifestyles and developmental adverse events, respectively.

Data and Code Sharing

We used publicly available ABCD 5.1 data (DOI: 10.15154/z563-zd24) 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-ABCD5.1.

Acknowledgements

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development[SM] (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 U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. 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/consortium_members/. 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. The ABCD data repository grows and changes over time. The ABCD data used in this report came from DOI:10.15154/z563-zd24. 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), the Neurological Foundation of New Zealand (grant number 2350 PRG) and the University of Otago.

Disclosures

The authors declare that there is no conflict of interest.

References

Abramovitch, A., Short, T., & Schweiger, A. (2021). The C Factor: Cognitive dysfunction as a transdiagnostic dimension in psychopathology. *Clinical Psychology Review*, *86*, 102007. https://doi.org/10.1016/j.cpr.2021.102007

Achenbach, T. M., Ivanova, M. Y., & Rescorla, L. A. (2017). Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. *Comprehensive Psychiatry*, *79*, 4–18. https://doi.org/10.1016/j.comppsych.2017.03.006

Alexander, A. L., Lee, J. E., Lazar, M., & Field, A. S. (2007). Diffusion tensor imaging of the brain. *Neurotherapeutics*, *4*(3), 316–329. https://doi.org/10.1016/j.nurt.2007.05.011

Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry*, *24*(6), 819–827. https://doi.org/10.1038/s41380-019-0394-4

Ang, Y.-S., Frontero, N., Belleau, E., & Pizzagalli, D. A. (2020). Disentangling vulnerability, state and trait features of neurocognitive impairments in depression. *Brain*, *143*(12), 3865–3877. https://doi.org/10.1093/brain/awaa314

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., … National Eye Institute, N. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), Article 7571. https://doi.org/10.1038/nature15393

Bagot, K. S., Matthews, S. A., Mason, M., Squeglia, L. M., Fowler, J., Gray, K., Herting, M., May, A., Colrain, I., Godino, J., Tapert, S., Brown, S., & Patrick, K. (2018). Current, future and potential use of mobile and wearable technologies and social media data in the ABCD study to increase understanding of contributors to child health. *Developmental Cognitive Neuroscience*, *32*, 121–129. https://doi.org/10.1016/j.dcn.2018.03.008

Baker, D. P., Eslinger, P. J., Benavides, M., Peters, E., Dieckmann, N. F., & Leon, J. (2015). The cognitive impact of the education revolution: A possible cause of the Flynn Effect on population IQ. *Intelligence*, *49*, 144–158. https://doi.org/10.1016/j.intell.2015.01.003

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., … WU-Minn HCP Consortium. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189. https://doi.org/10.1016/j.neuroimage.2013.05.033

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bauer, P. J., Dikmen, S. S., Heaton, R. K., Mungas, D., Slotkin, J., & Beaumont, J. L. (2013). III. NIH TOOLBOX COGNITION BATTERY (CB): MEASURING EPISODIC MEMORY. *Monographs of the Society for Research in Child Development*, *78*(4), 34–48. https://doi.org/10.1111/mono.12033

Beam, E., Potts, C., Poldrack, R. A., & Etkin, A. (2021). A data-driven framework for mapping domains of human neurobiology. *Nature Neuroscience*, *24*(12), 1733–1744. https://doi.org/10.1038/s41593-021-00948-9

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature*, *604*(7906), 525–533. https://doi.org/10.1038/s41586-022-04554-y

Bissett, P. G., Hagen, M. P., Jones, H. M., & Poldrack, R. A. (2021). Design issues and solutions for stop-signal data from the Adolescent Brain Cognitive Development (ABCD) study. *eLife*, *10*, e60185. https://doi.org/10.7554/eLife.60185

Bleck, T. P., Nowinski, C. J., Gershon, R., & Koroshetz, W. J. (2013). What is the NIH Toolbox, and what will it mean to neurology? *Neurology*, *80*(10), 874–875. https://doi.org/10.1212/WNL.0b013e3182872ea0

Bogdan, R., Baranger, D. A. A., & Agrawal, A. (2018). Polygenic Risk Scores in Clinical Psychology: Bridging Genomic Risk to Individual Differences. *Annual Review of Clinical Psychology*, *14*(1), 119–157. https://doi.org/10.1146/annurev-clinpsy-050817-084847

Bolt, T., Nomi, J. S., Yeo, B. T. T., & Uddin, L. Q. (2017). Data-Driven Extraction of a Nested Model of Human Brain Function. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(30), 7263–7277. https://doi.org/10.1523/JNEUROSCI.0323-17.2017

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bruni, O., Ottaviano, S., Guidetti, V., Romoli, M., Innocenzi, M., Cortesi, F., & Giannotti, F. (1996). The Sleep Disturbance Scale for Children (SDSC) Construct ion and validation of an instrument to evaluate sleep disturbances in childhood and adolescence. *Journal of Sleep Research*, *5*(4), 251–261. https://doi.org/10.1111/j.1365-2869.1996.00251.x

Calvin, C. M., Batty, G. D., Der, G., Brett, C. E., Taylor, A., Pattie, A., Čukić, I., & Deary, I. J. (2017). Childhood intelligence in relation to major causes of death in 68 year follow-up: Prospective population study. *BMJ*, j2708. https://doi.org/10.1136/bmj.j2708

Carlozzi, N. E., Tulsky, D. S., Kail, R. V., & Beaumont, J. L. (2013). VI. NIH TOOLBOX COGNITION BATTERY (CB): MEASURING PROCESSING SPEED. *Monographs of the Society for Research in Child Development*, *78*(4), 88–102. https://doi.org/10.1111/mono.12036

Carver, C. S., & Johnson, S. L. (2018). Impulsive reactivity to emotion and vulnerability to psychopathology. *American Psychologist*, *73*(9), 1067–1078. https://doi.org/10.1037/amp0000387

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., … ABCD Imaging Acquisition Workgroup. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Caspi, A., & Moffitt, T. E. (2018). All for One and One for All: Mental Disorders in One Dimension. *American Journal of Psychiatry*, *175*(9), 831–844. https://doi.org/10.1176/appi.ajp.2018.17121383

Centers for Disease Control and Prevention. (2023, May 2). *Youth Risk Behavior Surveillance System (YRBSS) | CDC*. https://www.cdc.gov/healthyyouth/data/yrbs/index.htm

Chilcoat, H. D., & Anthony, J. C. (1996). Impact of Parent Monitoring on Initiation of Drug Use through Late Childhood. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(1), 91–100. https://doi.org/10.1097/00004583-199601000-00017

Choi, S. W., Mak, T. S.-H., & O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), Article 9. https://doi.org/10.1038/s41596-020-0353-1

Clark, D. B., Fisher, C. B., Bookheimer, S., Brown, S. A., Evans, J. H., Hopfer, C., Hudziak, J., Montoya, I., Murray, M., Pfefferbaum, A., & Yurgelun-Todd, D. (2018). Biomedical ethics and clinical oversight in multisite observational neuroimaging studies with children and adolescents: The ABCD experience. *Developmental Cognitive Neuroscience*, *32*, 143–154. https://doi.org/10.1016/j.dcn.2017.06.005

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

Dadi, K., Varoquaux, G., Houenou, J., Bzdok, D., Thirion, B., & Engemann, D. (2021). Population modeling with machine learning can enhance measures of mental health. *GigaScience*, *10*(10), giab071. https://doi.org/10.1093/gigascience/giab071

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Daniel, M. H., & Wahlstrom, D. (2014). *Equivalence of Q-interactive™ and Paper Administrations of Cognitive Tasks: WISC®–V*.

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., Hagenaars, S. P., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Liewald, D. C. M., Okely, J. A., Ahola-Olli, A. V., Barnes, C. L. K., Bertram, L., Bis, J. C., Burdick, K. E., Christoforou, A., DeRosse, P., … Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-04362-x

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*(1), 453–482. https://doi.org/10.1146/annurev-psych-120710-100353

Deary, I. J., Pattie, A., & Starr, J. M. (2013). The Stability of Intelligence From Age 11 to Age 90 Years: The Lothian Birth Cohort of 1921. *Psychological Science*, *24*(12), 2361–2368. https://doi.org/10.1177/0956797613486487

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Dick, A. S., Garcia, N. L., Pruden, S. M., Thompson, W. K., Hawes, S. W., Sutherland, M. T., Riedel, M. C., Laird, A. R., & Gonzalez, R. (2019). No evidence for a bilingual executive function advantage in the ABCD study. *Nature Human Behaviour*, *3*(7), Article 7. https://doi.org/10.1038/s41562-019-0609-3

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756), 20170284. https://doi.org/10.1098/rstb.2017.0284

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-11112-0

Dupuy, D., Helbert, C., & Franco, J. (2015). DiceDesign and DiceEval: Two R Packages for Design and Analysis of Computer Experiments. *Journal of Statistical Software*, *65*, 1–38. https://doi.org/10.18637/jss.v065.i11

Duyme, M., Dumaret, A.-C., & Tomkiewicz, S. (1999). How can we boost IQs of "dull children"?: A late adoption study. *Proceedings of the National Academy of Sciences*, *96*(15), 8790–8794. https://doi.org/10.1073/pnas.96.15.8790

East-Richard, C., R. -Mercier, A., Nadeau, D., & Cellard, C. (2020). Transdiagnostic neurocognitive deficits in psychiatry: A review of meta-analyses. *Canadian Psychology / Psychologie Canadienne*, *61*(3), 190–214. https://doi.org/10.1037/cap0000196

Echeverria, S. E., Diez-Roux, A. V., & Link, B. G. (2004). Reliability of self-reported neighborhood characteristics. *Journal of Urban Health*, *81*(4), 682–701. https://doi.org/10.1093/jurban/jth151

Engemann, D. A., Kozynets, O., Sabbagh, D., Lemaître, G., Varoquaux, G., Liem, F., & Gramfort, A. (2020). Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *eLife*, *9*, e54055. https://doi.org/10.7554/eLife.54055

Epskamp, S. (2015). semPlot: Unified Visualizations of Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 474–483. https://doi.org/10.1080/10705511.2014.937847

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Fassbender, C., Mukherjee, P., & Schweitzer, J. B. (2017). Minimizing noise in pediatric task-based functional MRI; Adolescents with developmental disabilities and typical development. *NeuroImage*, *149*, 338–347. https://doi.org/10.1016/j.neuroimage.2017.01.021

Federal Bureau Of Investigation. (2012). *Uniform Crime Reporting Program Data: County-Level Detailed Arrest and Offense Data, 2010: Version 2* [Dataset]. ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/ICPSR33523.V2

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., & Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/s0896-6273(02)00569-x

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*(1), 29–51. https://doi.org/10.1037/0033-2909.95.1.29

Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn effect* (1st expanded pbk. ed). Cambridge University Press.

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., Roalf, D. R., Satterthwaite, T. D., Gur, R. C., Gur, R. E., Schultz, R. T., Verma, R., & Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. https://doi.org/10.1016/j.neuroimage.2017.08.047

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22.

Frostenson, S., April 6, S. K. on, 2016, & ET, 8:50 a m. (n.d.). *Where is the lead exposure risk in your community?* Vox.Com. Retrieved June 27, 2023, from http://www.vox.com/a/lead-exposure-risk-map

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*, 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society: JINS*, *20*(6), 642–651. https://doi.org/10.1017/S1355617714000411

Goodman, R., Meltzer, H., & Bailey, V. (2003). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *International Review of Psychiatry (Abingdon, England)*, *15*(1–2), 173–177. https://doi.org/10.1080/0954026021000046137

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, *26*(1), 288–303. https://doi.org/10.1093/cercor/bhu239

Greenwell, B. (2023). *fastshap: Fast Approximate Shapley Values* (Version 0.1.0) [Computer software]. https://cran.r-project.org/web/packages/fastshap/index.html

Hagler, D. J., Ahmadi, M. E., Kuperman, J., Holland, D., McDonald, C. R., Halgren, E., & Dale, A. M. (2009). Automated white-matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy. *Human Brain Mapping*, *30*(5), 1535–1547. https://doi.org/10.1002/hbm.20619

Hagler, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., Sutherland, M. T., Casey, B. J., Barch, D. M., Harms, M. P., Watts, R., Bjork, J. M., Garavan, H. P., Hilmer, L., Pung, C. J., Sicat, C. S., Kuperman, J., Bartsch, H., Xue, F., … Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage*, *202*, 116091. https://doi.org/10.1016/j.neuroimage.2019.116091

Hankin, B. L., Snyder, H. R., & Gulley, L. D. (2016). Cognitive Risks in Developmental Psychopathology. In D. Cicchetti (Ed.), *Developmental Psychopathology* (1st ed., pp. 1–74). Wiley. https://doi.org/10.1002/9781119125556.devpsy308

Hartshorne, J. K., & Germine, L. T. (2015). When Does Cognitive Functioning Peak? The Asynchronous Rise and Fall of Different Cognitive Abilities Across the Life Span. *Psychological Science*, *26*(4), 433–443. https://doi.org/10.1177/0956797614567339

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang, P. (2010). Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Johnson, S. L., Turner, R. J., & Iwata, N. (2003). BIS/BAS Levels and Psychiatric Disorder: An Epidemiological Study. *Journal of Psychopathology and Behavioral Assessment*, *25*(1), 25–36. https://doi.org/10.1023/A:1022247919288

Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)*, *8*(1), 118–127. https://doi.org/10.1093/biostatistics/kxj037

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., Garnier-Villarreal, M., Selig, J., Boulton, A., Preacher, K., Coffman, D., Rhemtulla, M., Robitzsch, A., Enders, C., Arslan, R., Clinton, B., Panko, P., Merkle, E., Chesnut, S., … Johnson, A. R. (2022). *semTools: Useful Tools for Structural Equation Modeling* (Version 0.5-6) [Computer software]. https://cran.r-project.org/web/packages/semTools/index.html

Keller, A. S., Pines, A. R., Shanmugan, S., Sydnor, V. J., Cui, Z., Bertolero, M. A., Barzilay, R., Alexander-Bloch, A. F., Byington, N., Chen, A., Conan, G. M., Davatzikos, C., Feczko, E., Hendrickson, T. J., Houghton, A., Larsen, B., Li, H., Miranda-Dominguez, O., Roalf, D. R., … Satterthwaite, T. D. (2023). Personalized functional brain network topography is associated with individual differences in youth cognition. *Nature Communications*, *14*(1), 8411. https://doi.org/10.1038/s41467-023-44087-0

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Üstün, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Current Opinion in Psychiatry*, *20*(4). https://journals.lww.com/co-psychiatry/fulltext/2007/07000/age_of_onset_of_mental_disorders__a_review_of.10.aspx

Kessler, R. C., Avenevoli, S., Costello, E. J., Green, J. G., Gruber, M. J., Heeringa, S., Merikangas, K. R., Pennell, B.-E., Sampson, N. A., & Zaslavsky, A. M. (2009). Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *International Journal of Methods in Psychiatric Research*, *18*(2), 69–83. https://doi.org/10.1002/mpr.279

Kind, A. J. H., Jencks, S., Brock, J., Yu, M., Bartels, C., Ehlenbach, W., Greenberg, C., & Smith, M. (2014). Neighborhood Socioeconomic Disadvantage and 30-Day Rehospitalization. *Annals of Internal Medicine*, *161*(11), 765–774. https://doi.org/10.7326/M13-2946

Kirlic, N., Colaizzi, J. M., Cosgrove, K. T., Cohen, Z. P., Yeh, H.-W., Breslin, F., Morris, A. S., Aupperle, R. L., Singh, M. K., & Paulus, M. P. (2021). Extracurricular Activities, Screen Media Activity, and Sleep May Be Modifiable Factors Related to Children's Cognitive Functioning: Evidence From the ABCD Study®. *Child Development*. https://doi.org/10.1111/cdev.13578

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, *12*(1), 20–27. https://doi.org/10.1006/nimg.2000.0593

Kuhn, M., & Wickham, H. (2023). *Tidymodels* [R]. tidymodels. https://github.com/tidymodels/tidymodels (Original work published 2018)

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), Article 8. https://doi.org/10.1038/s41588-018-0147-3

Lencz, T., Knowles, E., Davies, G., Guha, S., Liewald, D. C., Starr, J. M., Djurovic, S., Melle, I., Sundet, K., Christoforou, A., Reinvang, I., Mukherjee, S., DeRosse, P., Lundervold, A., Steen, V. M., John, M., Espeseth, T., Räikkönen, K., Widen, E., … Malhotra, A. K. (2014). Molecular genetic evidence for overlap between general cognitive ability and risk for schizophrenia: A report from the Cognitive Genomics consorTium (COGENT). *Molecular Psychiatry*, *19*(2), 168–174. https://doi.org/10.1038/mp.2013.166

Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience*, *32*, 67–79. https://doi.org/10.1016/j.dcn.2018.02.006

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Makowski, C., Brown, T. T., Zhao, W., Hagler Jr, D. J., Parekh, P., Garavan, H., Nichols, T. E., Jernigan, T. L., & Dale, A. M. (2024). Leveraging the adolescent brain cognitive development study to improve behavioral prediction from neuroimaging in smaller replication samples. *Cerebral Cortex*, *34*(6), bhae223. https://doi.org/10.1093/cercor/bhae223

Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., … Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, *603*(7902), 654–660. https://doi.org/10.1038/s41586-022-04492-9

Merikangas, K. R., Avenevoli, S., Costello, E. J., Koretz, D., & Kessler, R. C. (2009). National Comorbidity Survey Replication Adolescent Supplement (NCS-A): I. Background and Measures. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(4), 367–379. https://doi.org/10.1097/CHI.0b013e31819996f1

Michelini, G., Barch, D. M., Tian, Y., Watson, D., Klein, D. N., & Kotov, R. (2019). Delineating and validating higher-order dimensions of psychopathology in the Adolescent Brain Cognitive Development (ABCD) study. *Translational Psychiatry*, *9*(1), 261. https://doi.org/10.1038/s41398-019-0593-4

Mihalik, A., Brudfors, M., Robu, M., Ferreira, F. S., Lin, H., Rau, A., Wu, T., Blumberg, S. B., Kanber, B., Tariq, M., Garcia, M. E., Zor, C., Nikitichev, D. I., Mourão-Miranda, J., & Oxtoby, N. P. (2019). ABCD Neurocognitive Prediction Challenge 2019: Predicting Individual Fluid Intelligence Scores from Structural MRI Using Probabilistic Segmentation and Kernel Ridge Regression. In K. M. Pohl, W. K. Thompson, E. Adeli, & M. G. Linguraru (Eds.), *Adolescent Brain Cognitive Development Neurocognitive Prediction* (pp. 133–142). Springer International Publishing. https://doi.org/10.1007/978-3-030-31901-4_16

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., Connor, R., Davis, S., Deakin, B., DeRubeis, R. J., Dubois, B., Geyer, M. A., Goodwin, G. M., Gorwood, P., Jay, T. M., Joëls, M., Mansuy, I. M., Meyer-Lindenberg, A., Murphy, D., … Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: Characteristics, causes and the quest for improved therapy. *Nature Reviews Drug Discovery*, *11*(2), 141–168. https://doi.org/10.1038/nrd3628

Molnar, C. (2019). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/

Moos, R. H., Insel, P. M., & Humphrey, B. (1974). *Preliminary manual for family environment scale, work environment scale, group environment scale*. Consulting Psychologists Press. http://books.google.com/books?id=1ihHAAAAMAAJ

Morris, S. E., & Cuthbert, B. N. (2012). Research Domain Criteria: Cognitive systems, neural circuits, and dimensions of behavior. *Dialogues in Clinical Neuroscience*, *14*(1), 29–37.

Morris, S. E., Sanislow, C. A., Pacheco, J., Vaidyanathan, U., Gordon, J. A., & Cuthbert, B. N. (2022). Revisiting the seven pillars of RDoC. *BMC Medicine*, *20*(1), 220. https://doi.org/10.1186/s12916-022-02414-0

Mowinckel, A. M., & Vidal-Piñeiro, D. (2020). Visualization of Brain Statistics With R Packages ggseg and ggseg3d. *Advances in Methods and Practices in Psychological Science*, *3*(4), 466–483. https://doi.org/10.1177/2515245920928009

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nielson, D., Pereira, F., Zheng, C., Migineishvili, N., Lee, J., Thomas, A., & Bandettini, P. (2018). *Detecting and harmonizing scanner differences in the ABCD study—Annual release 1.0*. https://doi.org/10.1101/309260

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2008). An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *40*(2), 457–466. https://doi.org/10.3758/BRM.40.2.457

Nimon, K., Lewis, M., Kane, R., & Haynes, R. M. (2017). Erratum to: An R package to compute commonality coefficients in the multiple regression case: An introduction to the package and a practical example. *Behavior Research Methods*, *49*(6), 2275–2275. https://doi.org/10.3758/s13428-017-0853-2

Pat, N., Riglin, L., Anney, R., Wang, Y., Barch, D. M., Thapar, A., & Stringaris, A. (2022). Motivation and Cognitive Abilities as Mediators Between Polygenic Scores and Psychopathology in Children. *Journal of the American Academy of Child and Adolescent Psychiatry*, *61*(6), 782-795.e3. https://doi.org/10.1016/j.jaac.2021.08.019

Pat, N., Wang, Y., Anney, R., Riglin, L., Thapar, A., & Stringaris, A. (2022). Longitudinally stable, brain-based predictive models mediate the relationships between childhood cognition and socio-demographic, psychological and genetic factors. *Human Brain Mapping*, *43*(18), 5520–5542. https://doi.org/10.1002/hbm.26027

Pat, N., Wang, Y., Bartonicek, A., Candia, J., & Stringaris, A. (2023). Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cerebral Cortex*, *33*(6), 2682–2703. https://doi.org/10.1093/cercor/bhac235

Pietschnig, J., & Voracek, M. (2015). One Century of Global IQ Gains: A Formal Meta-Analysis of the Flynn Effect (1909–2013). *Perspectives on Psychological Science*, *10*(3), 282–306. https://doi.org/10.1177/1745691615577701

Psaty, B. M., O'Donnell, C. J., Gudnason, V., Lunetta, K. L., Folsom, A. R., Rotter, J. I., Uitterlinden, A. G., Harris, T. B., Witteman, J. C. M., & Boerwinkle, E. (2009). Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of Prospective Meta-Analyses of Genome-Wide Association Studies From 5 Cohorts. *Circulation: Cardiovascular Genetics*, *2*(1), 73–80. https://doi.org/10.1161/CIRCGENETICS.108.829747

Quah, S. K. L., Jo, B., Geniesse, C., Uddin, L. Q., Mumford, J. A., Barch, D. M., Fair, D. A., Gotlib, I. H., Poldrack, R. A., & Saggar, M. (2025). A data-driven latent variable approach to validating the research domain criteria framework. *Nature Communications*, *16*(1), 830. https://doi.org/10.1038/s41467-025-55831-z

Rasero, J., Sentis, A. I., Yeh, F.-C., & Verstynen, T. (2021). Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. *PLOS Computational Biology*, *17*(3), e1008347. https://doi.org/10.1371/journal.pcbi.1008347

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed). Sage Publications.

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F., & van der Ende, J. (2010). Predicting adult emotional and behavioral problems from externalizing problem trajectories in a 24-year longitudinal study. *European Child & Adolescent Psychiatry*, *19*(7), 577–585. https://doi.org/10.1007/s00787-010-0088-6

Rindermann, H., Becker, D., & Coyle, T. R. (2017). Survey of expert opinion on intelligence: The FLynn effect and the future of intelligence. *Personality and Individual Differences*, *106*, 242–247. https://doi.org/10.1016/j.paid.2016.10.061

Rohart, F., Gautier, B., Singh, A., & Cao, K.-A. L. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Roth, A. E. (Ed.). (1988). *The Shapley Value: Essays in Honor of Lloyd S. Shapley*. Cambridge University Press. https://doi.org/10.1017/CBO9780511528446

Roza, S. J., Hofstra, M. B., Van Der Ende, J., & Verhulst, F. C. (2003). Stable Prediction of Mood and Anxiety Disorders Based on Behavioral and Emotional Problems in Childhood: A 14-Year Follow-Up During Childhood, Adolescence, and Young Adulthood. *American Journal of Psychiatry*, *160*(12), 2116–2121. https://doi.org/10.1176/appi.ajp.160.12.2116

Rundquist, E. A. (1936). Intelligence test scores and school marks of high school seniors in 1929 and 1933. *School & Society*, *43*, 301–304.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and Analysis of Computer Experiments. *Statistical Science*, *4*(4), 409–423. https://doi.org/10.1214/ss/1177012413

Santner, T. J., Williams, B. J., & Notz, W. I. (n.d.). *The Design and Analysis of Computer Experiments*.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), Article 7. https://doi.org/10.1038/s41588-018-0152-6

Schaefer, E. S. (1965). A configurational analysis of children's reports of parent behavior. *Journal of Consulting Psychology*, *29*(6), 552–557. https://doi.org/10.1037/h0022702

Snellen, H. (1862). *Letterproeven, tot bepaling der gezigtsscherpte*. J. Greven.

Sripada, C., Angstadt, M., Rutherford, S., Taxali, A., & Shedden, K. (2020). Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping*, *41*(12), 3186–3197. https://doi.org/10.1002/hbm.25007

Sripada, C., Angstadt, M., Taxali, A., Clark, D. A., Greathouse, T., Rutherford, S., Dickens, J. R., Shedden, K., Gard, A. M., Hyde, L. W., Weigard, A., & Heitzeg, M. (2021). Brain-wide functional connectivity patterns support general cognitive ability and mediate effects of socioeconomic status in youth. *Translational Psychiatry*, *11*(1), 1–8. https://doi.org/10.1038/s41398-021-01704-0

Stover, P. J., Harlan, W. R., Hammond, J. A., Hendershot, T., & Hamilton, C. M. (2010). PhenX: A toolkit for interdisciplinary genetics research. *Current Opinion in Lipidology*, *21*(2), 136. https://doi.org/10.1097/MOL.0b013e3283377395

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biological Psychiatry*, *88*(11), 818–828. https://doi.org/10.1016/j.biopsych.2020.02.016

Tetereva, A., Li, J., Deng, J. D., Stringaris, A., & Pat, N. (2022). Capturing brain-cognition relationship: Integrating task-based fMRI across tasks markedly boosts prediction and test-retest reliability. *NeuroImage*, *263*, 119588. https://doi.org/10.1016/j.neuroimage.2022.119588

Tetereva, A., & Pat, N. (2024). Brain age has limited utility as a biomarker for capturing fluid cognition in older individuals. *eLife*, *12*, RP87297. https://doi.org/10.7554/eLife.87297.4

Thompson, W. K., Barch, D. M., Bjork, J. M., Gonzalez, R., Nagel, B. J., Nixon, S. J., & Luciana, M. (2019). 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. *Developmental Cognitive Neuroscience*, *36*, 100606. https://doi.org/10.1016/j.dcn.2018.12.004

Trahan, L. H., Stuebing, K. K., Fletcher, J. M., & Hiscock, M. (2014). The Flynn effect: A meta-analysis. *Psychological Bulletin*, *140*(5), 1332–1360. https://doi.org/10.1037/a0037173

Uban, K. A., Horton, M. K., Jacobus, J., Heyser, C., Thompson, W. K., Tapert, S. F., Madden, P. A. F., & Sowell, E. R. (2018). Biospecimens and the ABCD study: Rationale, methods of collection, measurement and early data. *Developmental Cognitive Neuroscience*, *32*, 97–106. https://doi.org/10.1016/j.dcn.2018.03.005

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. *NeuroImage*, *80*, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041

Vieira, B. H., Pamplona, G. S. P., Fachinello, K., Silva, A. K., Foss, M. P., & Salmon, C. E. G. (2022). On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. *Intelligence*, *93*, 101654. https://doi.org/10.1016/j.intell.2022.101654

Vonesh, E. F., Chinchilli, V. M., & Pu, K. (1996). Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models. *Biometrics*, *52*(2), 572–587. https://doi.org/10.2307/2532896

Wainschtein, P., Jain, D., Zheng, Z., Aslibekyan, S., Becker, D., Bi, W., Brody, J., Carlson, J. C., Correa, A., Du, M. M., Fernandez-Rhodes, L., Ferrier, K. R., Graff, M., Guo, X., He, J., Heard-Costa, N. L., Highland, H. M., Hirschhorn, J. N., Howard-Claudio, C. M., … NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium. (2022). Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nature Genetics*, *54*(3), 263–273. https://doi.org/10.1038/s41588-021-00997-7

Whelan, R., Conrod, P. J., Poline, J.-B., Lourdusamy, A., Banaschewski, T., Barker, G. J., Bellgrove, M. A., Büchel, C., Byrne, M., Cummins, T. D. R., Fauth-Bühler, M., Flor, H., Gallinat, J., Heinz, A., Ittermann, B., Mann, K., Martinot, J.-L., Lalor, E. C., Lathrop, M., … Garavan, H. (2012). Adolescent impulsivity phenotypes characterized by distinct brain networks. *Nature Neuroscience*, *15*(6), Article 6. https://doi.org/10.1038/nn.3092

Whiteside, S. P., & Lynam, D. R. (2003). Understanding the role of impulsivity and externalizing psychopathology in alcohol abuse: Application of the UPPS Impulsive Behavior Scale. *Experimental and Clinical Psychopharmacology*, *11*(3), 210–217. https://doi.org/10.1037/1064-1297.11.3.210

Williams, R. L. (2013). Overview of the Flynn effect. *Intelligence*, *41*(6), 753–764. https://doi.org/10.1016/j.intell.2013.04.010

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1

Wongupparaj, P., Wongupparaj, R., Kumari, V., & Morris, R. G. (2017). The Flynn effect for verbal and visuospatial short-term and working memory: A cross-temporal meta-analysis. *Intelligence*, *64*, 71–80. https://doi.org/10.1016/j.intell.2017.07.006

Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*, 1–17. https://doi.org/10.18637/jss.v077.i01

Yang, R., & Jernigan, Terry. (n.d.). *Adolescent Brain Cognitive Development Study (ABCD)—Annual Release 4.0* [Dataset]. NIMH Data Repositories. https://doi.org/10.15154/1523041

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Zapolski, T. C. B., Stairs, A. M., Settles, R. F., Combs, J. L., & Smith, G. T. (2010). The Measurement of Dispositions to Rash Action in Children. *Assessment*, *17*(1), 116–125. https://doi.org/10.1177/1073191109351372

Zhao, W., Makowski, C., Hagler, D. J., Garavan, H. P., Thompson, W. K., Greene, D. J., Jernigan, T. L., & Dale, A. M. (2023). Task fMRI paradigms may capture more behaviorally relevant information than resting-state functional connectivity. *NeuroImage*, *270*, 119946. https://doi.org/10.1016/j.neuroimage.2023.119946

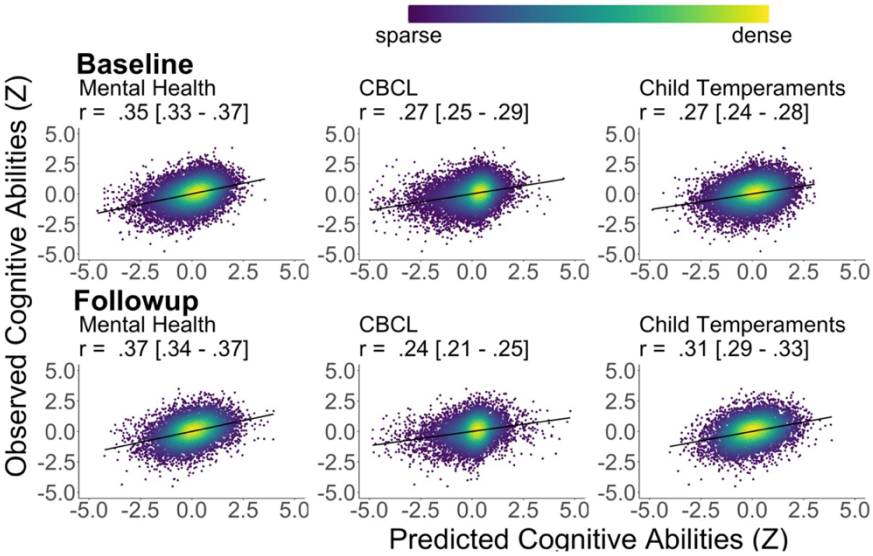
Zhi, D., Jiang, R., Pearlson, G., Fu, Z., Qi, S., Yan, W., Feng, A., Xu, M., Calhoun, V., & Sui, J. (2024). Triple Interactions Between the Environment, Brain, and Behavior in Children: An ABCD Study. *Developmental Circuit Alterations in Autism and Attention-Deficit/Hyperactivity Disorder*, *95*(9), 828–838. https://doi.org/10.1016/j.biopsych.2023.12.019

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *67*(2), 301–320.

Zucker, R. A., Gonzalez, R., Feldstein Ewing, S. W., Paulus, M. P., Arroyo, J., Fuligni, A., Morris, A. S., Sanchez, M., & Wills, T. (2018). Assessment of culture and environment in the Adolescent Brain and Cognitive Development Study: Rationale, description of measures, and early data. *Developmental Cognitive Neuroscience*, *32*, 107–120. https://doi.org/10.1016/j.dcn.2018.03.004

Figures

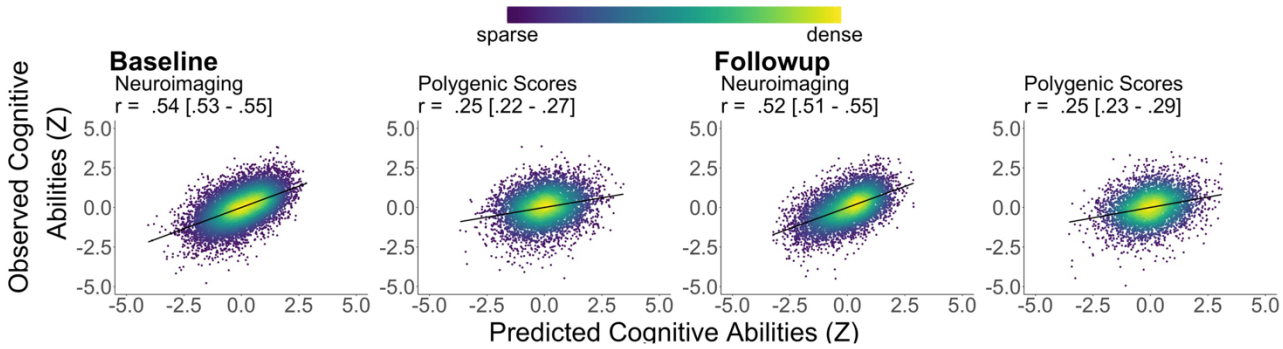
a) Performance of Mental Health in Predicting Cognitive Abilities



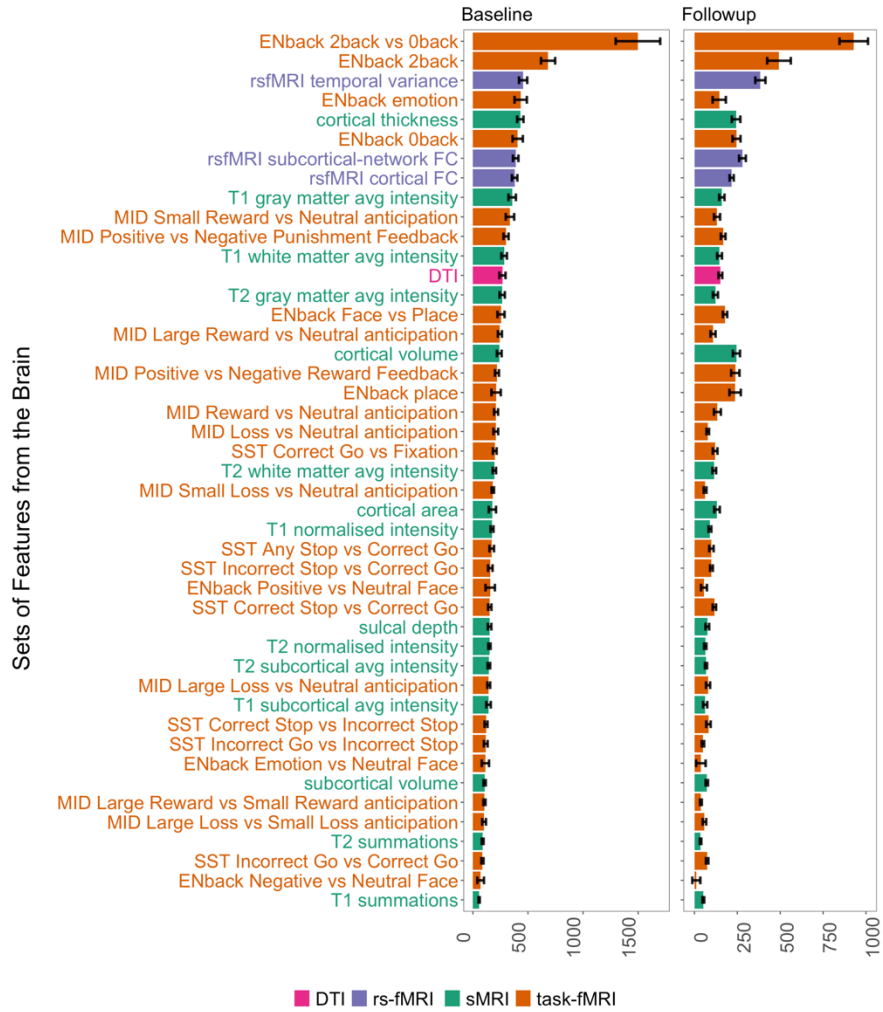
b) Feature Importance of Mental Health in Predicting Cognitive Abilities

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. Cognitive abilities are based on the second-order latent variable, the g-factor, based on a confirmatory factor analysis of six cognitive tasks. All data points are from test sets. *r* is the average Pearson's *r* across 21 test sites. The parentheses following the *r* indicate bootstrapped 95%CIs, calculated based on observed vs predicted cognitive abilities from all test sites combined. UPPS-P Impulsive and Behaviour Scale and the Behavioural Inhibition System/Behavioural Activation System (BIS/BAS) were used for child temperaments, conceptualised as risk factors for mental issues. Mental health includes features from CBCL and child temperaments. b) Feature importance of mental health, predicting cognitive abilities via PLS. 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 (in green), reflecting children's emotional and behavioural problems; UPPS-P = Urgency, Premeditation, Perseverance, Sensation seeking and Positive urgency Impulsive Behaviour Scale; BAS = Behavioural Activation System (in orange).

a) Performance of Neuroimaging Measures and Polygenic Scores in Predicting Cognitive Abilities



b) Feature Importance of Neuroimaging Measures in Predicting Cognitive Abilities



c) Feature Importance of Polygenic Scores in Predicting Cognitive Abilities

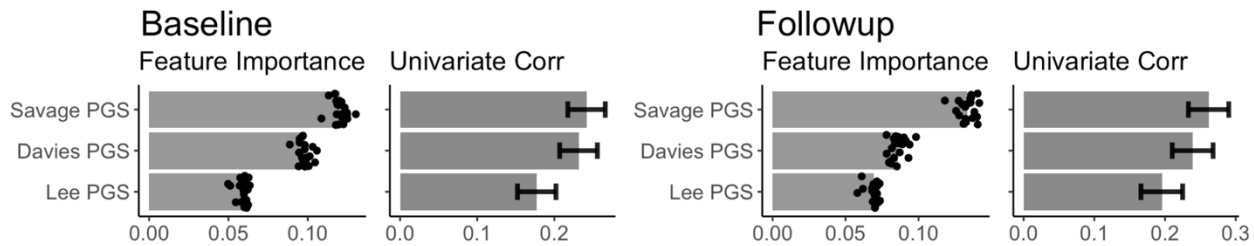


Figure 2. Predictive models predicting cognitive abilities from neuroimaging via opportunistic stacking and polygenic scores via Elastic Net. a) Scatter plots between observed vs predicted cognitive abilities based on neuroimaging and polygenic scores. Cognitive abilities are based on the second-order latent variable, the g-factor, based on a confirmatory factor analysis of six cognitive tasks. The parentheses following the *r* indicate the bootstrapped 95%CIs, calculated based on observed vs predicted cognitive abilities from all test sites combined. All data points are from test sets. *r* is the average Pearson's *r* across 21 test sites. The parentheses following the *r* indicate bootstrapped 95%CIs, calculated based on observed vs predicted cognitive abilities from all test sites combined. 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. Different sets of neuroimaging features were filled with different colours: pink for dMRI, orange for fMRI, purple for rsMRI and green for sMRI. 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.

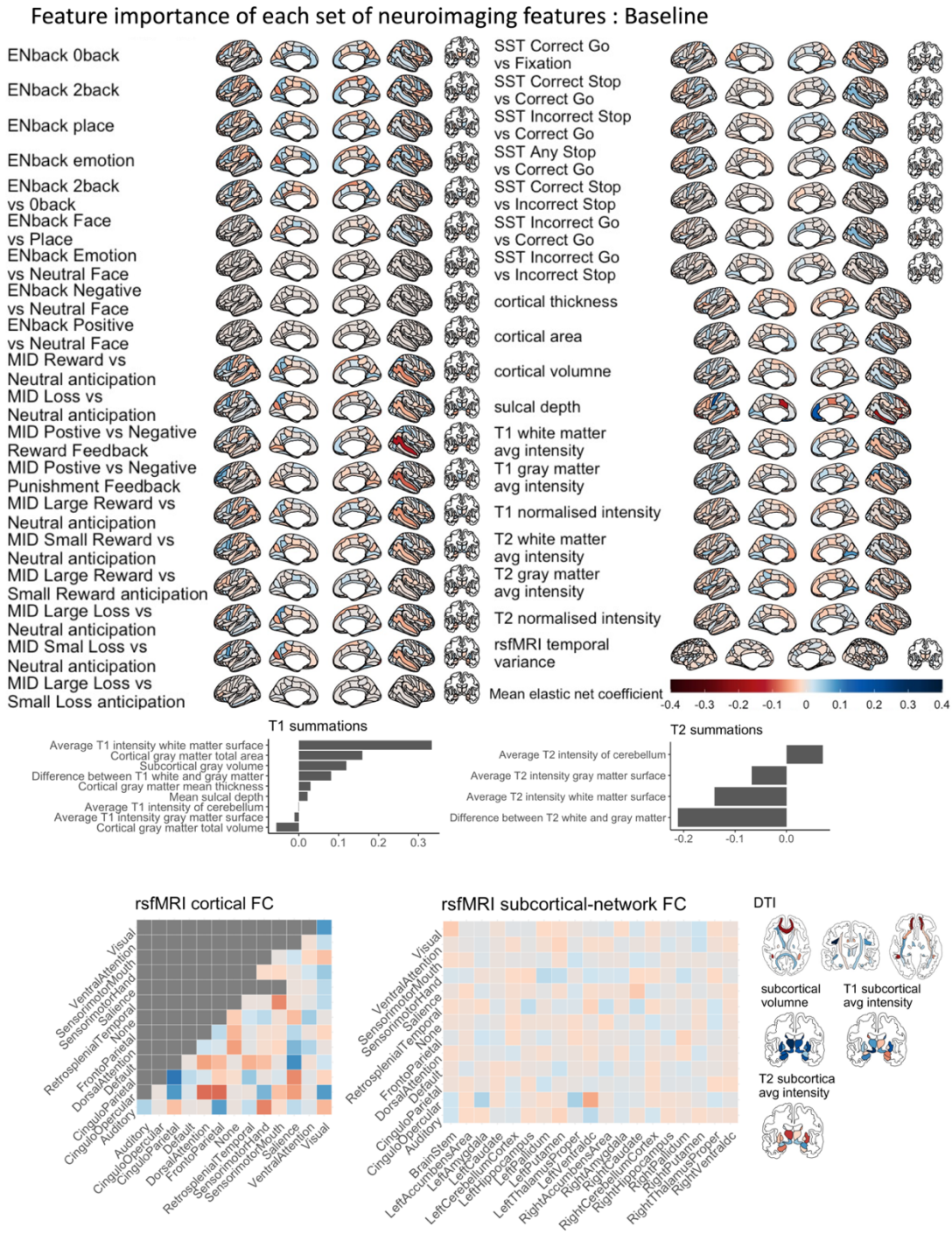
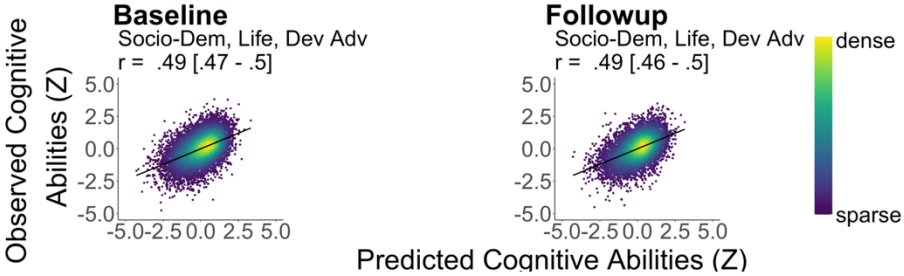


Figure 3. Feature importance of each set of neuroimaging features, predicting cognitive abilities in the baseline data. The feature importance was based on the Elastic Net coefficients, averaged across test sites. We did not order these sets of neuroimaging features according to their contribution to the stacking layer (see Figure 2). Larger versions of the feature importance for each set of neuroimaging features can be found in Supplementary Figures 4 - 13. MID = Monetary Incentive Delay task; SST = Stop Signal Task; DTI = Diffusion Tensor Imaging; FC = functional connectivity.

a) Performance of Socio-Demographics, Lifestyle and Developmental Adverse Events in Predicting Cognitive Abilities



b) Feature Importance of Socio-Demographics, Lifestyle and Developmental Adverse Events in Predicting Cognitive Abilities

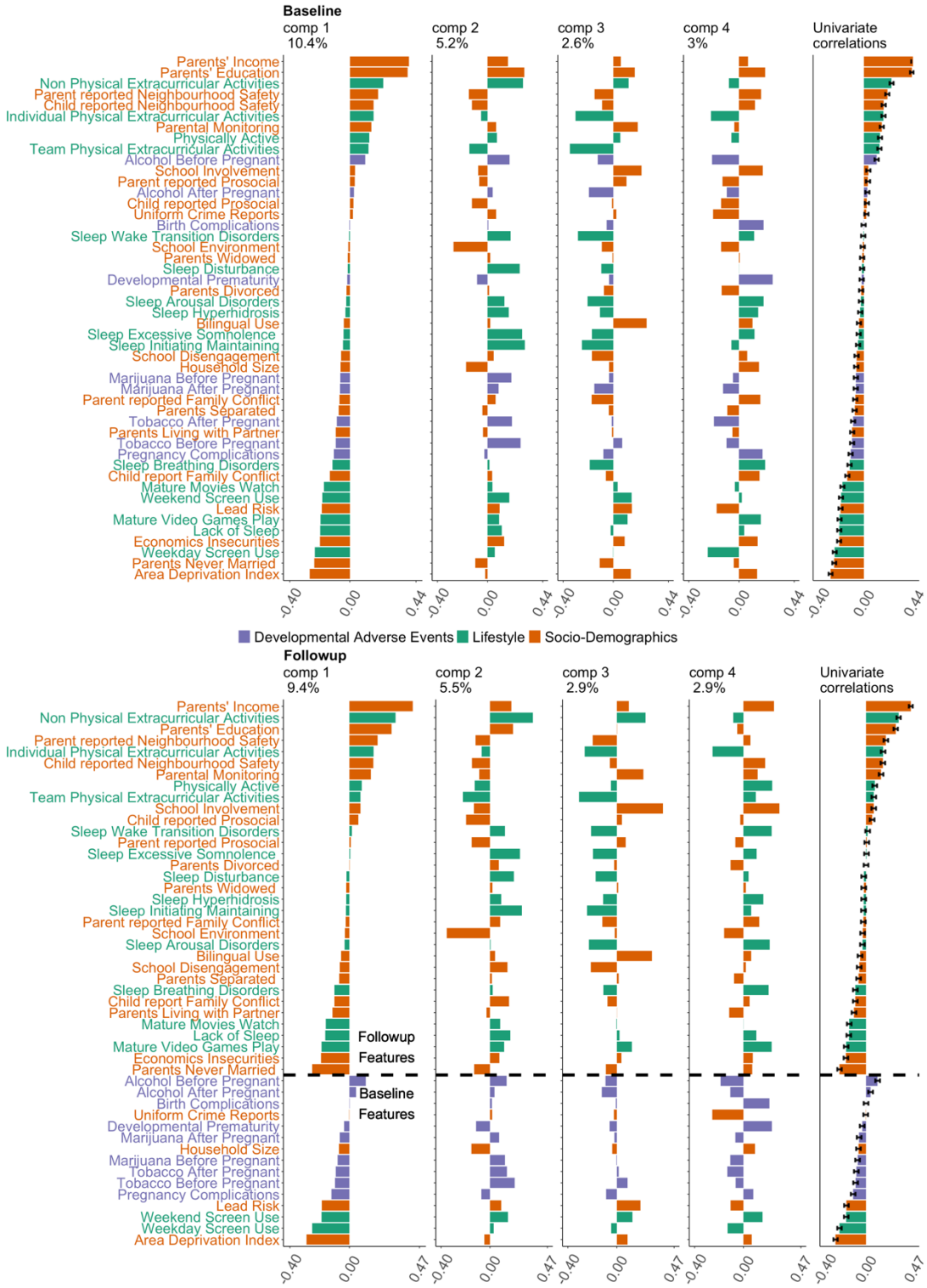


Figure 4. Predictive models, predicting cognitive abilities from socio-demographics, lifestyles and developmental adverse events via Partial Least Square (PLS). a) Scatter plots between observed vs predicted cognitive abilities based on socio-demographics, lifestyles and developmental adverse events. Cognitive abilities are based on the second-order latent variable, the g-factor, based on a confirmatory factor analysis of six cognitive tasks. All data points are from test sets. *r* is the average Pearson's *r* across 21 test sites. The parentheses following the *r* indicate bootstrapped 95%CIs, calculated based on observed vs predicted cognitive abilities from all test sites combined. b) Feature importance of socio-demographics, lifestyles and developmental adverse events, 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. Different types of environmental factors were filled with different colours: orange for socio-demographics, purple for developmental adverse events and green for lifestyle. A dashed horizontal line in the follow-up feature importance figure distinguishes whether the variables were collected at baseline or follow-up.

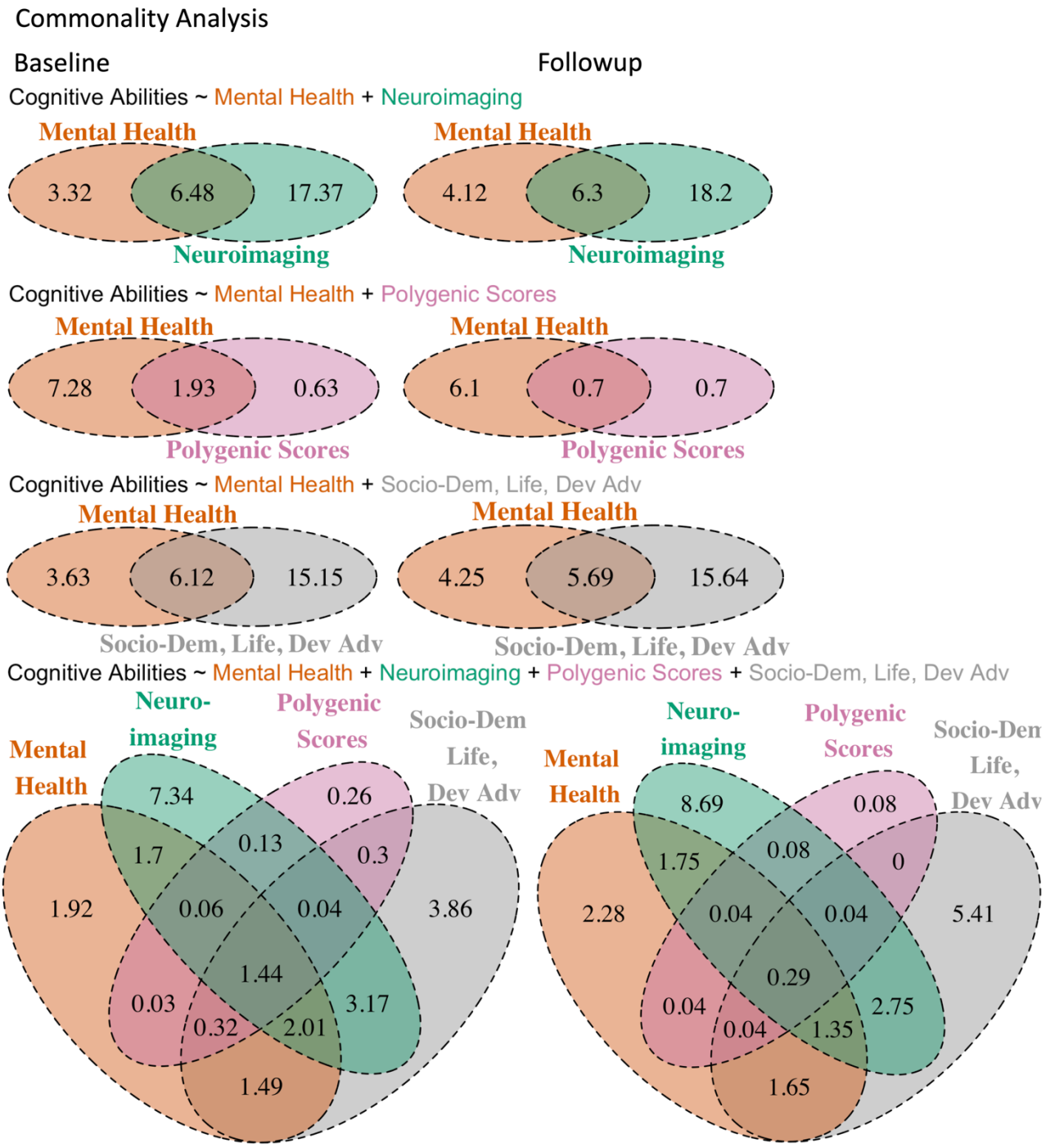


Figure 5. Venn diagrams showing common and unique effects of proxy measures of cognitive abilities based on mental health, neuroimaging, polygenic scores and/or socio-demographics, lifestyles and developmental adverse events in explaining cognitive abilities across test sites. We computed the common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