**Supplementary Information**

1. **METHOD SUPPLEMENT**
   1. **Data and Measures**

***Early life adversity***

A total of 21 questions provided a binary score such that an adverse experience was coded as present or not (1 or 0 respectively). One question was a Likert-type scale (‘Violence is not a problem in my neighbourhood’) which was coded as 1 (adversity present) if the parent ‘strongly disagreed’ with the statement. Peer-related events like bullying and victimisation were not considered as these tend to be more prevalent in adolescence (Bowes et al., 2015; Finkelhor et al., 2013). One-time traumas were similarly excluded from the analysis as these do not represent chronic forms of adversity (Danese & McEwen, 2012; Teicher & Samson, 2016). Finally, although household psychopathology is frequently included in adversity scales (e.g., Berman et al., 2022; Finkelhor et al., 2013), this item was excluded as an indicator of adversity as it is a likely shared genetic confound for mental illness (Teicher & Parigger, 2015).

***Cognition***

A total of 57 cognitive variables were included in this study. These broadly spanned the domains of general cognition (i.e., vocabulary, learning, memory, executive function), language, risk-taking and probabilistic reasoning, affective-cognition, and reward processing.

***NIH Toolbox Cognitive Battery*** *–* The NIH Toolbox Cognitive Battery (NIHTCB) was developed as part of the NIH Blueprint for Neuroscience Research (Weintraub et al., 2013) to provide standardised measures of seven cognitive domains identified as most important for adaptive function (Carlozzi et al., 2015). The cognitive battery consists of seven tasks that were designed to assess the domains of episodic memory; executive function; working memory; attention; processing speed; and language (Casaletto et al., 2015; Weintraub et al., 2013). All measures included in the battery have been validated across a range of ages, socio-demographic groups and levels of education (Casaletto et al., 2015; Weintraub et al., 2013). Further information on task administration, test-retest reliability, and validity are detailed elsewhere (Carlozzi et al., 2015; Weintraub et al., 2013; Zelazo et al., 2013). Testing was administered through an iPad with a research assistant present and lasted approximately 35 minutes. The majority of NIHTCB measures were taken from the third time point (T3) when the children were aged 12, except the Dimensional Change Card Sort and List Sort Working Memory tasks which were only available at Baseline (T1) when the children were aged 10. Age-corrected scores were used.

***Rey Auditory Verbal Learning Task –*** The Rey Auditory Verbal Learning Task (RAVLT) measures auditory learning, memory, recognition, and delayed recall. Participants are required to listen to a list of 15 random words, over five learning trials. A list of distractor words is then presented, after which they are then asked to recall as many words as possible from the distractor list. Next, participants are asked to recall the words presented in the initial learning list. Finally, long-term retention is assessed by asking participants to recall the initial learning list again after a 30-minute delay. Recognition is also assessed by asking participants if they recognised the word from the original list (yes/no) (Luciana et al., 2018). Normative data across childhood and adulthood is available for RAVLT (Strauss et al., 2006). The test is also sensitive to memory deficits in neurodevelopmental disorders (Vakil et al., 2012), affective disorders (Günther et al., 2004), and substance use (Solowij et al., 2011). Internal reliability estimates are high (coefficient alpha ~ 0.90) and test-retest reliability is acceptable (r-values ~ 0.60-0.70) (Van Den Burg & Kingma, 1999). The task was administered through an iPad using a customised computerised version created by Pearson Assessments (Luciana et al., 2018). T3 raw scores were used.

***Little Man Task –*** The Little Man Task (LMT) measures visuospatial reasoning, in particular mental rotation (Acker & Acker, 1982). The task involves the presentation of a cartoon figure holding a briefcase in either the left or right hand. The figure may appear in one of four positions: upside down or right side up, either facing the respondent or with his back to the respondent. Participants are required to indicate which hand is holding the briefcase using one of two buttons. Participants are urged to respond as quickly and accurately as possible. LMT has also been shown to map onto ABCD measures of fluid and crystallised reasoning (Luciana et al., 2018). The task was administered using an iPad. Raw scores and reaction times from T3 were used.

***Matrix Reasoning Task –*** A modified version of the Matrix Reasoning subtest from the Weschler Intelligence test for Children-V (WISC-V) was used as a measure of nonverbal reasoning, visual intelligence, perceptual organisation, and sequencing (Wechsler, 2014). The test is a well-validated and reliable measure of nonverbal reasoning (Wechsler, 2014). On each trial of the task, participants are presented with an incomplete array of visual stimuli, and participants are asked to select from a list of four alternatives to complete the series. The task consists of a maximum of 32 possible trials, stopping after three consecutive incorrect responses. The total number of correct items completed across trials is converted into a standard score using a normed database (*M* = 10.0, SD = 3.0). The ABCD sample is normally distributed with respect to the Matrix Reasoning normed standards (Luciana et al., 2018). The Matrix Reasoning Task falls within the Fluid Reasoning factor of the WISC-V and correlates positively with ABCD measures of fluid and crystallised reasoning, and with the RAVLT (Luciana et al., 2018). Split-half reliability for children aged 9-10 is strong (r= 0.87), and retest stability is acceptable (r= 0.78). The task is administered using an iPad and can be completed in under 10 minutes. Scaled scores T1 were used.

***Cash Choice Task –*** The Cash Choice Task (CCT) is a single-item question measuring delay of gratification (Wulfert et al., 2002). The research assistant asks the child: “Let’s pretend that a kind person wanted to give you some money. Would you rather have $75 in three days or $115 in 3 months?”. The participant must choose between these two options or a third ‘can’t decide’ option. The participant’s response is recorded using an iPad. Studies have shown a general pattern of decline of the immediate reward preference with age, while also maintaining a certain degree of within-person stability across time. Choice selection is also associated with externalising behaviour and substance use in older children and adolescents (Anokhin et al., 2011; Sparks et al., 2014). Item response at Baseline (T1) was used.

***Delay Discounting Task –*** The Delay Discounting Task (DDT) measures reward processing by asking the child to choose between a small hypothetical reward given immediately versus a larger hypothetical reward given at various points in the future (Koffarnus & Bickel, 2014). The choices are presented in seven blocks consisting of six choices in each block (42 total choices). Each block is defined by the time interval which would have to elapse when the larger hypothetical reward would be given: 6 hours; 1 day; 1 week; 1 month; 3 months; 1 year; and 5 years. Each block automatically adjusts the value of the small immediate reward based on the participant’s previous choices. In other words, the immediate reward increasees in value if the future reward was previously chosen or decreases in value if the immediate reward was previously chosen. The task thus converges on an ‘indifference point’ where the subjective value of the small-immediate reward is equal to the subjective value of the delayed but larger reward (i.e., there is no difference between a $100 immediate reward today and a $110 reward in a week) (Luciana et al., 2018).

The indifference points are used to calculate a discounting constant *k* (Bickel et al., 2012; Odum, 2011). K-values from DDT show good test-retest reliability over short time intervals (Matusiewicz et al., 2013). Higher k-values indicate more severe discounting, which tends to decrease through the teenage years as adolescents become increasingly future-oriented (Steinberg et al., 2009). Higher delay discounting rates are also characteristic of individuals with greater impulsivity (De Wit, 2009), substance use disorders (Amlung et al., 2017; Stanger et al., 2012), and ADHD and schizophrenia (Bickel et al., 2012). Although the DDT uses hypothetical rewards, studies show that k-values derived from hypothetical rewards tend to correlate well with k-values derived from real monetary incentives within individuals ( *r*’s= 0.80+; Johnson & Bickel, 2002; Matusiewicz et al., 2013). DDT was administered using an iPad. Indifference scores and response times from Time 2 (T2) were used.

***Game of Dice Task –*** The Game of Dice Task (GDT) measures participants’ aversion/attraction to risky decisions and probabilistic reasoning (Brand et al., 2005). Participants are asked to predict the outcome of a dice roll by selecting among different options ranging from high probability but low payoff to options with low probability but high payoff. Participants are instructed to maximize their hypothetical capital within 18 throws of a virtual die. At the start of each trial, before the die is thrown, participants must bet on the outcome of the throw. Bets are split into four possible levels with the riskiest being a possible payout or loss of $1000. The lower the expected probability, the higher the reward. After the die is rolled, participants are shown feedback on whether they gained or lost money (Donati et al., 2019). The software records the number of winning and losing bets, the final account balance, and the proportion of high-risk and low-risk choices. The number of winning vs losing bets, the final account balance, and the proportion of high- vs low-risk choices from Time 3 (T3) were used.

***EN-Back Task* –** The EN-Back task is an emotional variant of the traditional N-back task which engages working memory, using a block design that adds elements of facial and emotional processing (Barch et al., 2013). The task is performed in an MRI scanner and is designed to investigate working memory, facial recognition, emotional processing, and emotional bias, including bias towards threat using fMRI contrast subtraction. The current analysis used only the behavioural measures derived from this task, including accuracy and response time. Participants are shown a series of stimuli and are asked to indicate whether each one was the same or different from the stimulus presented *N* items earlier (i.e., *N*-back). The EN-back contains two conditions of high and low working memory load: a 2-back and 0-back respectively. Participants are required to indicate their responses on a two-button keypad, with one button indicating that the stimulus is a ‘match’ and the other indicating a ‘no match’. The task consists of two runs of eight blocks and four 15s periods containing just a fixation cross. Each block contained ten trials lasting 2.5s each and was preceded by an instruction screen indicating the condition for each upcoming block. Within each block, two trials were targets, two to three were non-target lures, and the remainder were non-lures (i.e., stimuli only presented once). In total, there were 160 trials with 96 unique images of four different types of stimuli: happy, fearful, or neutral racially diverse adult human faces, as well as images of places as a fourth stimulus type. Additionally, a post-scan recognition memory test was performed asking participants if they recognised previously presented stimuli from a set of 48 new and 48 previously seen images (12 from each stimulus type; 96 total). Participants were asked to rate each picture as either ‘old’ or ‘new’ (Chaarani et al., 2021). The task shows good reliability across individuals (Drobyshevsky et al., 2006), time (Caceres et al., 2009), and is developmentally appropriate (Casey et al., 2018). Accuracy and reaction times from Time 3 (T3) were used.

***Emotional Faces Stroop Task –*** The Emotional Faces Stroop (EF Stroop) task is a variant of the classic Stroop task which requires individuals to attend to less salient stimulus cues while ignoring more salient or automatically processed cues (Stroop, 1935), with an added emotional component. The task is designed to investigate executive control, the ability to maintain attention on the task at hand without interference from distracting information. The classic Stroop task presents a colour-word stimulus combination, with colour being the task-relevant dimension and the word being the irrelevant dimension (distractor). Incongruent trials are those in which the ink colour and word do not match (e.g., the word ‘red’ but in blue ink), whereas congruent trials are those in which the ink colour and word align. In contrast, the modified EF Stroop task that was used in this study involves comparing performance on word identification in the presence of task-irrelevant faces (Başgöze et al., 2015). In this task, participants must decide whether a word describes a ‘good’ feeling (i.e., happy, joyful, glad), or a ‘bad’ feeling (i.e., angry, mad, upset). This choice is made in the context of a face positioned behind the word, whose expression can be either congruent or incongruent to the valence of the word. Incongruent trials higher level of executive control and are associated with longer reaction times and higher error rates (Luciana et al., 2018). Incongruent facial expressions are thus potent affective distractors, particularly in adolescence due to changes in social and emotional processing (Casey, 2015; Hare et al., 2008). The task was designed so that the proportion of congruent to incongruent trials varies across blocks (50/50 vs 25/75). This manipulation was included as incongruent trials may serve as external reminders of the task goal: the need for greater executive control, and studies generally show a developmental shift from such stimulus-driven reactive control to more internally driven proactive control over adolescence (Munakata et al., 2012). The task was performed in an MRI scanner, although the current analysis used only the behavioural measures derived from this task, including accuracy and response time. Stroop performance typically increases throughout adolescence, suggesting maturational changes in executive control over this period (Casey et al., 2008; Duell et al., 2018). Accuracy and response times from Time 2 (T2) were used.

* 1. **Participants**

Missing adversity data was coded as “0” because sensitivity analyses revealed that either coding it as 1 or imputing it resulted in an overestimation of adversity in the sample relative to population prevalence estimates (Finkelhor et al., 2005; McLaughlin et al., 2012; Struck et al., 2020). In the first sensitivity analysis, we coded missing data as 1 instead of 0. This resulted in an additional 524 participants with adversity exposures, and an ELA group representing 27% of the total sample. This was much higher than population prevalence estimates would suggest, meaning this approach was heavily biased unlikely representative of real-world data.

Next, we used a multiple imputation package for mixed-type data, missForest in R (Stekhoven et al., 2012) to impute the missing values. Imputing missing values resulted in an additional 512 participants with adversity exposures (1251 in; 26.5% of the total sample), again much higher than population prevalence estimates (Finkelhor et al., 2005; McLaughlin et al., 2012; Struck et al., 2020). Imputation algorithms are heavily biased towards rare cases with binary data (e.g., exposures to ELA). This means adversity is likely over-estimated, explaining why using this method resulted in an unusually high number of children classified as having adversity relative to population prevalence estimates (Finkelhor et al., 2005; McLaughlin et al., 2012; Struck et al., 2020). There were other reasons for not using imputation. First, it would have increased the standard error, which would have been particularly problematic in the case of RF analyses as it can lead to model overfitting (a common problem with machine-learning methods). Second, the data were not missing at random. We tested for associations between missingness and several key cognitive and demographic variables. Missingness was associated interview ethnicity (*p*<.001); parental education (*p*<.001); parental income (*p*<.001) and 3 out of 5 measures of cognition that were tested (*p*s= .001-.05). The missing at random assumption was therefore not plausible, meaning that imputation would be heavily biased. Although imputation is beneficial in some cases, it must be balanced against the possible risks of inducing bias and overfitting, particularly in the case of non-normally distributed binary data (Sterne et al., 2009). While some procedures can handle non-normally distributed data better than others (Van Buuren et al., 1999), it is an ongoing area of development (Horton et al., 2007; Bernaards et al, 2007) that currently has no well-defined solution (Lee & Carlin, 2016; Sullivan et al., 2017). For these reasons, we decided imputation was not appropriate for our data.

* 1. **Analysis**

***Random Forest Classification***

***Comparing Different Feature Selection Methods –*** Multivariate pattern analysis techniques were used to identify the most salient cognitive differences between the ELA and NOA groups. RF provides mean decrease accuracy (MDA) scores to estimate the relative importance and predictive power of each cognitive variable in the model. This is calculated as the loss in prediction accuracy when a given feature is removed from the RF model using the ‘importance’ function in the ‘randomForest’ package in R (Liaw & Wiener, 2022). It should be noted that while RF can capture complex non-linear interactions (Garge et al., 2013), its “black box” and randomised nature means that features that have high predictive power in the context of a model may hold little explanatory value in statistical terms (for a review see Grömping, 2009). Further, because MDA scores are relative to the rest of the features in the model and can be biased in the presence of multicollinearity, they do not provide insight into the nature or strength of the relationship between a given input feature and the output the model is trying to predict (Strobl et al., 2007, 2008). Finally, RF can be biased in calculating variable importance when features vary in the scale of measurement and number of categories, resulting in a high selection of noise variables (Hothorn et al., 2012; Lu & Petkova, 2014). Thus, RF feature importance metrics ought not to be used for statistical inference, but rather can help with model interpretability and to identify which features are powering predictions in a given model.

In contrast, elastic net regularisation (ENR), a variant of least absolute shrinkage and selection operator (Lasso), provides variable importance metrics that are more interpretable in linear and statistical terms, although they may not capture unspecified or complex interactions as in RF (Lu & Petkova, 2014). Further, ENR importance metrics can be biased when features are highly correlated, leading to the removal of important features even though they may be involved in the process of interest. This is because ENR handles multicollinearity by removing the influence of features that are highly correlated and that do not improve model accuracy. It does so by first shrinking unimportant coefficients to zero using Lasso regularisation. Next, it reduces the size of coefficients that are highly correlated with other predictors using Ridge regression (Hastie et al., 2009; Zou & Hastie, 2005). Lambda is used as a tuning parameter that determines how much to regularise, whereas alpha is a hyperparameter that determines the balance between Lasso and Ridge in the model. An alpha of 0 produces a Ridge model, whereas a value of 1 produces a Lasso model. The ENR model was implemented using the ‘glmnet’ package in R (Friedman et al., 2010), using 10-fold cross-validation repeated 5 times on the training set. Receiver operating characteristic curve (ROC) was used to select the optimal alpha and lambda parameters. The final values used for the model were alpha= 0.1 and lambda= 0.01369358, which produce a model closer to Ridge.

Overall, RF and ENR feature importance metrics provide information at different levels of prediction and explanation to identify cognitive features that differ significantly between the ELA and NOA groups (Kuhn & Johnson, 2013). Variables which are consistently selected across different feature selection methods may hold greater predictive power in the real world, although caution is warranted as discrepancies may also stem from fundamental differences between models as outlined above (Kuhn & Johnson, 2013). We therefore report on RF feature selection in the main text to help with model interpretation, but also report ENR importance metrics in the Supplement for completeness.

1. **RESULTS SUPPLEMENT** 
   1. **Identifying subgroups**

***Subgroups were not driven by demographic factors -*** There were no significant differences in the number of boys or girls in the two NOA subgroups, *Χ²*(1)= 0.05, *p*= .82. NOA-1 (*M*= 11.92, *SD*= 0.65). NOA-low was significantly younger than NOA-high (*M*= 12.05, *SD*= 0.64), *t*(5214)= -6.741, *p* < .001, *d*= -0.21. Despite being younger, boys in NOA-low (*M*= 1.917, *SD*= 0.871), had faster pubertal development than boys in NOA-high (*M*= 1.827, *SD*= 0.854), *t*(2671)= 2.443, *p*= .015, *d*= 0.104 (Bonferroni adjusted alpha level of 0.025 per test (0.05/2)). There were no significant subgroup differences in pubertal development in girls, *t*(2299)= 0.65, *p*= .52 in the two NOA subgroups.

There were no significant differences in the number of boys or girls in the two ELA subgroups, *Χ²*(1)= 0.61, *p*= .44. There were also no significant age differences between ELA-high (*M*= 11.98, *SD*= 0.64) and ELA-low (*M*= 11.88, *SD*= 0.65), *t*(314)= 1.806, *p*= .07). Using a Bonferroni adjusted alpha level of 0.025 per test (0.05/2), there were no significant subgroup differences in pubertal development in girls, *t*(314)= 0.488, *p*= .63, nor in boys, *t*(360)= -0.927, *p*= .36.

1. **TABLES SUPPLEMENT**

**Table S1.**Demographics for the ABCD cohort

|  |  |
| --- | --- |
| *N* | 11,876 |
| Age mean *(SD)* | 9.9 *(0.6)* |
| % Female | 47.8% |
|  |  |
| *Race/Ethnicity* |  |
| % White | 52.2% |
| % African American | 15.1% |
| % Hispanic | 3.9% |
| % Asian | 2.2% |
| % Other/Multi-racial | 25.9% |
|  |  |
| *Household characteristics* |  |
| % Married caregivers | 65.5% |
| % College-level education | 59.4% |
|  |  |
| *Household income* |  |
| < $25k | 17.2% |
| $25k - $49.99k | 22.4% |
| $50k -$74.99k | 14.5% |
| $75k -$99.99k | 30.5% |
| $100k + | 15.4% |

*Notes*. Demographics reported at baseline (T1). Age reported in years.

**Table S2**. Measures of adversity

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Category** | **Dimension** | **Timescale** |
| Child was shot, stabbed, or beaten brutally by a grown up in the home | Physical Abuse | Threat | Lifetime (age 0-9) |
| Child was beaten to the point of having bruises by a grown up in the home | Physical Abuse | Threat | Lifetime (age 0-9) |
| A family member threatened to kill your child | Physical Abuse | Threat | Lifetime (age 0-9) |
| Child was shot, stabbed, or beaten brutally by a non-family member | Physical Abuse | Threat | Lifetime (age 0-9) |
| A non-family member threatened to kill your child | Physical Abuse | Threat | Lifetime (age 0-9) |
| A grown up in the home touched your child in his or her privates, had your child touch their privates, or did other sexual things to your child | Sexual Abuse | Threat | Lifetime (age 0-9) |
| An adult outside your family touched your child in his or her privates, had your child touch their privates or did other sexual things to your child | Sexual Abuse | Threat | Lifetime (age 0-9) |
| A peer forced your child to do something sexually | Sexual Abuse | Threat | Lifetime (age 0-9) |
| Child witnessed the grownups in the home push, shove or hit one another | Domestic Violence | Threat | Lifetime (age 0-9) |
| Child witnessed death or mass destruction in a war zone | Community Violence | Threat | Lifetime (age 0-9) |
| Child witnessed someone shot or stabbed in the community | Community Violence | Threat | Lifetime (age 0-9) |
| Neighborhood violence is a problem | Community Violence | Threat | Lifetime (age 0-9) |
| Family needed food but couldn't afford to buy it or couldn't afford to go out to get it | Material Deprivation | Deprivation | Past 12 months |
| Family was evicted from home for not paying the rent or mortgage | Material Deprivation | Deprivation | Past 12 months |
| Family had services turned off by the gas or electric company, or the oil company wouldn't deliver oil because payments were not made? | Material Deprivation | Deprivation | Past 12 months |
| Someone in family needed to see a doctor or go to the hospital but didn't go because could not afford it | Material Deprivation | Deprivation | Past 12 months |
| Someone in family needed a dentist but couldn't go because could not afford it | Material Deprivation | Deprivation | Past 12 months |
| Household substance abuse: biological mother (Isolated self, arguments, drunk a lot) | Household Substance Abuse | Deprivation | Lifetime (age 0-9) |
| Household substance abuse: biological father (Isolated self, arguments, drunk a lot) | Household Substance Abuse | Deprivation | Lifetime (age 0-9) |
| Household substance abuse: biological father (Isolated self, arguments, was high a lot) | Household Substance Abuse | Deprivation | Lifetime (age 0-9) |
| Household substance abuse: biological mother (Isolated self, arguments, was high a lot) | Household Substance Abuse | Deprivation | Lifetime (age 0-9) |

*Notes*. Adversity measures taken at baseline assessment (T1). Parental-report used.

**Table S3**. Measures of cognition

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Measure** | **Study Variable Name** | **Domain** | **Timepoint** | **Description** |
| Cash Choice Task | Cash Choice Task | Reward processing | T1 | Would you rather have $75 in three days or $115 in 3 months? |
| Delay Discounting | DDIS (delayed choice RT) | Reward processing | T2 | Mean latency (in ms) of 'delayed' choices |
| Delay Discounting | DDIS (immediate choice RT) | Reward processing | T2 | Mean latency (in ms) of 'immediate' choices |
| Delay Discounting | DDIS (immediate choice validity) | Reward processing | T2 | Tally of immediate choices in the three validity questions presented after the "3months", "1year", and "5years" delay blocks. |
| Delay Discounting | DDIS (1d delay indifference) | Reward processing | T2 | 1-day delay "indifference point" |
| Delay Discounting | DDIS (1mo delay indifference) | Reward processing | T2 | 1 month delay "indifference point" |
| Delay Discounting | DDIS (1wk delay indifference) | Reward processing | T2 | 1-week delay "indifference point" |
| Delay Discounting | DDIS (1yr delay indifference) | Reward processing | T2 | 1 year delay "indifference point" |
| Delay Discounting | DDIS (3mo delay indifference) | Reward processing | T2 | 3-month delay "indifference point" |
| Delay Discounting | DDIS (5yr delay indifference) | Reward processing | T2 | 5-year delay "indifference point" |
| Delay Discounting | DDIS (6h delay indifference) | Reward processing | T2 | 6-hour delay "indifference point" |
| Game of Dice Task | Game of Dice (safe vs. risky bets) | Risk-taking and prob. reasoning | T3 | Reflects the number of safe vs. risky choices. |
| Game of Dice Task | Game of Dice (account balance) | Risk-taking and prob. reasoning | T3 | Account balance at end of game |
| Game of Dice Task | Game of Dice (losing bets) | Risk-taking and prob. reasoning | T3 | Adds the number of losing bets |
| Game of Dice Task | Game of Dice (winning bets) | Risk-taking and prob. reasoning | T3 | Adds the number of winning bets |
| Little Man Task | Little Man Task (RT) | General cognition | T3 | Average reaction time for trials. Does not include timed out trials. |
| Little Man Task | Little Man Task (score) | General cognition | T3 | Percentage correct of all 32 presented trials |
| NIH Toolbox Dimensional Change Card Sort | Dimensional Change Card Sort | General cognition | T1 | Age-corrected score |
| NIH Toolbox List Sort Working Memory | List Sort Working Memory | General cognition | T1 | Age-corrected score |
| NIH Toolbox Flanker Inhibitory Control & Attention | Flanker Inhibitory Control | General cognition | T3 | Age-corrected score |
| NIH Toolbox Pattern Comparison Processing Speed | Pattern Comparison Processing Speed | General cognition | T3 | Age-corrected score |
| NIH Toolbox Picture Sequence Memory | Picture Sequence Memory | General cognition | T3 | Age-corrected score |
| NIH Toolbox Picture Vocabulary | Picture Vocabulary | General cognition | T3 | Age-corrected score |
| NIH Toolbox Oral Reading Recognition | Oral Reading Recognition | General cognition | T3 | Age-corrected score |
| Rey Auditory Verbal Learning Task | RAVLT (long delay score) | General cognition | T3 | Total correct on long delay (trial 7) |
| Rey Auditory Verbal Learning Task | RAVLT (learning score) | General cognition | T3 | Total learning score (sum of learning trials 1-5) |
| Rey Auditory Verbal Learning Task | RAVLT (distraction score) | General cognition | T3 | Total correct on immediate distraction trial (trial B) |
| Rey Auditory Verbal Learning Task | RAVLT (short delay score) | General cognition | T3 | Total correct on short delay (trial 6) |
| Matrix Reasoning | Matrix Reasoning (total score) | General cognition | T1 | Total scaled score |
| Emotional Faces Stroop | EF Stroop (75% congruent angry accuracy) | Socio-emotional processing | T2 | Proportion correct of congruent 'angry face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (75% congruent angry RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in congruent 'angry face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% congruent angry accuracy) | Socio-emotional processing | T2 | Proportion correct of congruent 'angry face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% congruent angry RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in congruent 'angry face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (75% congruent happy accuracy) | Socio-emotional processing | T2 | Proportion correct of congruent 'happy face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (75% congruent happy RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in congruent 'happy face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% congruent happy accuracy) | Socio-emotional processing | T2 | Proportion correct of congruent 'happy face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% congruent happy RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in congruent 'happy face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (25% incongruent angry accuracy) | Socio-emotional processing | T2 | Proportion correct of incongruent 'angry face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (25% incongruent angry RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in incongruent 'angry face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% incongruent angry accuracy) | Socio-emotional processing | T2 | Proportion correct of incongruent 'angry face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% incongruent angry RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in incongruent 'angry face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (25% incongruent happy accuracy) | Socio-emotional processing | T2 | Proportion correct of incongruent 'happy face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (25% incongruent happy RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in incongruent 'happy face' trials in "more congruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% incongruent happy accuracy) | Socio-emotional processing | T2 | Proportion correct of incongruent 'happy face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional Faces Stroop | EF Stroop (50% incongruent happy RT) | Socio-emotional processing | T2 | Mean correct response times (in ms) in incongruent 'happy face' trials in "equal numbers of congruent/incongruent pairings" block |
| Emotional n-back task | En-Back (mean RT to neutral faces) | Socio-emotional processing | T3 | Average reaction time for all correct responses to neutral face stimuli |
| Emotional n-back task | En-Back (accuracy neutral faces) | Socio-emotional processing | T3 | The rate of correct responses to neutral face stimuli |
| Emotional n-back task | En-Back (SD of RT to neutral faces) | Socio-emotional processing | T3 | Standard deviation of the reaction time for all correct responses to neutral face stimuli |
| Emotional n-back task | En-Back (mean RT to negative faces) | Socio-emotional processing | T3 | Average reaction time for all correct responses to negative face stimuli |
| Emotional n-back task | En-Back (accuracy negative faces) | Socio-emotional processing | T3 | The rate of correct responses to negative face stimuli |
| Emotional n-back task | En-Back (SD of RT to negative faces) | Socio-emotional processing | T3 | Standard deviation of the reaction time for all correct responses to negative face stimuli |
| Emotional n-back task | En-Back (mean RT to positive faces) | Socio-emotional processing | T3 | Average reaction time for all correct responses to positive face stimuli |
| Emotional n-back task | En-Back (accuracy positive faces) | Socio-emotional processing | T3 | The rate of correct responses to positive face stimuli |
| Emotional n-back task | En-Back (SD of RT to positive faces) | Socio-emotional processing | T3 | Standard deviation of the reaction time for all correct responses to positive face stimuli |
| Emotional n-back task | En-Back (bias negative faces) | Socio-emotional processing | T3 | Response bias for negative faces |
| Emotional n-back task | En-Back (bias neutral faces) | Socio-emotional processing | T3 | Response bias for neutral faces |
| Emotional n-back task | En-Back (bias positive faces) | Socio-emotional processing | T3 | Response bias for positive faces |

**Table S4**. Measures of mental health

|  |  |  |
| --- | --- | --- |
| **Study Variable Name** | **ABCD Variable Name** | **Measure** |
| Anxious-Depressed | cbcl\_scr\_syn\_anxdep\_t | CBCL-syndrome scale |
| Withdrawn-Depressed | cbcl\_scr\_syn\_withdep\_t | CBCL-syndrome scale |
| Somatic Complains | cbcl\_scr\_syn\_somatic\_t | CBCL-syndrome scale |
| Social Problems | cbcl\_scr\_syn\_social\_t | CBCL-syndrome scale |
| Thought Problems | cbcl\_scr\_syn\_thought\_t | CBCL-syndrome scale |
| Attention Problems | cbcl\_scr\_syn\_attention\_t | CBCL-syndrome scale |
| Rule Breaking | cbcl\_scr\_syn\_rulebreak\_t | CBCL-syndrome scale |
| Aggressive Behaviour | cbcl\_scr\_syn\_aggressive\_t | CBCL-syndrome scale |
| Sluggish Cognitive Tempo | cbcl\_scr\_07\_sct\_t | CBCL-syndrome scale |
| Obsessive-Compulsive Problems | cbcl\_scr\_07\_ocd\_t | CBCL-syndrome scale |
| Stress Problems | cbcl\_scr\_07\_stress\_t | CBCL-syndrome scale |
| Internalising Composite | cbcl\_scr\_syn\_internal\_t | CBCL-syndrome scale |
| Externalising Composite | cbcl\_scr\_syn\_external\_t | CBCL-syndrome scale |
| Total Problems | cbcl\_scr\_syn\_totprob\_t | CBCL-syndrome scale |
| Depressive Disorder | cbcl\_scr\_dsm5\_depress\_t | CBCL DSM-oriented scale |
| Anxiety Disorder | cbcl\_scr\_dsm5\_anxdisord\_t | CBCL DSM-oriented scale |
| Somatic Disorder | cbcl\_scr\_dsm5\_somaticpr\_t | CBCL DSM-oriented scale |
| ADHD | cbcl\_scr\_dsm5\_adhd\_t | CBCL DSM-oriented scale |
| Oppositional-Defiant Disorder | cbcl\_scr\_dsm5\_opposit\_t | CBCL DSM-oriented scale |
| Conduct Disorder | cbcl\_scr\_dsm5\_conduct\_t | CBCL DSM-oriented scale |
| Psychotic Symptoms | pps\_y\_ss\_severity\_score | Prodromal Psychosis scale |

*Notes*. Measures taken at the third data-collection point (T3). CBCL = Child Behaviour Checklist. ADHD = attention deficit hyperactivity disorder. DSM = Diagnostic and Statistical Manual of Mental Disorders.

**Table S5**. Differences in cognition between the ELA and NOA groups

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | | **SE Diff** | | **Cohen's d** | **Lower CI** | **Upper CI** | **FDR-adj *p*** |  |
| Picture Vocabulary | 17.48 | 9.62 | 0.55 | | 0.66 | | 0.57 | 0.74 | <.001 | \*\*\* |
| Oral Reading Recognition | 12.81 | 7.82 | 0.61 | | 0.49 | | 0.41 | 0.57 | <.001 | \*\*\* |
| Matrix Reasoning (total score) | 12.23 | 1.39 | 0.11 | | 0.48 | | 0.40 | 0.56 | <.001 | \*\*\* |
| Flanker Inhibitory Control | 10.58 | 5.81 | 0.55 | | 0.41 | | 0.33 | 0.49 | <.001 | \*\*\* |
| RAVLT (long delay score) | 9.70 | 1.05 | 0.11 | | 0.38 | | 0.30 | 0.46 | <.001 | \*\*\* |
| Picture Sequence Memory | 9.47 | 5.99 | 0.63 | | 0.37 | | 0.29 | 0.45 | <.001 | \*\*\* |
| En-Back (accuracy neutral faces) | 8.25 | 0.05 | 0.01 | | 0.36 | | 0.28 | 0.43 | <.001 | \*\*\* |
| RAVLT (short delay score) | 9.06 | 0.93 | 0.10 | | 0.35 | | 0.27 | 0.43 | <.001 | \*\*\* |
| RAVLT (learning score) | 8.57 | 3.66 | 0.43 | | 0.34 | | 0.26 | 0.42 | <.001 | \*\*\* |
| En-Back (accuracy negative faces) | 7.61 | 0.05 | 0.01 | | 0.33 | | 0.25 | 0.41 | <.001 | \*\*\* |
| En-Back (accuracy positive faces) | 7.58 | 0.05 | 0.01 | | 0.32 | | 0.24 | 0.40 | <.001 | \*\*\* |
| Dimensional Change Card Sort | 7.98 | 4.48 | 0.56 | | 0.31 | | 0.23 | 0.38 | <.001 | \*\*\* |
| List Sort Working Memory | 7.68 | 4.08 | 0.53 | | 0.30 | | 0.22 | 0.38 | <.001 | \*\*\* |
| RAVLT (distraction score) | 7.48 | 0.51 | 0.07 | | 0.29 | | 0.22 | 0.37 | <.001 | \*\*\* |
| Game of Dice (account balance) | 6.85 | 1089.54 | 159.03 | | 0.28 | | 0.21 | 0.36 | <.001 | \*\*\* |
| DDIS (6h delay indifference) | 6.29 | 7.01 | 1.11 | | 0.27 | | 0.19 | 0.35 | <.001 | \*\*\* |
| Game of Dice (safe vs. risky bets) | 6.26 | 2.50 | 0.40 | | 0.25 | | 0.17 | 0.33 | <.001 | \*\*\* |
| DDIS (1wk delay indifference) | 5.78 | 7.24 | 1.25 | | 0.24 | | 0.16 | 0.32 | <.001 | \*\*\* |
| DDIS (1d delay indifference) | 5.42 | 6.05 | 1.12 | | 0.23 | | 0.15 | 0.31 | <.001 | \*\*\* |
| DDIS (1mo delay indifference) | 5.08 | 6.93 | 1.36 | | 0.21 | | 0.13 | 0.29 | <.001 | \*\*\* |
| Game of Dice (winning bets) | 4.69 | 0.54 | 0.12 | | 0.19 | | 0.11 | 0.27 | <.001 | \*\*\* |
| Pattern Comparison Processing Speed | 4.79 | 3.65 | 0.76 | | 0.19 | | 0.11 | 0.26 | <.001 | \*\*\* |
| DDIS (3mo delay indifference) | 4.33 | 6.07 | 1.40 | | 0.18 | | 0.10 | 0.25 | <.001 | \*\*\* |
| DDIS (immediate choice RT) | 4.33 | 157.30 | 36.37 | | 0.17 | | 0.09 | 0.24 | <.001 | \*\*\* |
| EF Stroop (75% congruent angry accuracy) | 3.63 | 0.01 | 0.00 | | 0.16 | | 0.08 | 0.23 | <.001 | \*\*\* |
| EF Stroop (25% incongruent happy accuracy) | 3.62 | 0.02 | 0.01 | | 0.15 | | 0.07 | 0.22 | <.001 | \*\*\* |
| EF Stroop (50% congruent angry accuracy) | 3.26 | 0.01 | 0.00 | | 0.14 | | 0.07 | 0.22 | <.001 | \*\*\* |
| EF Stroop (50% incongruent angry accuracy) | 3.38 | 0.02 | 0.00 | | 0.14 | | 0.06 | 0.21 | <.001 | \*\*\* |
| EF Stroop (50% incongruent happy accuracy) | 3.08 | 0.02 | 0.01 | | 0.13 | | 0.05 | 0.21 | <.001 | \*\*\* |
| DDIS (1yr delay indifference) | 2.82 | 4.08 | 1.45 | | 0.11 | | 0.04 | 0.19 | <.001 | \*\*\* |
| EF Stroop (75% congruent happy accuracy) | 2.65 | 0.01 | 0.00 | | 0.11 | | 0.03 | 0.19 | 0.001 | \*\* |
| EF Stroop (25% incongruent angry accuracy) | 2.53 | 0.02 | 0.01 | | 0.10 | | 0.03 | 0.18 | 0.002 | \*\* |
| EF Stroop (50% congruent happy accuracy) | 2.61 | 0.01 | 0.00 | | 0.10 | | 0.03 | 0.18 | 0.003 | \*\* |
| DDIS (delayed choice RT) | 2.62 | 84.09 | 32.08 | | 0.10 | | 0.02 | 0.18 | 0.003 | \*\* |
| DDIS (5yr delay indifference) | 0.41 | 0.59 | 1.44 | | 0.02 | | -0.06 | 0.09 | 0.008 | \*\* |
| En-Back (bias positive faces) | 0.37 | 0.00 | 0.01 | | 0.02 | | -0.06 | 0.09 | 0.010 | \*\* |
| En-Back (bias neutral faces) | -0.28 | 0.00 | 0.01 | | -0.01 | | -0.09 | 0.07 | 0.0120 |  |
| En-Back (bias negative faces) | -0.60 | -0.01 | 0.01 | | -0.02 | | -0.10 | 0.05 | 0.0123 |  |
| EF Stroop (75% congruent angry RT) | -1.00 | -6.26 | 6.25 | | -0.04 | | -0.12 | 0.04 | 0.0128 |  |
| Little Man Task (score) | -1.12 | -0.28 | 0.25 | | -0.05 | | -0.13 | 0.03 | 0.0132 |  |
| Little Man Task (RT) | -1.30 | -26.99 | 20.78 | | -0.05 | | -0.13 | 0.02 | 0.0139 |  |
| Cash Choice Task | -1.43 | -0.03 | 0.02 | | -0.06 | | -0.13 | 0.02 | 0.0149 |  |
| EF Stroop (50% congruent angry RT) | -1.44 | -9.85 | 6.86 | | -0.06 | | -0.14 | 0.02 | 0.0186 |  |
| EF Stroop (25% incongruent happy RT) | -1.80 | -14.86 | 8.27 | | -0.07 | | -0.15 | 0.01 | 0.0259 |  |
| En-Back (mean RT to negative faces) | -1.91 | -9.99 | 5.25 | | -0.08 | | -0.16 | 0.00 | 0.0481 |  |
| En-Back (mean RT to neutral faces) | -2.05 | -9.80 | 4.77 | | -0.08 | | -0.16 | 0.00 | 0.0496 |  |
| EF Stroop (75% congruent happy RT) | -2.08 | -12.50 | 6.00 | | -0.08 | | -0.16 | -0.01 | 0.0691 |  |
| EF Stroop (50% incongruent happy RT) | -2.34 | -16.40 | 7.02 | | -0.09 | | -0.17 | -0.02 | 0.0867 |  |
| EF Stroop (25% incongruent angry RT) | -2.47 | -19.83 | 8.03 | | -0.10 | | -0.18 | -0.02 | 0.1756 |  |
| En-Back (mean RT to positive faces) | -2.58 | -13.27 | 5.15 | | -0.11 | | -0.18 | -0.03 | 0.1768 |  |
| En-Back (SD of RT to negative faces) | -2.65 | -7.11 | 2.68 | | -0.11 | | -0.19 | -0.03 | 0.2168 |  |
| EF Stroop (50% incongruent angry RT) | -2.76 | -18.54 | 6.73 | | -0.11 | | -0.19 | -0.03 | 0.2894 |  |
| EF Stroop (50% congruent happy RT) | -3.12 | -19.74 | 6.34 | | -0.12 | | -0.20 | -0.05 | 0.3409 |  |
| En-Back (SD of RT to positive faces) | -3.64 | -9.30 | 2.56 | | -0.15 | | -0.22 | -0.07 | 0.5784 |  |
| Game of Dice (losing bets) | -4.50 | -0.52 | 0.12 | | -0.18 | | -0.26 | -0.11 | 0.7047 |  |
| DDIS (immediate choice validity) | -5.17 | -0.20 | 0.04 | | -0.21 | | -0.29 | -0.13 | 0.7217 |  |
| En-Back (SD of RT to neutral faces) | -5.59 | -15.07 | 2.70 | | -0.23 | | -0.30 | -0.15 | 0.7810 |  |

*Notes*. Welch's t-test. Lower and upper CI represent the 95% confidence interval for Cohen's d. P-values adjusted for false-discovery using the Hochberg procedure with a significance value of .01. \*\* *p* < .01. \*\*\**p* < .001.

**Table S6**. Differences in mental health between the ELA and NOA groups

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | **SE Diff** | **Cohen's d** | **Lower CI** | **Upper CI** | **FDR-adj *p*** |  |
| Anxious-Depressed | -8.67 | -2.80 | 0.32 | -0.42 | -0.51 | -0.33 | < .001 | \*\*\* |
| Withdrawn-Depressed | -10.19 | -3.29 | 0.32 | -0.50 | -0.59 | -0.41 | < .001 | \*\*\* |
| Somatic Complains | -9.71 | -3.04 | 0.31 | -0.46 | -0.55 | -0.38 | < .001 | \*\*\* |
| Social Problems | -11.85 | -3.33 | 0.28 | -0.59 | -0.68 | -0.50 | < .001 | \*\*\* |
| Thought Problems | -9.92 | -3.19 | 0.32 | -0.48 | -0.57 | -0.40 | < .001 | \*\*\* |
| Attention Problems | -11.10 | -3.64 | 0.33 | -0.55 | -0.64 | -0.46 | < .001 | \*\*\* |
| Rule Breaking | -12.66 | -3.19 | 0.25 | -0.64 | -0.73 | -0.55 | < .001 | \*\*\* |
| Aggressive Behaviour | -11.16 | -3.56 | 0.32 | -0.57 | -0.66 | -0.48 | < .001 | \*\*\* |
| Sluggish Cognitive Tempo | -9.71 | -2.74 | 0.28 | -0.47 | -0.56 | -0.39 | < .001 | \*\*\* |
| Obsessive-Compulsive Problems | -7.88 | -2.50 | 0.32 | -0.38 | -0.47 | -0.30 | < .001 | \*\*\* |
| Stress Problems | -11.55 | -3.82 | 0.33 | -0.58 | -0.67 | -0.49 | < .001 | \*\*\* |
| Internalising Composite | -12.44 | -6.13 | 0.49 | -0.56 | -0.65 | -0.47 | < .001 | \*\*\* |
| Externalising Composite | -14.74 | -7.10 | 0.48 | -0.68 | -0.77 | -0.59 | < .001 | \*\*\* |
| Total Problems | -15.59 | -8.18 | 0.53 | -0.71 | -0.80 | -0.62 | < .001 | \*\*\* |
| Depressive Disorder | -11.61 | -3.92 | 0.34 | -0.57 | -0.66 | -0.48 | < .001 | \*\*\* |
| Anxiety Disorder | -8.97 | -3.03 | 0.34 | -0.44 | -0.52 | -0.35 | < .001 | \*\*\* |
| Somatic Disorder | -8.48 | -2.73 | 0.32 | -0.40 | -0.48 | -0.31 | < .001 | \*\*\* |
| ADHD | -11.02 | -3.32 | 0.30 | -0.54 | -0.63 | -0.45 | < .001 | \*\*\* |
| Oppositional-Defiant Disorder | -11.43 | -3.32 | 0.29 | -0.57 | -0.65 | -0.48 | < .001 | \*\*\* |
| Conduct Disorder | -12.59 | -3.66 | 0.29 | -0.63 | -0.72 | -0.54 | < .001 | \*\*\* |
| Psychotic Symptoms | -8.56 | -3.73 | 0.44 | -0.40 | -0.48 | -0.32 | < .001 | \*\*\* |

*Notes*. Welch's t-test. Lower and upper CI represent the 95% confidence interval for Cohen's d. P-values adjusted for false-discovery using the Hochberg procedure with a significance value of .01. \*\*\**p* < .001.

**Table S7**. Variable importance rank results for ENR and RF by decreasing order of importance

|  |  |  |
| --- | --- | --- |
| **Rank** | **Elastic Net Regularisation** | **Random Forest** |
| 1 | Cash Choice Task | Picture Vocabulary |
| 2 | En-Back (accuracy negative faces) | Little Man Task (score) |
| 3 | En-Back (accuracy neutral faces) | DDIS (delayed choice RT) |
| 4 | EF Stroop (75% congruent happy accuracy) | DDIS (3mo delay indifference) |
| 5 | EF Stroop (50% incongruent angry accuracy) | DDIS (1d delay indifference) |
| 6 | Matrix Reasoning (total score) | Matrix Reasoning (total score) |
| 7 | DDIS (immediate choice validity) | DDIS (6h delay indifference) |
| 8 | RAVLT (long delay score) | Oral Reading Recognition |
| 9 | Picture Vocabulary | Picture Sequence Memory |
| 10 | EF Stroop (25% incongruent angry accuracy) | List Sort Working Memory |
| 11 | List Sort Working Memory | DDIS (5yr delay indifference) |
| 12 | Oral Reading Recognition | En-Back (SD of RT to negative faces) |
| 13 | RAVLT (short delay score) | DDIS (1mo delay indifference) |
| 14 | Picture Sequence Memory | DDIS (1wk delay indifference) |
| 15 | Dimensional Change Card Sort | DDIS (immediate choice RT) |
| 16 | Flanker Inhibitory Control | En-Back (SD of RT to positive faces) |
| 17 | Little Man Task (score) | RAVLT (learning score) |
| 18 | RAVLT (distraction score) | Dimensional Change Card Sort |
| 19 | Game of Dice (safe vs. risky bets) | En-Back (accuracy neutral faces) |
| 20 | En-Back (SD of RT to neutral faces) | En-Back (SD of RT to neutral faces) |
| 21 | RAVLT (learning score) | Little Man Task (RT) |
| 22 | DDIS (6h delay indifference) | DDIS (1yr delay indifference) |
| 23 | DDIS (1mo delay indifference) | En-Back (accuracy negative faces) |
| 24 | DDIS (3mo delay indifference) | Pattern Comparison Processing Speed |
| 25 | EF Stroop (75% congruent angry RT) | EF Stroop (50% incongruent happy RT) |
| 26 | En-Back (SD of RT to positive faces) | EF Stroop (50% incongruent angry RT) |
| 27 | DDIS (1yr delay indifference) | Game of Dice (account balance) |
| 28 | DDIS (1d delay indifference) | EF Stroop (50% incongruent angry accuracy) |
| 29 | DDIS (1wk delay indifference) | EF Stroop (50% congruent angry RT) |
| 30 | DDIS (delayed choice RT) | EF Stroop (75% congruent angry accuracy) |
| 31 | Game of Dice (account balance) | Flanker Inhibitory Control |
| 32 | EF Stroop (50% congruent angry RT) | DDIS (immediate choice validity) |
| 33 | DDIS (immediate choice RT) | EF Stroop (25% incongruent happy accuracy) |
| 34 | DDIS (5yr delay indifference) | EF Stroop (75% congruent happy RT) |
| 35 | Game of Dice (losing bets) | EF Stroop (25% incongruent happy RT) |
| 36 | Game of Dice (winning bets) | EF Stroop (75% congruent angry RT) |
| 37 | Little Man Task (RT) | En-Back (accuracy positive faces) |
| 38 | Pattern Comparison Processing Speed | RAVLT (short delay score) |
| 39 | EF Stroop (75% congruent angry accuracy) | En-Back (mean RT to neutral faces) |
| 40 | EF Stroop (50% congruent angry accuracy) | EF Stroop (50% congruent happy accuracy) |
| 41 | EF Stroop (75% congruent happy RT) | EF Stroop (75% congruent happy accuracy) |
| 42 | EF Stroop (50% congruent happy accuracy) | EF Stroop (50% congruent angry accuracy) |
| 43 | EF Stroop (50% congruent happy RT) | Cash Choice Task |
| 44 | EF Stroop (25% incongruent angry RT) | Game of Dice (safe vs. risky bets) |
| 45 | EF Stroop (50% incongruent angry RT) | RAVLT (distraction score) |
| 46 | EF Stroop (25% incongruent happy accuracy) | EF Stroop (25% incongruent angry accuracy) |
| 47 | EF Stroop (25% incongruent happy RT) | EF Stroop (25% incongruent angry RT) |
| 48 | EF Stroop (50% incongruent happy accuracy) | En-Back (bias negative faces) |
| 49 | EF Stroop (50% incongruent happy RT) | EF Stroop (50% congruent happy RT) |
| 50 | En-Back (mean RT to neutral faces) | Game of Dice (losing bets) |
| 51 | En-Back (mean RT to negative faces) | En-Back (mean RT to positive faces) |
| 52 | En-Back (SD of RT to negative faces) | EF Stroop (50% incongruent happy accuracy) |
| 53 | En-Back (mean RT to positive faces) | En-Back (bias positive faces) |
| 54 | En-Back (accuracy positive faces) | Game of Dice (winning bets) |
| 55 | En-Back (bias negative faces) | RAVLT (long delay score) |
| 56 | En-Back (bias neutral faces) | En-Back (mean RT to negative faces) |
| 57 | En-Back (bias positive faces) | En-Back (bias neutral faces) |

*Notes*. Variable importance rank was assessed using Elastic Net Regularisation (ENR) and Random Forest (RF) to identify the most salient cognitive differences between the ELA and NOA groups. Differences between these two methods are outlined in the main text.

**Table S8**. Differences in cognition between the NOA-low and NOA-high subgroups

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | **SE Diff** | **Cohen's d** | **Lower CI** | **Upper CI** | **FDR-adj *p*** |  |
| Picture Vocabulary | -49.31 | -18.82 | 0.38 | -1.48 | -1.55 | -1.41 | <.001 | \*\*\* |
| Oral Reading Recognition | -38.80 | -17.07 | 0.44 | -1.18 | -1.25 | -1.11 | <.001 | \*\*\* |
| En-Back (accuracy neutral faces) | -31.75 | -0.08 | 0.00 | -0.80 | -0.86 | -0.74 | <.001 | \*\*\* |
| RAVLT (learning score) | -31.13 | -8.56 | 0.28 | -0.89 | -0.95 | -0.83 | <.001 | \*\*\* |
| RAVLT (long delay score) | -30.72 | -2.26 | 0.07 | -0.89 | -0.96 | -0.83 | <.001 | \*\*\* |
| En-Back (accuracy positive faces) | -29.83 | -0.08 | 0.00 | -0.76 | -0.82 | -0.70 | <.001 | \*\*\* |
| RAVLT (short delay score) | -30.25 | -2.10 | 0.07 | -0.88 | -0.95 | -0.82 | <.001 | \*\*\* |
| List Sort Working Memory | -29.74 | -11.71 | 0.39 | -0.89 | -0.96 | -0.83 | <.001 | \*\*\* |
| En-Back (accuracy negative faces) | -28.55 | -0.08 | 0.00 | -0.74 | -0.80 | -0.68 | <.001 | \*\*\* |
| Matrix Reasoning (total score) | -28.07 | -2.23 | 0.08 | -0.84 | -0.91 | -0.78 | <.001 | \*\*\* |
| Picture Sequence Memory | -25.98 | -12.03 | 0.46 | -0.78 | -0.85 | -0.72 | <.001 | \*\*\* |
| Game of Dice (account balance) | -21.22 | -1907.30 | 89.89 | -0.59 | -0.66 | -0.53 | <.001 | \*\*\* |
| RAVLT (distraction score) | -21.02 | -1.07 | 0.05 | -0.64 | -0.70 | -0.58 | <.001 | \*\*\* |
| DDIS (6h delay indifference) | -20.45 | -10.27 | 0.50 | -0.52 | -0.58 | -0.46 | <.001 | \*\*\* |
| Game of Dice (safe vs. risky bets) | -18.64 | -5.16 | 0.28 | -0.55 | -0.61 | -0.49 | <.001 | \*\*\* |
| Pattern Comparison Processing Speed | -18.34 | -10.49 | 0.57 | -0.55 | -0.61 | -0.49 | <.001 | \*\*\* |
| DDIS (1d delay indifference) | -17.62 | -9.84 | 0.56 | -0.47 | -0.53 | -0.41 | <.001 | \*\*\* |
| En-Back (SD of RT to neutral faces) | 17.48 | 31.11 | 1.78 | 0.52 | 0.46 | 0.58 | <.001 | \*\*\* |
| En-Back (SD of RT to positive faces) | 17.44 | 30.40 | 1.74 | 0.52 | 0.46 | 0.58 | <.001 | \*\*\* |
| Dimensional Change Card Sort | -17.12 | -8.07 | 0.47 | -0.53 | -0.59 | -0.47 | <.001 | \*\*\* |
| DDIS (1wk delay indifference) | -15.86 | -10.99 | 0.69 | -0.44 | -0.50 | -0.37 | <.001 | \*\*\* |
| Flanker Inhibitory Control | -15.65 | -6.37 | 0.41 | -0.48 | -0.54 | -0.42 | <.001 | \*\*\* |
| En-Back (SD of RT to negative faces) | 15.25 | 26.52 | 1.74 | 0.45 | 0.39 | 0.51 | <.001 | \*\*\* |
| DDIS (1mo delay indifference) | -14.66 | -11.84 | 0.81 | -0.41 | -0.47 | -0.35 | <.001 | \*\*\* |
| EF Stroop (50% incongruent happy accuracy) | -14.15 | -0.04 | 0.00 | -0.39 | -0.45 | -0.33 | <.001 | \*\*\* |
| Game of Dice (winning bets) | -14.15 | -1.08 | 0.08 | -0.41 | -0.47 | -0.35 | <.001 | \*\*\* |
| Game of Dice (losing bets) | 14.13 | 1.08 | 0.08 | 0.41 | 0.35 | 0.47 | <.001 | \*\*\* |
| EF Stroop (50% incongruent angry accuracy) | -13.86 | -0.04 | 0.00 | -0.38 | -0.44 | -0.32 | <.001 | \*\*\* |
| EF Stroop (50% congruent angry RT) | 13.78 | 64.34 | 4.67 | 0.41 | 0.35 | 0.47 | <.001 | \*\*\* |
| EF Stroop (50% congruent happy RT) | 13.72 | 63.06 | 4.60 | 0.41 | 0.35 | 0.47 | <.001 | \*\*\* |
| EF Stroop (75% congruent happy RT) | 12.36 | 52.88 | 4.28 | 0.37 | 0.31 | 0.43 | <.001 | \*\*\* |
| EF Stroop (50% incongruent angry RT) | 12.30 | 60.65 | 4.93 | 0.37 | 0.31 | 0.43 | <.001 | \*\*\* |
| En-Back (mean RT to positive faces) | 12.20 | 42.15 | 3.45 | 0.36 | 0.30 | 0.42 | <.001 | \*\*\* |
| EF Stroop (75% congruent angry accuracy) | -11.61 | -0.02 | 0.00 | -0.31 | -0.37 | -0.25 | <.001 | \*\*\* |
| EF Stroop (25% incongruent happy accuracy) | -11.60 | -0.04 | 0.00 | -0.32 | -0.38 | -0.26 | <.001 | \*\*\* |
| EF Stroop (50% congruent angry accuracy) | -11.51 | -0.03 | 0.00 | -0.32 | -0.38 | -0.26 | <.001 | \*\*\* |
| EF Stroop (50% congruent happy accuracy) | -11.48 | -0.03 | 0.00 | -0.32 | -0.38 | -0.26 | <.001 | \*\*\* |
| EF Stroop (50% incongruent happy RT) | 11.49 | 55.40 | 4.82 | 0.34 | 0.28 | 0.40 | <.001 | \*\*\* |
| EF Stroop (75% congruent happy accuracy) | -11.11 | -0.02 | 0.00 | -0.30 | -0.36 | -0.24 | <.001 | \*\*\* |
| EF Stroop (25% incongruent angry accuracy) | -10.95 | -0.04 | 0.00 | -0.31 | -0.37 | -0.25 | <.001 | \*\*\* |
| En-Back (mean RT to neutral faces) | 10.87 | 37.68 | 3.47 | 0.32 | 0.26 | 0.38 | <.001 | \*\*\* |
| EF Stroop (75% congruent angry RT) | 10.78 | 49.04 | 4.55 | 0.32 | 0.26 | 0.38 | <.001 | \*\*\* |
| DDIS (immediate choice validity) | 10.72 | 0.29 | 0.03 | 0.32 | 0.26 | 0.38 | <.001 | \*\*\* |
| DDIS (3mo delay indifference) | -10.48 | -9.43 | 0.90 | -0.30 | -0.36 | -0.24 | <.001 | \*\*\* |
| En-Back (mean RT to negative faces) | 10.35 | 35.96 | 3.48 | 0.31 | 0.25 | 0.37 | <.001 | \*\*\* |
| EF Stroop (25% incongruent angry RT) | 9.69 | 56.23 | 5.81 | 0.29 | 0.23 | 0.35 | <.001 | \*\*\* |
| EF Stroop (25% incongruent happy RT) | 9.30 | 54.94 | 5.91 | 0.28 | 0.22 | 0.34 | <.001 | \*\*\* |
| DDIS (immediate choice RT) | -8.53 | -257.65 | 30.21 | -0.26 | -0.32 | -0.20 | <.001 | \*\*\* |
| Little Man Task (RT) | 7.19 | 89.83 | 12.50 | 0.21 | 0.15 | 0.27 | <.001 | \*\*\* |
| DDIS (1yr delay indifference) | -7.04 | -7.02 | 1.00 | -0.21 | -0.27 | -0.15 | <.001 | \*\*\* |
| Cash Choice Task | -2.98 | -0.05 | 0.02 | -0.09 | -0.15 | -0.03 | 0.003 | \*\* |
| En-Back (bias negative faces) | 2.51 | 0.02 | 0.01 | 0.08 | 0.02 | 0.14 | 0.013 |  |
| DDIS (delayed choice RT) | -2.17 | -49.20 | 22.65 | -0.06 | -0.12 | 0.00 | 0.032 |  |
| En-Back (bias positive faces) | 1.33 | 0.01 | 0.01 | 0.04 | -0.02 | 0.10 | 0.195 |  |
| DDIS (5yr delay indifference) | -0.78 | -0.80 | 1.02 | -0.02 | -0.08 | 0.04 | 0.441 |  |
| Little Man Task (score) | -0.79 | -0.11 | 0.13 | -0.02 | -0.08 | 0.04 | 0.446 |  |
| En-Back (bias neutral faces) | 0.68 | 0.00 | 0.01 | 0.02 | -0.04 | 0.08 | 0.497 |  |

*Notes*. Welch's t-test. Lower and upper CI represent the 95% confidence interval for Cohen's d. P-values adjusted for false-discovery using the Hochberg procedure with a significance value of .01. \*\**p* < .01. \*\*\**p* < .001.

**Table S9**. Differences in cognition between the ELA-low and ELA-high subgroups

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | **SE Diff** | **Cohen’s d** | **Lower CI** | **Upper CI** | **FDR-adj *p*** |  |
| Picture Vocabulary | 20.29 | 16.75 | 0.83 | 1.57 | 1.36 | 1.77 | <.001 | \*\*\* |
| Oral Reading Recognition | 18.96 | 18.15 | 0.96 | 1.49 | 1.29 | 1.69 | <.001 | \*\*\* |
| List Sort Working Memory | 13.57 | 13.67 | 1.01 | 1.13 | 0.94 | 1.32 | <.001 | \*\*\* |
| Matrix Reasoning (total score) | 13.07 | 2.84 | 0.22 | 1.12 | 0.92 | 1.31 | <.001 | \*\*\* |
| RAVLT (short delay score) | 10.97 | 2.13 | 0.19 | 0.92 | 0.73 | 1.10 | <.001 | \*\*\* |
| RAVLT (long delay score) | 11.05 | 2.42 | 0.22 | 0.96 | 0.77 | 1.15 | <.001 | \*\*\* |
| En-Back (accuracy neutral faces) | 10.84 | 0.17 | 0.02 | 1.06 | 0.86 | 1.26 | <.001 | \*\*\* |
| En-Back (accuracy negative faces) | 10.82 | 0.17 | 0.02 | 1.07 | 0.86 | 1.27 | <.001 | \*\*\* |
| RAVLT (learning score) | 10.18 | 8.82 | 0.87 | 0.88 | 0.69 | 1.07 | <.001 | \*\*\* |
| Dimensional Change Card Sort | 9.64 | 9.55 | 0.99 | 0.77 | 0.58 | 0.95 | <.001 | \*\*\* |
| En-Back (accuracy positive faces) | 10.15 | 0.17 | 0.02 | 1.01 | 0.81 | 1.21 | <.001 | \*\*\* |
| Picture Sequence Memory | 9.11 | 11.02 | 1.21 | 0.76 | 0.57 | 0.94 | <.001 | \*\*\* |
| RAVLT (distraction score) | 7.19 | 0.99 | 0.14 | 0.61 | 0.43 | 0.79 | <.001 | \*\*\* |
| Flanker inhibitory Control | 6.76 | 8.04 | 1.19 | 0.61 | 0.43 | 0.79 | <.001 | \*\*\* |
| Pattern Comparison Processing Speed | 6.18 | 9.99 | 1.62 | 0.54 | 0.36 | 0.72 | <.001 | \*\*\* |
| EF Stroop (75% congruent happy RT) | -6.17 | -81.23 | 13.17 | -0.54 | -0.72 | -0.37 | <.001 | \*\*\* |
| EF Stroop (75% congruent angry RT) | -6.10 | -80.10 | 13.13 | -0.53 | -0.70 | -0.35 | <.001 | \*\*\* |
| DDIS (5yr delay indifference) | -5.82 | -18.94 | 3.25 | -0.52 | -0.70 | -0.34 | <.001 | \*\*\* |
| EF Stroop (25% incongruent happy RT) | -5.64 | -99.88 | 17.72 | -0.49 | -0.67 | -0.31 | <.001 | \*\*\* |
| En-Back (mean RT to positive faces) | -5.31 | -65.85 | 12.40 | -0.49 | -0.66 | -0.31 | <.001 | \*\*\* |
| EF Stroop (50% congruent happy RT) | -5.28 | -74.97 | 14.21 | -0.47 | -0.65 | -0.29 | <.001 | \*\*\* |
| En-Back (mean RT to negative faces) | -4.88 | -61.94 | 12.70 | -0.45 | -0.62 | -0.27 | <.001 | \*\*\* |
| EF Stroop (50% congruent angry RT) | -4.61 | -70.58 | 15.30 | -0.41 | -0.58 | -0.23 | <.001 | \*\*\* |
| EF Stroop (50% incongruent angry accuracy) | 4.61 | 0.06 | 0.01 | 0.45 | 0.27 | 0.63 | <.001 | \*\*\* |
| EF Stroop (50% incongruent happy RT) | -4.33 | -66.93 | 15.46 | -0.38 | -0.55 | -0.20 | <.001 | \*\*\* |
| EF Stroop (25% incongruent angry RT) | -4.26 | -70.00 | 16.45 | -0.36 | -0.53 | -0.18 | <.001 | \*\*\* |
| DDIS (6h delay indifference) | 4.09 | 11.39 | 2.79 | 0.38 | 0.20 | 0.55 | <.001 | \*\*\* |
| DDIS (1yr delay indifference) | -4.01 | -12.94 | 3.23 | -0.35 | -0.53 | -0.18 | <.001 | \*\*\* |
| EF Stroop (50% congruent happy accuracy) | 4.02 | 0.03 | 0.01 | 0.37 | 0.20 | 0.55 | <.001 | \*\*\* |
| EF Stroop (50% incongruent happy accuracy) | 3.96 | 0.05 | 0.01 | 0.36 | 0.18 | 0.53 | <.001 | \*\*\* |
| EF Stroop (75% congruent angry accuracy) | 3.94 | 0.04 | 0.01 | 0.38 | 0.20 | 0.55 | <.001 | \*\*\* |
| DDIS (immediate choice RT) | 3.83 | 317.05 | 82.85 | 0.34 | 0.17 | 0.52 | <.001 | \*\*\* |
| En-Back (mean RT to neutral faces) | -3.82 | -44.09 | 11.55 | -0.35 | -0.53 | -0.18 | <.001 | \*\*\* |
| EF Stroop (75% congruent happy accuracy) | 3.62 | 0.03 | 0.01 | 0.33 | 0.16 | 0.51 | 0.001 | \*\* |
| EF Stroop (25% incongruent angry accuracy) | 3.57 | 0.06 | 0.02 | 0.33 | 0.16 | 0.51 | 0.001 | \*\* |
| DDIS (1d delay indifference) | 3.14 | 8.42 | 2.68 | 0.29 | 0.11 | 0.46 | 0.003 | \*\* |
| EF Stroop (25% incongruent happy accuracy) | 2.83 | 0.04 | 0.02 | 0.26 | 0.09 | 0.44 | 0.008 | \*\* |
| EF Stroop (50% congruent angry accuracy) | 2.72 | 0.03 | 0.01 | 0.25 | 0.07 | 0.42 | 0.011 |  |
| DDIS (3mo delay indifference) | -2.60 | -8.11 | 3.11 | -0.23 | -0.40 | -0.05 | 0.015 |  |
| En-Back (SD of RT to positive faces) | -2.52 | -16.29 | 6.47 | -0.23 | -0.41 | -0.06 | 0.019 |  |
| En-Back (SD of RT to neutral faces) | -2.45 | -15.30 | 6.26 | -0.22 | -0.39 | -0.05 | 0.021 |  |
| EF Stroop (50% incongruent angry RT) | -2.31 | -38.46 | 16.67 | -0.21 | -0.39 | -0.04 | 0.030 |  |
| DDIS (immediate choice validity) | 2.26 | 0.21 | 0.09 | 0.20 | 0.03 | 0.38 | 0.033 |  |
| Game of Dice (account balance) | 2.10 | 757.74 | 360.74 | 0.19 | 0.01 | 0.36 | 0.048 |  |
| Game of Dice (safe vs. risky bets) | 2.09 | 1.80 | 0.87 | 0.18 | 0.01 | 0.35 | 0.048 |  |
| Cash Choice Task | -2.01 | -0.10 | 0.05 | -0.18 | -0.35 | -0.01 | 0.056 |  |
| Game of Dice (winning bets) | 1.88 | 0.48 | 0.26 | 0.16 | -0.01 | 0.34 | 0.074 |  |
| DDIS (delayed choice RT) | 1.56 | 127.93 | 82.06 | 0.15 | -0.03 | 0.32 | 0.143 |  |
| DDIS (1mo delay indifference) | -1.44 | -4.32 | 3.00 | -0.13 | -0.30 | 0.05 | 0.174 |  |
| Game of Dice (losing bets) | -1.30 | -0.34 | 0.26 | -0.11 | -0.29 | 0.06 | 0.225 |  |
| En-Back (SD of RT to negative faces) | -1.08 | -7.64 | 7.06 | -0.10 | -0.28 | 0.07 | 0.314 |  |
| En-Back (bias negative faces) | -0.80 | -0.02 | 0.02 | -0.07 | -0.24 | 0.10 | 0.466 |  |
| DDIS (1wk delay indifference) | 0.72 | 2.07 | 2.87 | 0.06 | -0.11 | 0.24 | 0.507 |  |
| Little Man Task (RT) | 0.46 | 24.53 | 53.19 | 0.04 | -0.13 | 0.22 | 0.681 |  |
| Little Man Task (score) | -0.20 | -0.11 | 0.55 | -0.02 | -0.19 | 0.16 | 0.868 |  |
| En-Back (bias neutral faces) | 0.07 | 0.00 | 0.02 | 0.01 | -0.17 | 0.18 | 0.948 |  |
| En-Back (bias positive faces) | 0.07 | 0.00 | 0.02 | 0.01 | -0.17 | 0.18 | 0.959 |  |

*Notes*. Welch’s t-test. Lower and upper CI represent the 95% confidence interval for Cohen’s d. P-values adjusted for false-discovery using the Hochberg procedure with a significance value of .01. \*\**p* < .01. \*\*\**p* < .001.

**Table S10**. Differences in mental health between the NOA-low and NOA-high subgroups

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | **SE Diff** | **Cohen's d** | **Lower CI** | **Upper CI** | **FDR-adj *p*** |  |
| Psychotic Symptoms | 8.87 | 1.43 | 0.16 | 0.25 | 0.19 | 0.31 | <.001 | \*\*\* |
| Attention Problems | 6.58 | 0.93 | 0.14 | 0.20 | 0.14 | 0.27 | <.001 | \*\*\* |
| Conduct Disorder | 5.79 | 0.66 | 0.12 | 0.18 | 0.11 | 0.24 | <.001 | \*\*\* |
| Rule Breaking | 5.63 | 0.55 | 0.10 | 0.17 | 0.11 | 0.23 | <.001 | \*\*\* |
| ADHD | 5.39 | 0.76 | 0.14 | 0.17 | 0.10 | 0.23 | <.001 | \*\*\* |
| Social Problems | 5.42 | 0.63 | 0.12 | 0.17 | 0.10 | 0.23 | <.001 | \*\*\* |
| Externalising Composite | 3.94 | 1.10 | 0.28 | 0.13 | 0.06 | 0.19 | <.001 | \*\*\* |
| Aggressive Behaviour | 3.49 | 0.42 | 0.12 | 0.11 | 0.04 | 0.17 | 0.001 | \*\*\* |
| Oppositional-Defiant Disorder | 3.38 | 0.45 | 0.13 | 0.11 | 0.04 | 0.17 | 0.002 | \*\* |
| Stress Problems | 2.89 | 0.41 | 0.14 | 0.09 | 0.03 | 0.15 | 0.008 | \*\* |
| Total Problems | 2.80 | 0.92 | 0.33 | 0.09 | 0.03 | 0.15 | 0.009 | \*\* |
| Somatic Disorder | 2.76 | 0.51 | 0.18 | 0.09 | 0.02 | 0.15 | 0.010 | \*\* |
| Somatic Complains | 2.43 | 0.40 | 0.17 | 0.08 | 0.01 | 0.14 | 0.023 |  |
| Depressive Disorder | 1.14 | 0.19 | 0.16 | 0.04 | -0.03 | 0.10 | 0.356 |  |
| Anxiety Disorder | 0.71 | 0.12 | 0.17 | 0.02 | -0.04 | 0.09 | 0.558 |  |
| Withdrawn-Depressed | 0.61 | 0.10 | 0.16 | 0.02 | -0.04 | 0.08 | 0.598 |  |
| Sluggish Cognitive Tempo | 0.24 | 0.03 | 0.14 | 0.01 | -0.06 | 0.07 | 0.850 |  |
| Thought Problems | -0.23 | -0.04 | 0.16 | -0.01 | -0.07 | 0.06 | 0.816 |  |
| Anxious-Depressed | -0.86 | -0.14 | 0.17 | -0.03 | -0.09 | 0.04 | 0.480 |  |
| Internalising Composite | -1.11 | -0.35 | 0.31 | -0.04 | -0.10 | 0.03 | 0.348 |  |
| Obsessive-Compulsive Problems | -2.54 | -0.43 | 0.17 | -0.08 | -0.15 | -0.02 | 0.02 |  |

*Notes*. Welch's t-test. Lower and upper CI represent the 95% confidence interval for Cohen's d. P-values adjusted for false-discovery using the Hochberg procedure with a significance value of .01. \*\**p* < .01. \*\*\**p* < .001.

**Table S11**. Differences in mental health between the ELA-low and ELA-high subgroups

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | **SE Diff** | **Cohen's d** | **Lower CI** | **Upper CI** | ***p*** |  |
| Withdrawn-Depressed | 3.39 | 2.21 | 0.65 | 0.31 | 0.11 | 0.50 | <.001 | \*\*\* |
| Conduct Disorder | -2.76 | -2.02 | 0.73 | -0.28 | -0.47 | -0.08 | 0.006 |  |
| Rule Breaking | -2.75 | -1.76 | 0.64 | -0.28 | -0.47 | -0.08 | 0.007 |  |
| Internalising Composite | 2.59 | 3.05 | 1.18 | 0.26 | 0.06 | 0.45 | 0.010 |  |
| Psychotic Symptoms | -1.71 | -2.00 | 1.17 | -0.16 | -0.34 | 0.01 | 0.088 |  |
| Depressive Disorder | 1.55 | 1.19 | 0.76 | 0.15 | -0.05 | 0.34 | 0.122 |  |
| Somatic Complains | 1.49 | 1.11 | 0.75 | 0.15 | -0.05 | 0.34 | 0.137 |  |
| Attention Problems | -1.48 | -1.22 | 0.83 | -0.15 | -0.34 | 0.05 | 0.140 |  |
| Externalising Composite | -1.44 | -1.70 | 1.17 | -0.15 | -0.34 | 0.05 | 0.150 |  |
| Somatic Disorder | 1.26 | 0.96 | 0.76 | 0.12 | -0.07 | 0.32 | 0.208 |  |
| Anxious-Depressed | 1.25 | 0.93 | 0.74 | 0.12 | -0.07 | 0.32 | 0.213 |  |
| Obsessive-Compulsive Problems | 1.14 | 0.84 | 0.74 | 0.11 | -0.08 | 0.30 | 0.257 |  |
| Aggressive Behaviour | -0.97 | -0.71 | 0.73 | -0.09 | -0.29 | 0.10 | 0.333 |  |
| Anxiety Disorder | 0.94 | 0.72 | 0.77 | 0.09 | -0.10 | 0.28 | 0.346 |  |
| Oppositional-Defiant Disorder | -0.83 | -0.58 | 0.70 | -0.08 | -0.28 | 0.11 | 0.405 |  |
| Sluggish Cognitive Tempo | 0.75 | 0.51 | 0.69 | 0.07 | -0.12 | 0.27 | 0.457 |  |
| ADHD | -0.74 | -0.52 | 0.71 | -0.07 | -0.27 | 0.12 | 0.458 |  |
| Total Problems | 0.60 | 0.79 | 1.33 | 0.06 | -0.13 | 0.26 | 0.551 |  |
| Social Problems | -0.49 | -0.32 | 0.66 | -0.05 | -0.24 | 0.15 | 0.628 |  |
| Thought Problems | 0.25 | 0.20 | 0.77 | 0.03 | -0.17 | 0.22 | 0.800 |  |
| Stress Problems | 0.23 | 0.17 | 0.75 | 0.02 | -0.17 | 0.22 | 0.822 |  |

*Notes*. Welch's t-test. Lower and upper CI represent the 95% confidence interval for Cohen's d. P-value significance assessed using Bonferroni correction (α < .002). \*\*\**p* < .001.

**Table S12**. MANOVA post-hoc tests for group by subgroup interaction on mental health

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Mean Sq** | **F** | ***p*** |  |
| Withdrawn-Depressed | 492.10 | 16.53 | <.001 | \*\*\* |
| Rule Breaking | 135.50 | 9.30 | 0.002 | \*\* |
| Conduct Disorder | 170.40 | 8.57 | 0.003 | \*\* |
| Somatic Complains | 213.10 | 6.69 | 0.009 | \*\* |
| Internalising Composite | 676.50 | 6.52 | 0.011 | \* |
| Depressive Disorder | 174.10 | 5.49 | 0.019 | \* |
| Somatic Disorder | 200.50 | 5.33 | 0.021 | \* |
| Total Problems | 272.00 | 2.34 | 0.126 |  |
| Anxiety Disorder | 65.30 | 1.98 | 0.159 |  |
| Anxious-Depressed | 57.30 | 1.82 | 0.178 |  |
| Sluggish Cognitive Tempo | 27.40 | 1.20 | 0.274 |  |
| Stress Problems | 31.30 | 1.14 | 0.285 |  |
| Obsessive-Compulsive Problems | 15.83 | 0.52 | 0.472 |  |
| Social Problems | 9.00 | 0.46 | 0.500 |  |
| Externalising Composite | 32.70 | 0.37 | 0.541 |  |
| Aggressive Behaviour | 8.10 | 0.37 | 0.542 |  |
| Attention Problems | 8.10 | 0.29 | 0.590 |  |
| ADHD | 5.00 | 0.19 | 0.659 |  |
| Thought Problems | 2.30 | 0.08 | 0.780 |  |
| Oppositional-Defiant Disorder | 1.60 | 0.07 | 0.787 |  |
| Psychotic Symptoms | 1.50 | 0.03 | 0.861 |  |

*Notes*. MANOVA post-hoc tests for interaction effect of group (ELA vs NOA) on subgroup (low vs high) for mental health. \**p* < .05. \*\**p* < .01. \*\*\**p* < .001.

**Table S13**. Supplemental analysis of differences in mental health between the NOA-low and NOA-high subgroups with sub-sampling

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | ***t*** | **Mean Diff** | **SE Diff** | **Cohen's d** | **Lower CI** | **Upper CI** | **FDR-adj *p*** |  |
| Externalising Composite | 4.28 | 3.23 | 0.76 | 0.36 | 0.19 | 0.53 | <.001 | \*\*\* |
| Rule Breaking | 4.19 | 1.07 | 0.26 | 0.33 | 0.16 | 0.50 | <.001 | \*\*\* |
| Conduct Disorder | 3.80 | 1.20 | 0.32 | 0.30 | 0.13 | 0.48 | <.001 | \*\* |
| Total Problems | 3.25 | 2.99 | 0.92 | 0.28 | 0.11 | 0.45 | 0.009 | \*\* |
| Social Problems | 3.19 | 1.03 | 0.32 | 0.26 | 0.09 | 0.43 | 0.007 | \*\* |
| Aggressive Behaviour | 2.99 | 1.03 | 0.35 | 0.24 | 0.07 | 0.41 | 0.010 | \*\* |
| Oppositional-Defiant Disorder | 2.97 | 1.09 | 0.37 | 0.24 | 0.07 | 0.42 | 0.010 | \*\* |
| ADHD | 2.84 | 1.11 | 0.39 | 0.24 | 0.06 | 0.41 | 0.013 |  |
| Attention Problems | 2.74 | 1.13 | 0.41 | 0.23 | 0.06 | 0.40 | 0.015 |  |
| Psychotic Symptoms | 2.60 | 1.27 | 0.49 | 0.20 | 0.04 | 0.36 | 0.020 |  |
| Stress Problems | 2.07 | 0.85 | 0.41 | 0.17 | 0.00 | 0.35 | 0.074 |  |
| Somatic Complains | 1.79 | 0.81 | 0.45 | 0.15 | -0.02 | 0.32 | 0.131 |  |
| Somatic Disorder | 1.70 | 0.86 | 0.50 | 0.15 | -0.03 | 0.32 | 0.144 |  |
| Thought Problems | 1.56 | 0.72 | 0.46 | 0.13 | -0.04 | 0.30 | 0.179 |  |
| Depressive Disorder | 1.29 | 0.58 | 0.45 | 0.11 | -0.06 | 0.28 | 0.275 |  |
| Withdrawn-Depressed | 1.29 | 0.54 | 0.42 | 0.11 | -0.06 | 0.28 | 0.259 |  |
| Sluggish Cognitive Tempo | 0.46 | 0.18 | 0.39 | 0.04 | -0.13 | 0.21 | 0.795 |  |
| Internalising Composite | 0.43 | 0.38 | 0.87 | 0.04 | -0.13 | 0.21 | 0.776 |  |
| Anxiety Disorder | 0.34 | 0.16 | 0.48 | 0.03 | -0.14 | 0.20 | 0.810 |  |
| Anxious-Depressed | 0.02 | 0.01 | 0.48 | 0.00 | -0.17 | 0.17 | 1.036 |  |
| Obsessive-Compulsive Problems | 0.01 | 0.00 | 0.48 | 0.00 | -0.17 | 0.17 | 0.996 |  |

*Notes*. Supplemental analysis performed on randomly selected NOA sample of *n*= 739 to match the sample size of ELA group. Welch's t-test. Lower and upper CI represent the 95% confidence interval for Cohen's d. P-values adjusted for false-discovery using the Hochberg procedure with a significance value of .01. \*\**p* < .01. \*\*\**p* < .001.

1. **FIGURES SUPPLEMENT**

**Figure S1.** Differences in cognition between the ELA and NOA groups

*Chart

Description automatically generated*

*Notes*. Higher scores indicate greater accuracy or longer response time. Error bars represent the 95% confidence interval for Cohen's d. Game of Dice (losing bets) was multiplied by -1 so that higher scores represent better performance in the figure. Only significant differences shown.

**Figure S2.** Differences in mental health between the ELA and NOA groups

Chart, bar chart

Description automatically generated

*Notes*. Higher scores indicate greater symptoms. Error bars represent the 95% confidence interval for Cohen's d. All 21 measures were significant, including the prodromal psychosis scale which is not shown in figure due to differences in scale.

**Figure S3**. Hierarchical clustering analysis indicated an optimal 2-cluster solution for the NOA group

Chart, line chart

Description automatically generated

Chart, line chart

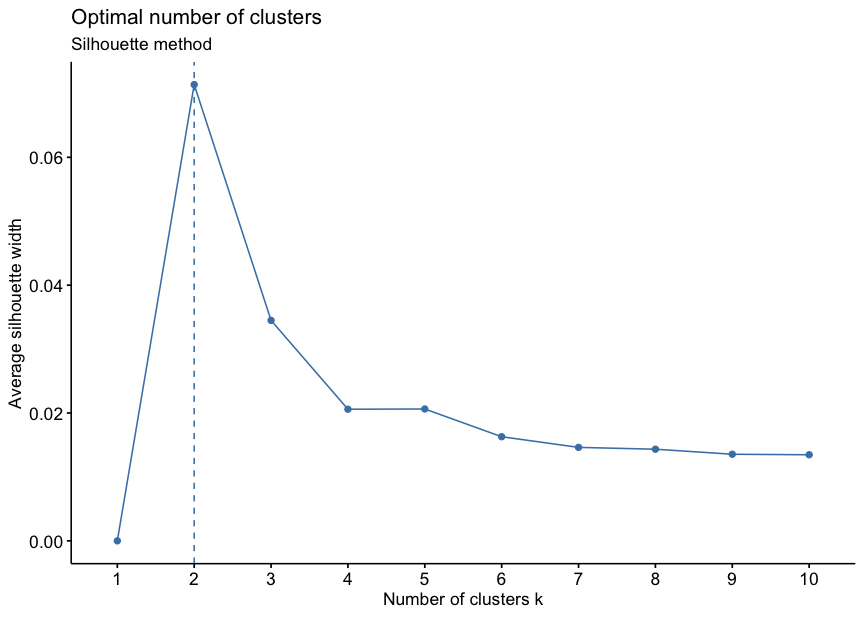
Description automatically generated

Chart, histogram

Description automatically generated

**Figure S4.** Hierarchical clustering analysis indicated an optimal 2-cluster solution for the ELA group

Chart, line chart

Description automatically generated

Chart, histogram

Description automatically generated

**Figure S5**. Hierarchical cluster internal validation measures indicated an optimal 2-cluster solution for the NOA group

**Chart

Description automatically generated**

*Notes*. Hierarchical clustering internal validation measures compared across different three clustering methods: hierarchical, k-means and pam. Connectivity= Connectivity indicates the degree of connectedness of the clusters, as determined by the k-nearest neighbors. Lower connectivity values indicate superior clustering solution. Dunn= Dunn Index is the ratio between the smallest distance between observations not in the same cluster to the largest intra-cluster distance. Higher Dunn values indicate superior clustering solution. Silhouette= Silhouette width measures the degree of confidence in a particular clustering assignment. Higher Silhouette values indicate superior clustering solution.

**Figure S6**. Hierarchical cluster internal validation measures indicated an optimal 2-cluster solution for the ELA group

Chart

Description automatically generated

*Notes*. Hierarchical clustering internal validation measures compared across different three clustering methods: hierarchical, k-means and pam. Connectivity= Connectivity indicates the degree of connectedness of the clusters, as determined by the k-nearest neighbors. Lower connectivity values indicate superior clustering solution. Dunn= Dunn Index is the ratio between the smallest distance between observations not in the same cluster to the largest intra-cluster distance. Higher Dunn values indicate superior clustering solution. Silhouette= Silhouette width measures the degree of confidence in a particular clustering assignment. Higher Silhouette values indicate superior clustering solution.

**Supplementary References**

Bernaards CA, Belin TR, Schafer JL. Robustness of a multivariate normal approximation for imputation of incomplete binary data. Stat Med2007; 26:1368-82

Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. Am Stat, 2007;61:79-90.

Lee, K. J., & Carlin, J. B. (2017). Multiple imputation in the presence of non-normal data. *Statistics in medicine*, *36*(4), 606–617. <https://doi.org/10.1002/sim.7173>

Sterne J A C, White I R, Carlin J B, Spratt M, Royston P, Kenward M G et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls BMJ 2009; 338: b2393 doi:10.1136/bmj.b2393

Sullivan, T.R., Lee, K.J., Ryan, P. *et al.* Multiple imputation for handling missing outcome data when estimating the relative risk. *BMC Med Res Methodol* **17**, 134 (2017). https://doi.org/10.1186/s12874-017-0414-5

Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med, 1999;18:681-94.