doi:10.1093/brain/awaa314 BRAIN 2020: 143; 3865–3877 | 3865

BRAIN
A JOURNAL OF NEUROLOGY

Disentangling vulnerability, state and trait features of neurocognitive impairments in depression

Yuen-Siang Ang,[1,2] Nicole Frontero,[2] Emily Belleau[1,2] and Diego A. Pizzagalli[1,2,3]

Depression is a debilitating disorder that often starts manifesting in early childhood and peaks in onset during adolescence. Neurocognitive impairments have emerged as clinically important characteristics of depression, but it remains controversial which domains specifically index pre-existing vulnerability, state-related or trait-related markers. Here, we disentangled these effects by analysing the Adolescent Brain Cognitive Development dataset ($n = 4626$). Using information of participants' current and past mental disorders, as well as family mental health history, we identified low-risk healthy ($n = 2100$), high-risk healthy ($n = 2023$), remitted depressed ($n = 401$) and currently depressed children ($n = 102$). Factor analysis of 11 cognitive variables was performed to elucidate latent structure and canonical correlation analyses conducted to probe regional brain volumes reliably associated with the cognitive factors. Bayesian model comparison of various *a priori* hypotheses differing in how low-risk healthy, high-risk healthy, remitted depressed and currently depressed children performed in various cognitive domains was performed. Factor analysis revealed three domains: language and reasoning, cognitive flexibility and memory recall. Deficits in language and reasoning ability, as well as in volumes of associated regions such as the middle temporal and superior frontal gyrus, represented state- and trait-related markers of depression but not pre-existing vulnerability. In contrast, there was no compelling evidence of impairments in other domains. These findings—although cross-sectional and specific to 9–10-year-old children—might have important clinical implications, suggesting that cognitive dysfunction may not be useful targets of preventive interventions. Depressed patients, even after remission, might also benefit from less commonly used treatments such as cognitive remediation therapy.

1 Department of Psychiatry, Harvard Medical School, Boston, MA 02115, USA
2 Center for Depression, Anxiety and Stress Research, McLean Hospital, Belmont, MA 02478, USA
3 McLean Imaging Center, McLean Hospital, Belmont, MA 02478, USA

Correspondence to: Diego A. Pizzagalli, PhD
Department of Psychiatry, Harvard Medical School, McLean Hospital, 115 Mill St, Room 233C, Belmont, MA 02478, USA
E-mail: dap@mclean.harvard.edu

Keywords: depression; affective disorders; child psychiatry; imaging; computational psychiatry

Abbreviations: ABCD = Adolescent Brain Cognitive Development; CCA = canonical correlation analysis; CD = currently depressed; HRH = high-risk healthy; KSADS-5 = Kiddie Schedule for Affective Disorders and Schizophrenia for DSM-5; LRH = low-risk healthy; RD = remitted depressed

Received April 15, 2020. Revised July 15, 2020. Accepted July 31, 2020. Advance access publication November 11, 2020

Introduction

Depression is a debilitating psychiatric illness affecting more than 300 million individuals globally (Smith, 2014). This prevalent disorder often starts manifesting in early childhood (Birmaher *et al.*, 2004) and peaks in onset during adolescence (Hankin *et al.*, 1998; Lewinsohn *et al.*, 1999; Kessler *et al.*, 2005; Thapar *et al.*, 2012). Highlighting familiarity, offspring of depressed parents are three to four times more likely to suffer from depression than those of healthy parents (Sullivan *et al.*, 2000; Rice *et al.*, 2002; Rasic *et al.*, 2014). To develop effective interventions, it is crucial to identify and distinguish between vulnerability markers that predict risk for developing the disorder, state markers that reflect pathophysiological processes of the illness, and trait markers that represent barriers to recovery.

A promising, yet highly debated, marker implicated in all three mechanisms is disrupted cognition (Allott *et al.*, 2016). Case-control studies have highlighted widespread deficits in domains such as conceptual reasoning, verbal fluency and memory in both current (Austin *et al.*, 2001; McDermott and Ebmeier, 2009; Maalouf *et al.*, 2011; Lee *et al.*, 2012; McIntyre *et al.*, 2013; Snyder, 2013; Baune *et al.*, 2014; Rock *et al.*, 2014; Goodall *et al.*, 2018) and remitted depression (Smith *et al.*, 2006; Hasselbalch *et al.*, 2011; Bora *et al.*, 2013; Rock *et al.*, 2014; Semkovska *et al.*, 2019), but conflicting results exist (Korhonen *et al.*, 2002; Liu *et al.*, 2002; Robertson *et al.*, 2003; Clark *et al.*, 2005; Favre *et al.*, 2009; Klimkeit *et al.*, 2011; Peters *et al.*, 2017). Whether cognitive functioning may improve in remission relative to the acute phase is also equivocal (Maalouf *et al.*, 2010, 2011; Beaujean *et al.*, 2013; Schmid and Hammar, 2013; Bloch *et al.*, 2015). Moreover, these studies were insufficient to dissociate state- and trait-related deficiencies occurring as a consequence of the disorder from vulnerability-related impairments. In other words, might cognitive dysfunctions have already existed before the onset of depression? Investigations into this have provided mixed results (Allott *et al.*, 2016). Some studies have reported that unaffected first-degree relatives of depressed individuals have lower cognitive functioning compared to controls with no familial history of depression (Winters *et al.*, 1981; Christensen *et al.*, 2006; Belleau *et al.*, 2013; Hughes *et al.*, 2013; Hsu *et al.*, 2014; MacKenzie *et al.*, 2019), but others did not find any difference (Klimes-Dougan *et al.*, 2006; Micco *et al.*, 2009; Santucci *et al.*, 2014). Furthermore, individuals at familial risk for depression tend to exhibit higher levels of subclinical depressive symptoms (Christensen *et al.*, 2007), but many prior investigations failed to account for it (Winters *et al.*, 1981; Klimes-Dougan *et al.*, 2006; Micco *et al.*, 2009; Belleau *et al.*, 2013; Hughes *et al.*, 2013; Hsu *et al.*, 2014; MacKenzie *et al.*, 2019). Taken together, it remains unclear to what extent cognitive impairments might be specific vulnerability, state and trait markers of depression.

Previous studies have been limited by at least one of the following reasons. First, the sample sizes of most empirical studies were small and, thus, lacked statistical power to detect small effect sizes (Korhonen *et al.*, 2002; Liu *et al.*, 2002; Robertson *et al.*, 2003; Clark *et al.*, 2005; Klimes-Dougan *et al.*, 2006; Smith *et al.*, 2006; Favre *et al.*, 2009; Micco *et al.*, 2009; Maalouf *et al.*, 2011; Santucci *et al.*, 2014; Bloch *et al.*, 2015; Peters *et al.*, 2017) or be interpreted with great confidence (Winters *et al.*, 1981; Christensen *et al.*, 2006; Maalouf *et al.*, 2010, 2011; Belleau *et al.*, 2013; Schmid and Hammar, 2013; Hsu *et al.*, 2014; Bloch *et al.*, 2015). Meta-analyses that pool multiple studies together possess greater statistical power (McDermott and Ebmeier, 2009; Lee *et al.*, 2012; Bora *et al.*, 2013; Snyder, 2013; Rock *et al.*, 2014; MacKenzie *et al.*, 2019), but interpretation is restricted by the task impurity problem (Phillips, 1997; Austin *et al.*, 2001). That is, cognitive paradigms often operate across multiple domains (e.g. a set shifting task will involve attention, working memory, visual and spatial processing), but distinct tasks adopted in different studies were grouped together and assumed to measure one similar construct (rather than based on latent variable methods). There is also greater diversity in patients stemming from differences in specific inclusion/exclusion criteria.

Second, prior studies were never designed to disentangle vulnerability versus state versus trait impairments in depression. Instead, they tended to focus on only one aspect, comparing cognitive ability between depressed patients versus controls (i.e. state-related) (Korhonen *et al.*, 2002; Liu *et al.*, 2002; Robertson *et al.*, 2003; Favre *et al.*, 2009; Lee *et al.*, 2012; Snyder, 2013; Baune *et al.*, 2014; Rock *et al.*, 2014), remitted patients versus controls (i.e. trait-related) (Clark *et al.*, 2005; Smith *et al.*, 2006; Hasselbalch *et al.*, 2011; Bora *et al.*, 2013; Rock *et al.*, 2014; Bloch *et al.*, 2015; Peters *et al.*, 2017), or healthy individuals at high versus low familial risk (i.e. vulnerability-related) (Christensen *et al.*, 2006; Klimes-Dougan *et al.*, 2006; Micco *et al.*, 2009; Belleau *et al.*, 2013; Hsu *et al.*, 2014; Santucci *et al.*, 2014; MacKenzie *et al.*, 2019). To the best of our knowledge, no empirical study has explored how individuals in the low-risk, high-risk, remitted and depressed groups might perform relative to one another. Furthermore, all previous studies have adopted frequentist statistics, which suffer from the asymmetrical inference problem and can only test whether there is sufficient evidence to reject the null hypothesis, i.e. one can never conclude the alternative hypothesis is true regardless of *P*-value.

Third, significant, albeit not homogenous, evidence suggests that acute and remitted depression are associated with brain volume reduction in many regions including the fronto-cingulate cortex, temporal cortex, hippocampus, amygdala and basal ganglia (Koolschijn *et al.*, 2009; Lorenzetti *et al.*, 2009; Bora *et al.*, 2012). Individuals with family history of depression were also found to have smaller prefrontal and hippocampal volumes (Amico, 2011), as well

as thinner grey matter across widespread areas in the right hemisphere (Peterson *et al.*, 2009), compared to subjects with no depressed first-degree relatives. However, it is unclear how these brain structural abnormalities might contribute to putative vulnerability, state- and/or trait-related cognitive impairments in depression. A few studies have reported that worse attention (Leung *et al.*, 2009; Li *et al.*, 2010), set-shifting (Vasic *et al.*, 2008), and cognitive control (Jung *et al.*, 2014) were associated with specific structural abnormalities, but sample sizes were relatively modest (n = 15–50) and they examined only depressed patients versus controls (Vasic *et al.*, 2008; Leung *et al.*, 2009; Jung *et al.*, 2014), or remitted versus non-remitted patients (Li *et al.*, 2010). In a larger study (n = 131), Peterson *et al.* (2009) found that cortical thinning in the right hemisphere predicted lower attention and visual memory performance in individuals at high familial risk for depression compared to those at low risk. This suggested a potential vulnerability impairment, although it should be noted that a substantial proportion (~40%) of participants were either acutely depressed or in remission when tested.

Here, we overcame the aforementioned limitations and conducted the largest empirical study (n = 4626) to date to clarify, using a Bayesian inference approach, whether cognitive impairments might represent specific vulnerability, state and trait markers of depression; and if so, which domains might be most affected and whether there might be an association with regional abnormalities in brain volume. This was achieved by analysing data in the second annual curated release of the Adolescent Brain Cognitive Development (ABCD) study (Volkow *et al.*, 2018).

Materials and methods

Children aged 9–10 years old were recruited at 21 sites across the USA (Garavan *et al.*, 2018). A comprehensive protocol was administered and, here, we focused on the following.

Clinical assessments

Kiddie Schedule for Affective Disorders and Schizophrenia for DSM-5

The Kiddie Schedule for Affective Disorders and Schizophrenia for DSM-5 (KSADS-5) is a well-established, reliable and valid assessment of lifetime mental disorders in youths (Kaufman *et al.*, 1997; Barch *et al.*, 2018). Children provided self-reports in selected domains including mood and anxiety disorders, sleep and suicidality. See Supplementary material for details.

Family history assessment module screener

This instrument probes parents/guardians for the presence/absence of symptoms associated with depression, mania, psychosis, substance and alcohol use disorder, as well as antisocial personality disorder in all first and second-degree biological relatives of the child (Rice *et al.*, 1995).

Achenbach Child Behavior Checklist

Parents were asked to complete the Achenbach Child Behavior Checklist (CBCL), which assesses a wide range of emotional and behavioural problems in youth. T-scores for subscales were analysed (Achenbach and Rescorla, 2001).

Neurocognitive battery

Picture Vocabulary Test

This is a measure of receptive vocabulary in the NIH Toolbox (Gershon *et al.*, 2014). Each participant was presented with four pictures and heard an audio recording of a word. The goal was to select the picture that best depicted the meaning of the word. Age-correct standard scores were analysed.

Oral Reading Recognition Test

The Oral Reading Recognition Test is part of the NIH Toolbox and assessed participants' ability to pronounce and read printed letters or words on a screen (Gershon *et al.*, 2014). Age-corrected standard scores were analysed.

Flanker Task

The Flanker Task is a measure of executive attention and inhibitory control in the NIH Toolbox (Zelazo *et al.*, 2013). Age-corrected standard scores were analysed. See Supplementary material for details.

Dimension Change Card Sort Task

The Dimension Change Card Sort Task is a measure of cognitive flexibility in the NIH Toolbox (Zelazo *et al.*, 2013). Age-corrected standard scores were analysed. See Supplementary material for details.

Picture Sequence Test

In this NIH Toolbox task, participants were presented with a sequence of pictures one at a time (Dikmen *et al.*, 2014). Every picture stayed on the screen and was accompanied by an audio that briefly described its content. The images were then jumbled up and participants had to place each picture back in the correct sequence order. The Picture Sequence Test consisted of three trials and scores were derived by taking the number of adjacent picture pairs remembered correctly across the trials. Age-corrected standard scores were analysed.

List Sorting Test

The List Sorting Test is part of the NIH Toolbox and participants were presented with a series of stimuli, each for 2 s (Tulsky *et al.*, 2014). They had to remember each stimulus, reorder the stimuli in terms of size and recite the names of the stimulus in that order. Age-corrected standard scores were analysed. See Supplementary material for details.

Pattern Comparison Processing Test

The Pattern Comparison Processing Test is a measure of processing speed in the NIH Toolbox (Carlozzi *et al.*, 2013). Participants were required to determine whether two visual patterns were identical. Patterns could vary in one of three dimensions, namely colour, adding/taking something away, or one versus many. The number of correct responses completed in 90 s was measured and age-corrected standard scores were analysed.

Rey Auditory Verbal Learning Test

In the Rey Auditory Verbal Learning Test (RAVLT), participants heard a target list of 15 nouns over five consecutive learning trials (Daniel *et al.*, 2014). After each trial, they were asked to list all the words they could remember. They were then presented with an interference list of 15 other nouns and had to freely recall the words from this list. Following that, participants were required to list words from the initial target list, and did so again after a 20-min delay. The measures of interest were (i) total recall score over the first five trials; and (ii) delayed recall score.

Matrix Reasoning Task

In this task, participants saw an array of pictures with one missing square and were required to select the missing piece out of a set of options (Daniel *et al.*, 2014). There were 32 trials and the total scaled score provided by the ABCD team was analysed.

Little Man Task

The Little Man Task (LMT) was a measure of visual spatial processing, specifically mental rotation (Acker and Acker, 1982). The task involved presentation of a male figure holding a briefcase appearing right side up, upside-down, facing the participant, or with his back towards the participant. There were 32 trials and the participant had to indicate whether the briefcase was held in the man's left or right hand. Following the recommendation of the ABCD team (Luciana *et al.*, 2018), we analysed efficiency = (% correct) / (average reaction time for accurate response) due to greater sensitivity in this measure.

Structural MRI

Participants completed a baseline MRI scan on a GE, Phillips or Siemens scanner, which included a high-resolution (1 mm isotropic voxels) T_1-weighted structural MRI image. Structural MRI data were processed by the ABCD staff using FreeSurfer v5.3.0 (http://surfer.nmr.mgh.harvard.edu/). The structural data were processed using a standardized pipeline that included removal of non-brain tissue, cortical parcellation and segmentation of grey and white matter subcortical structures. Participants' structural data were excluded if they had poor quality T_1-weighted images, FreeSurfer output failed quality control procedures, or if incidental findings were reported based on a neuroradiological report. Analyses focused on 68 cortical volumes based on the DK atlas (APARC ROI in FreeSurfer) as well as 14 subcortical regions of interest based on the ASEG atlas, namely left and right: caudate, putamen, accumbens, pallidum, thalamus, hippocampus and amygdala.

Factor analysis

An intercorrelation examination revealed that some cognitive variables are correlated with a group of other metrics, but not with measures outside that group (Supplementary Fig. 1). This suggests that more general, underlying cognitive factors might be present. Using the *psych* package in R (Revelle, 2018), the optimal number of factors was first determined via Horn's parallel analysis. Next, a factor analysis was performed by using ordinary least squares to find the minimum residual solution. The solution was assessed by scree plot, root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR). Values of RMSEA and SRMR < 0.05 are generally considered to be good (Hu and Bentler, 1999). We note that a factor analysis on the cognitive battery had been conducted previously (Thompson *et al.*, 2019), but it was necessary to perform a factor analysis in this study because different variables were examined in the RAVLT. Thompson and coworkers only considered a single overall score, which is the sum of number of words recalled across all trials and thought to be a less pure measure aggregating several cognitive processes (Vakil *et al.*, 2010). In this study, we decided to use the more standard measures of total recall over the first five learning trials and delayed recall instead.

Canonical correlation analysis

To examine whether cognitive factors might be reliably associated with a set of regional brain volumes, we conducted canonical correlation analysis (CCA) using the *candisc* package in R (Friendly and Fox, 2017). Effects of intracranial volume and gender were first regressed out from regional volumes (age was not included due to narrow range). The matrix of residual brain volumes, as well as a matrix of cognitive factor scores, were then inputted into a CCA. CCA takes these two multidimensional datasets and computes vectors in each subspace (i.e. canonical variates) that maximally correlate with each other to form a canonical pair. Multiple canonical pairs that are orthogonal to one another can be found depending on the dimensionality of the input data.

The reliability of canonical pairs were determined by three methods (Dinga *et al.*, 2019). The first was through Wilks' lambda statistic with cut-off of $P < 0.05$. Second, 10-fold cross-validation was conducted. The sample was randomly partitioned into 10 approximately equal subsets. One subset was retained as validation data for testing while the remaining nine were used as training data. During cross-validation, CCA was performed on the training data. Coefficients from the training set were then used to compute canonical variates and correlations in the test set. Ten rounds of cross-validation were performed, with each of the 10 subsets being used exactly once as validation data. The average out-of-sample correlation for each component was calculated and compared against the original value. Finally, we conducted permutation testing. The dataset was randomly shuffled in order to mismatch subject indices between the brain and cognitive matrices. CCA was then performed on this random dataset. Twenty thousand permutations were run to generate a null distribution of canonical correlations for comparison to the original value.

Bayesian model comparison

To examine group differences in cognitive factors and canonical variates, Bayesian linear mixed effect models were constructed using the *BayesFactor* package in R (Morey and Rouder, 2018) and default multivariate Cauchy priors, which have been shown to be appropriate for many real-world experimental designs common in psychological science and possess desirable properties, including location and scale invariance, consistency in Bayes Factor (BF) approaching the appropriate bound as sample size tends towards infinity, and consistency in information (Rouder *et al.*, 2012). Multiple models that differed in equality constraints between groups were computed. This allowed us to derive BF, which quantify how much more likely one model explained the observed data compared to another model.

A larger BF indicates stronger evidence in favour of one model over another. In contrast, BF close to 1 suggest data insensitivity, that is, more data are required as there is insufficient evidence for either theory. Parameters were estimated by simulating posterior distributions via Markov Chain Monte Carlo sampling.

Bayesian correlations

Bayesian correlations were performed using the *brms* package in R (Bürkner, 2017). For robustness, we assumed a multivariate *t*-distribution for the data and applied a weakly informative gamma prior on its degrees of freedom, i.e. $\nu \sim \Gamma(2, 0.1)$, with 2 and 0.1 representing the associated parameters adopted for the prior (Juárez and Steel, 2010). This can be thought of as a generalized correlation model that can flexibly adapt to the level of noise and incorporates conventional bivariate normal Pearson's correlation within it as a special case when ν is large. Estimation is conducted in Stan (Carpenter *et al.*, 2017) using the Hamiltonian Monte Carlo algorithm extension, No-U-Turn Sampler (Hoffman and Gelman, 2014).

Data availability

Data used in the preparation of this article were obtained from the ABCD Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). The ABCD data repository grows and changes over time. The ABCD data used in this report came from the Curated Annual Release 2.0 (doi: 10.15154/1503209).

Results

Participant demographics

Using information from the KSADS-5 and Family History Assessment Module Screener (FHAM-S), we identified: (i) low-risk healthy children (LRH, $n = 2100$), who fulfilled no diagnostic criteria on the KSADS-5 and had both biological parents without any history of mental illnesses; (ii) high-risk healthy children (HRH, $n = 2023$), who did not fulfil any diagnostic criteria on the KSADS-5 but had at least one biological parent with history of depression; (iii) remitted depressed children (RD, $n = 401$), who fulfilled KSADS-5 criteria for past (but not current) major, persistent or unspecified depressive disorder. Parental history of depression was not an inclusion criterion; and (iv) currently depressed children (CD, $n = 102$), who had major, persistent or unspecified depressive disorder based on KSADS-5. Parental history of depression was not an inclusion criterion (Table 1).

Three-factor latent structure revealed in the ABCD cognitive battery

A simple three-factor model provided the most parsimonious account of the data. This structure had good model fit [RMSEA = 0.037 with 90% confidence interval (CI) of 0.032–0.042, SRMR = 0.02] (Hu and Bentler, 1999) and was supported by Horn's parallel analysis (Horn, 1965), which showed that the observed eigenvalues of three factors were larger than the mean eigenvalues of 50 uncorrelated random datasets. We adopted a cut-off of >0.30 for variable loadings and labelled each factor according to what we believe best characterized the component variables (Table 2). Factor 1 comprised the Picture Vocabulary Test and Oral Reading Recognition Test, which assessed language comprehension, as well as the List Sorting Test and Matrix Reasoning Task, which tested reasoning skills such as ability to sort items based on a certain criterion and identify missing pieces in a pattern. Accordingly, it was labelled 'language and reasoning'. Factor 2 consisted of the Flanker Task, Dimension Change Card Sort Task, Pattern Comparison Processing Test and Little Man Task, which assessed ability to think flexibly and shift between different concepts. Hence, we designated it as 'cognitive flexibility'. Factor 3 comprised the Picture Sequence Test, as well as RAVLT total and delayed recall metrics that measured ability to recall a target list of words over the first five trials and after 20 min, respectively. Thus, it was labelled 'memory recall'.

Remitted and currently depressed children were more impaired in language and reasoning than healthy children

To assess group differences in these cognitive domains, we examined eight *a priori* models that each included a random effect of site but varied in equality constraints between groups: (i) LRH=HRH=RD=CD, null hypothesis; (ii) LRH≠HRH=RD=CD, non-specific vulnerability/state/trait effect; (iii) LRH=HRH≠RD=CD, non-specific state/trait, but no vulnerability, effect; (iv) LRH=HRH=RD≠CD, specific state, but no vulnerability and trait, effect; (v) LRH≠HRH≠RD=CD, specific vulnerability and non-specific state/trait effects; (vi) LRH≠HRH=RD≠CD, non-specific vulnerability/trait and specific state effects; (vii) LRH=HRH≠RD≠CD, specific state and trait, but no vulnerability, effects; and (viii) LRH≠HRH≠RD≠CD, specific vulnerability, state and trait effects.

For language and reasoning, Model 7 (LRH=HRH≠RD≠CD) was best (BF = 1.06×10^3) relative to the null hypothesis (Model 1: LRH=HRH=RD=CD). However, it only slightly outperformed the second-best model (Model 3: LRH=HRH≠RD=CD, BF = 1.08). Compared to the third-best model (Model 8: LRH≠HRH≠RD≠CD), Models 7 and 3 were 4.95 and 4.57 times more likely, respectively, to explain the observed data, which indicated these two models offered the most parsimonious account (Supplementary Tables 1 and 2). Thus, LRH and HRH did not differ in language and reasoning abilities, suggesting the absence of a pre-existing vulnerability in this domain. In contrast, RD and CD performed worse than the other two groups, with evidence reflecting either a non-specific state/trait impairment or unique state and trait effects. To facilitate

Table 1 Demographic and clinical details of participants

	LRH	HRH	RD	CD	BF_{10}
n	2100	2023	401	102	–
Age, months, mean (SD)	119.1 (7.3)	118.7 (7.4)	119.2 (7.4)	118.7 (7.7)	0.011
Sex at birth, male: female[a]	1078:1022	1010:1013	220:181	65:37	0.016
Major depressive disorder, *n* (%)[b]	0 (0.0)	0 (0.0)	192 (47.9)	81 (79.4)	–
Persistent depressive disorder, *n* (%)[b]	0 (0.0)	0 (0.0)	4 (1.0)	0 (0.0)	–
Unspecified depressive disorder, *n* (%)[b]	0 (0.0)	0 (0.0)	209 (52.1)	22 (21.6)	–
Maternal depression history, *n* (%)[c]	0 (0.0)	1498 (74.0)	131 (32.7)	35 (34.3)	–
Paternal depression history, *n* (%)[c]	0 (0.0)	921 (45.5)	64 (16.0)	19 (18.6)	–

[a]Bayesian contingency table test assuming joint multinomial sampling scheme.
[b]Diagnoses are based on the KSADS-5. Number in the RD and CD groups add up to slightly more than total because four RD and one CD participant fulfilled the diagnostic criteria for two categories.
[c]The same child might have both mother and father with history of depression, hence the sum of *n*'s in the maternal and paternal cells do not equal the number of unique individuals. There are 2023 (100%), 151 (37.6%) and 41 (40.1%) unique children with parental history of depression in the HRH, RD and CD groups, respectively.

Table 2 Loadings from factor analysis of the ABCD cognitive battery

	Factor 1	Factor 2	Factor 3
Matrix Reasoning Task	**0.46**	0.17	0.24
Picture Vocabulary Test	**0.66**	0.14	0.17
Oral Reading Recognition Test	**0.64**	0.16	0.15
List Sorting Test	**0.48**	0.24	0.29
Flanker Task	0.18	**0.58**	0.08
Dimensional Change Card Sort	0.17	**0.64**	0.16
Pattern Comparison Processing	0.06	**0.61**	0.11
Little Man Task	0.16	**0.33**	0.10
RAVLT Total Recall	0.28	0.17	**0.78**
RAVLT Delayed Recall	0.19	0.12	**0.82**
Picture Sequence Test	0.24	0.20	**0.40**

RAVLT = Rey Auditory Verbal Learning Task. Loadings >0.30 are highlighted in bold.

interpretation, we plotted estimated marginal means of the fully unconstrained model in Fig. 1A. Findings were similar even after correcting for differences in internalizing and externalizing syndromes (Supplementary material). At the request of an anonymous reviewer, we also conducted additional analyses which suggested that besides depression, impairments in language and reasoning might also be related to attention-deficit hyperactivity disorder (ADHD) (Supplementary material).

Results of Bayesian model comparison for other cognitive factors

There was no compelling evidence for vulnerability, state or trait impairment in cognitive flexibility (Fig. 1B) and memory recall (Fig. 1C). See Supplementary material for details.

Brain-cognition canonical correlation analysis

Next, we conducted a CCA to examine whether the three cognitive factors might be reliably associated with a set of regional brain volumes. All three canonical variates (CVs) were significant according to Wilks' lambda statistic (CV1: $r = 0.307$, $P < 0.001$; CV2: $r = 0.189$, $P < 0.001$; CV3: $r = 0.169$, $P < 0.01$). However, CCA is prone to overfitting. To evaluate reliability, we adopted a 10-fold cross-validation approach and found that while the average out-of-sample canonical correlation for CV1 was moderate at 0.244, those for CVs 2 and 3 were very low at 0.070 and 0.073, respectively. This suggests that only the first CV is reliable (Fig. 2). To support this, CCA on 20 000 permuted datasets revealed that the canonical correlation for CV1 was highly unlikely to occur under the generated null distribution ($P < 0.001$, Supplementary Fig. 3).

CV1 predominantly defined language and reasoning, which was correlated with a set of brain volumes showing the highest loadings in regions such as the middle temporal gyrus, pars orbitalis, superior frontal gyrus and superior parietal cortex (Fig. 2).

State and trait deficits in regional brain volumes were correlated with language and reasoning

With respect to CV1 scores along the brain dimension, we found that Model 4 (LRH=HRH=RD≠CD) had the best fit relative to null (Model 1: LRH=HRH=RD=CD, BF = 5.94), although it should be noted that it was only modestly better than the second-best (Model 7: LRH=HRH≠RD≠CD, BF = 1.95) and third-best models (Model 3: LRH=HRH≠RD=CD, BF = 1.98). No other alternative hypotheses outperformed the null model (Supplementary Tables 13 and 14). Hence, there was no difference in brain CV1 scores between LRH and HRH, indicating the absence of a pre-existing vulnerability. However, a state-related impairment is present as CD children had considerably smaller volumes in brain regions associated with language and reasoning. More data are required to determine if deficiencies in these areas might be a trait marker (Fig. 3A). Results for CV1 cognition were, unsurprisingly,

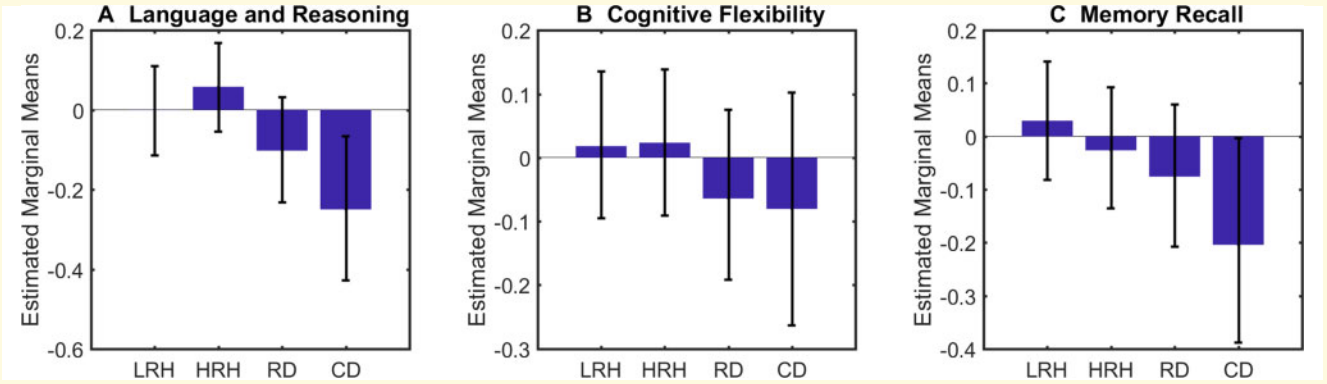

Figure 1 Group comparison of scores across the three cognitive factors. (**A**) For the language and reasoning factor, Bayesian model comparison found that Models 7 (LRH=HRH≠RD≠CD) and 3 (LRH=HRH≠RD=CD) performed the best, indicating the absence of any vulnerability impairment. However, more data will be required to determine whether these reflected a non-specific state/trait impairment, or unique state and trait effects. (**B**) In terms of cognitive flexibility, the null Model 1 (LRH=HRH=RD=CD) was comparable to Model 3 (LRH=HRH≠RD=CD) (BF = 1.01), but outperformed all other alternative models. (**C**) For memory recall, Models 4 (LRH=HRH=RD≠CD), 6 (LRH≠HRH=RD≠CD) and 3 (LRH=HRH≠RD=CD) were only anecdotally better than the null model (LRH=HRH=RD=CD). The fit of all other models were worse than the null hypothesis. Vertical lines indicate lower and upper bound of the 95% highest posterior density.

very similar to those found earlier for the language and reasoning factor (Fig. 3B and Supplementary material).

Accounting for parental history of depression in remitted and currently depressed children

Parental depression was not an inclusion criterion for RD and CD groups, but 37.6% of remitted and 40.1% of currently depressed children have at least one parent with a history of depression. To account for this, we compared three models that each included group (RD versus CD) and a random effect of site, but differed in whether they comprised parental depression and/or its interaction with group. For all cognitive factors and canonical variates, the data were more likely to occur under a model with only group, compared to a model that additionally included parental depression and the full model that also included the interaction term (Supplementary Table 18). This suggests that RD and CD children performed similarly regardless of parental depressive history.

Gender effects in cognitive factors and canonical variates

At the request of an anonymous reviewer, potential gender effects were also tested. For each cognitive factor and canonical variate, we took the best-fitting model that contained only group and a random effect of site and examined whether adding gender and/or its interaction with group would explain the data better. In terms of memory recall, there was very strong evidence for the presence of an additional effect of gender (but not interaction) whereby females performed better than males (BF > 100, Supplementary Tables 19 and 20), but no compelling evidence for gender differences in the other cognitive factors and canonical variates (Supplementary Table 19).

Children who were more withdrawn/depressed had lower cognitive flexibility

Given the heterogeneity of depression, we also conducted Bayesian correlations in the CD group to investigate how different CBCL depressive syndromes—specifically, anxious depression, withdrawn depression and somatic complaints—might be related to the cognitive factors and canonical variates. Interestingly, an examination of the posterior distribution revealed that the correlation coefficient between withdrawn depression and cognitive flexibility was between –0.43 and –0.02 (median = –0.21) with 95% probability. This suggests that depressed children who were more withdrawn were likely to have lower cognitive flexibility. There was no evidence for the presence of other reliable correlations (Supplementary Fig. 4).

Discussion

The degree to which neurocognitive impairments might represent pre-existing vulnerability, state or trait markers of depression is highly debated and remains unclear. To clarify this, we analysed the large ABCD dataset and adopted a Bayesian inference approach to compare several *a priori* hypotheses differing in their postulates of how LRH, HRH, RD and CD children performed in various cognitive domains. This method presents several distinct advantages (Wagenmakers *et al.*, 2018). Unlike frequentist inference, there is perfect inferential symmetry with no bias against the

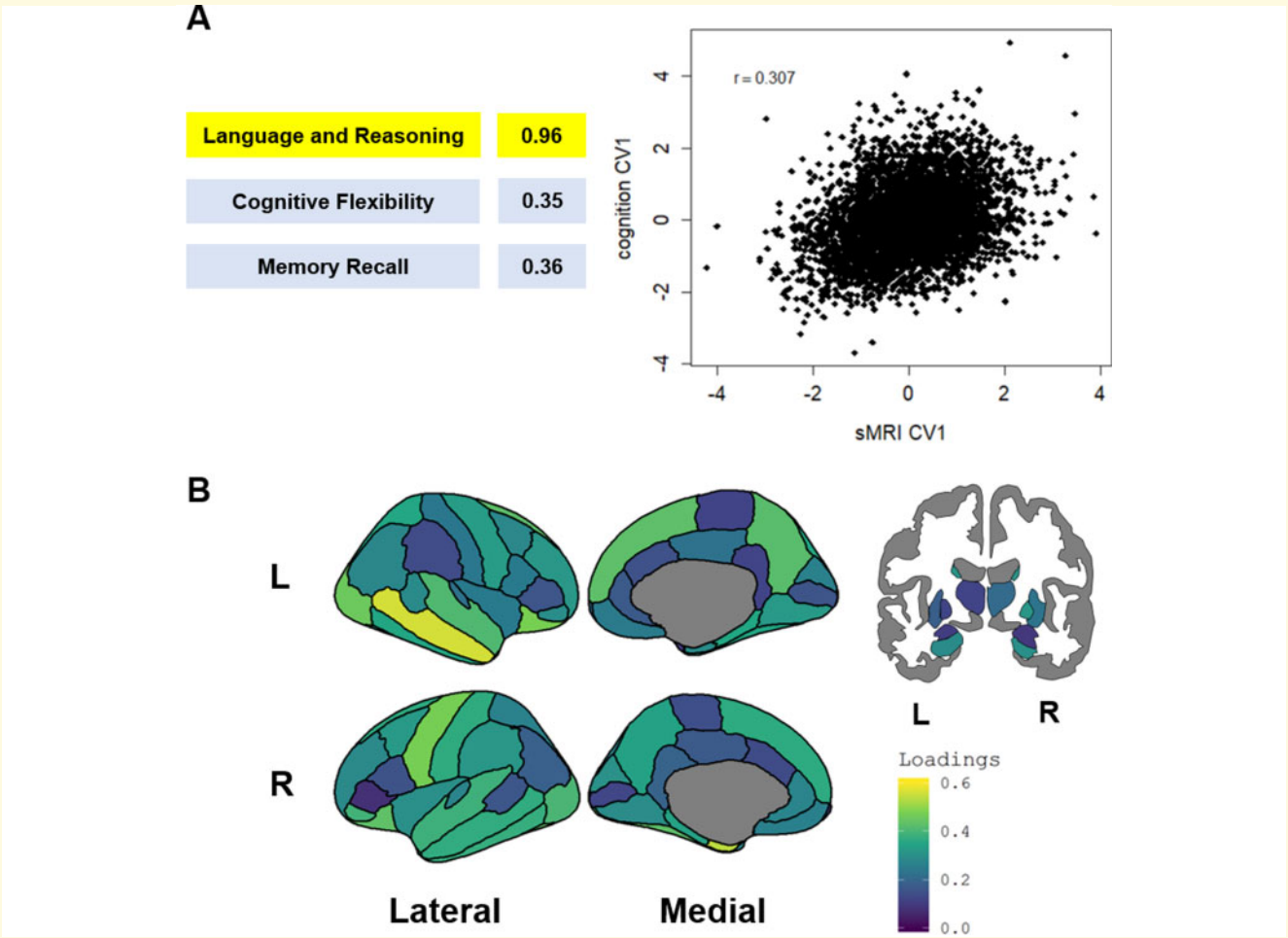

Figure 2 Results of canonical correlation analyses. (**A**) *Left*: As can be seen by the loadings of the cognitive factors, the first canonical variate along the cognition dimension is dominated by language and reasoning. *Right*: Correlation of the first brain-cognition canonical pair. (**B**) Illustration of loadings of regional brain volumes on the first canonical variate. The most strongly loading regions included those implicated in language processing, such as the middle temporal gyrus and inferior frontal cortex, as well as areas thought to be involved in reasoning, such as the superior frontal gyrus and superior parietal cortex. The Desikan-Killiany and automatic subcortical atlases in Freesurfer were used.

null hypothesis. Bayes factors are also easily interpretable and quantify the strength of evidence provided by the data for competing models without any dependence on sample size or number of comparisons made.

Our first key finding was that there were no pre-existing vulnerabilities in language and reasoning, cognitive flexibility and memory recall. Prior studies comparing unaffected first-degree relatives of depressed individuals to controls with no familial history of depression have yielded mixed results. Some have reported deficits in executive function (Winters *et al.*, 1981; Christensen *et al.*, 2006; Belleau *et al.*, 2013; Hughes *et al.*, 2013), language processing (Christensen *et al.*, 2006; Hsu *et al.*, 2014), memory (Christensen *et al.*, 2006) and cognitive flexibility (Hsu *et al.*, 2014), but others found no difference in similar domains (Klimes-Dougan *et al.*, 2006; Micco *et al.*, 2009; Hsu *et al.*, 2014; Santucci *et al.*, 2014). MacKenzie *et al.* (2019) recognized that these investigations had very small samples sizes; hence, they conducted a meta-analysis and identified the presence of pre-existing vulnerabilities across all measures of cognition in domains including language, intelligence and memory. This appears incongruent to our results, although several caveats in their study should be noted. First, they did not account for potential differences in subclinical symptoms of psychopathology. In contrast, we showed that there is no difference in cognition between children at low and high familial risk of depression even after controlling for individual differences in internalizing and externalizing problems. Second, cognitive paradigms are often impure and operate across multiple domains (Phillips, 1997; Austin *et al.*, 2001). Despite this task impurity problem, the numerous different tasks that were included in the meta-analysis were each classified into various cognitive domains. In contrast, we used a latent variable method to extract common variance across a consistent cognitive battery completed by every participant. Third, Mackenzie *et al.* (2019) concluded that general impairment

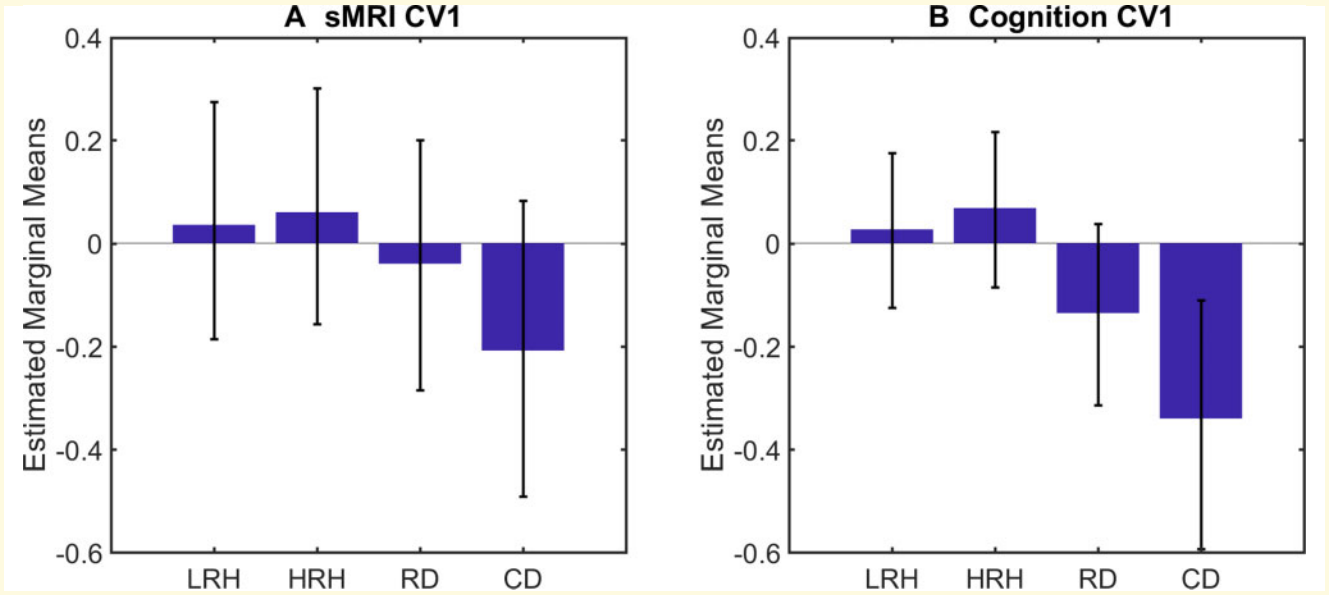

Figure 3 Group comparison of scores along the first canonical variate. (**A**) For structural MRI (sMRI) CV1, Bayesian model comparison found that Models 4 (LRH=HRH=RD≠CD), 7 (LRH=HRH≠RD≠CD) and 3 (LRH=HRH≠RD=CD) provided the most parsimonious account of the data. Thus, a state deficit, but no pre-existing vulnerability, is present. However, more data are required to determine if deficiencies in these areas might be a trait marker. (**B**) The results for cognition CV1 are similar to those for the language and reasoning factor. Models 7 (LRH=HRH≠RD≠CD) and 3 (LRH=HRH≠RD=CD) performed the best, indicating the absence of any pre-existing vulnerability, but more data will be required to determine whether these reflected a non-specific state/trait impairment, or unique state and trait effects. Models included a random effect of site. Vertical lines indicate lower and upper bound of the 95% highest posterior density.

in cognition is a vulnerability feature of depression on the basis of significant frequentist tests, even though what they have really demonstrated is the lack of sufficient evidence to reject the null hypothesis. In other words, the absence of evidence had been interpreted incorrectly as the evidence of absence. In contrast, our Bayesian model comparison analyses have demonstrated that the most parsimonious accounts of the data for language and reasoning, cognitive flexibility and memory recall all indicate the absence of a pre-existing vulnerability in these domains.

Second, we found that impairments in language and reasoning were state and trait features of depression. However, there is insufficient evidence to determine whether it represented a non-specific state/trait deficit, or unique state and trait effects. Even though the sample sizes of RD and CD children ($n = 503$) were substantially larger than previous empirical studies, more data (probably for CD) are required to disentangle them. Nevertheless, we can confidently conclude that state and trait deficits are present in this domain, which are in line with prior literature reporting worse language and executive functioning (e.g. reasoning and planning) performances in current and remitted depression (McDermott and Ebmeier, 2009; Lee *et al.*, 2012; Bora *et al.*, 2013; Snyder, 2013; Baune *et al.*, 2014; Rock *et al.*, 2014). Based on our results, one might be tempted to speculate that scar impairments in language and reasoning, which are independent of pre-existing vulnerability, are also present in depression. Indeed, we found that RD children performed worse in this domain than LRH and HRH children. Moreover, only 37.7% of RD children have parental history of depression (Table 1), suggesting that most of these individuals had low pre-existing vulnerability. These observations provide some support for the scarring hypothesis, although future studies with a longitudinal design are needed to confirm this. Interestingly, there was no compelling evidence for the existence of vulnerability, state and trait deficits in cognitive flexibility and memory recall (Supplementary material).

To the best of our knowledge, this is also the first study to find reliable regional brain volume correlates of language and reasoning, as well as dissociate vulnerability, state and trait effects in them. Peterson *et al.* (2009) reported that cortical thinning in the right hemisphere predicted lower attention and visual memory performance in individuals at high familial risk for depression compared to those at low risk. However, it is unclear to what degree this reflects a pre-existing vulnerability as ∼40% of their sample were acutely depressed or in remission. Lower performance in attention (Leung *et al.*, 2009; Li *et al.*, 2010), set-shifting (Vasic *et al.*, 2008), and cognitive control (Jung *et al.*, 2014) have also been associated with specific structural abnormalities, albeit in relatively small sample sizes ($n = 15–50$); moreover, these studies examined only depressed patients versus controls (Vasic *et al.*, 2008; Leung *et al.*, 2009; Jung *et al.*, 2014), or remitted versus non-remitted patients (Li *et al.*, 2010). Using canonical correlation via traditional Wilks' lambda, 10-fold

cross-validation and permutation testing, we found a reliable canonical variate of brain volumes that was associated predominantly with language and reasoning. The most strongly loading regions included those implicated in language processing, such as the middle temporal gyrus (Stemmer and Whitaker, 2008; Petrides, 2014; Friederici, 2015) and inferior frontal gyrus (Stemmer and Whitaker, 2008; Petrides, 2014; Friederici, 2015), as well as areas implicated in reasoning, such as the superior frontal gyrus (Prabhakaran *et al.*, 1997; Perfetti *et al.*, 2009; Schilling *et al.*, 2013) and superior parietal gyrus (Prabhakaran *et al.*, 1997; Goel and Dolan, 2001; Knauff *et al.*, 2002, 2003; Wendelken, 2014). Consistent with behavioural findings, CD children had the smallest volumes in these areas (i.e. presence of state effect) while LRH and HRH children did not differ (i.e. absence of vulnerability effect). The volumes of RD children were comparable to LRH and HRH children, although it should be noted that evidence in favour of this hypothesis was only modestly better than two alternative models postulating the presence of a trait deficit.

In addition, robust Bayesian correlations in the CD group revealed a negative relationship between cognitive flexibility and withdrawn depression, but not anxious depression and somatic complaints. Accordingly, depressed children who were more withdrawn tended to have lower cognitive flexibility. These findings suggest it might be important to consider different symptom dimensions when examining the cognitive profile of depression. Interestingly, our results appear to contradict a previous study (Lundy *et al.*, 2010), which reported that children with better cognitive flexibility based on the Trail Making Tests had higher levels of anxious depression but not withdrawn depression. This discrepancy might have arisen due to paradigm differences and future work is needed to reconcile these differences.

Limitations in this study should be noted. First, we analysed a young pre-adolescent sample (9–10 years old) and thus, findings may not be generalizable to older samples, including adults. Despite this, our results might be valuable as the early adolescent phases represent a highly vulnerable period for onset of depression, but most of the prior literature has focused on adults (Baune *et al.*, 2014; Allott *et al.*, 2016). Moreover, our data are less likely to suffer from confounds that are associated with older populations and may impact on cognitive ability, such as long-term medication use, substance abuse, prior hospitalization, and electroconvulsive therapy. Nevertheless, the development of cognitive abilities in domains including language, reasoning and memory continues throughout adolescence and early adulthood (Luna *et al.*, 2004; Rosselli *et al.*, 2014; Cromer *et al.*, 2015). Thus, it is possible that cognitive vulnerabilities might not yet have emerged in our sample. Furthermore, depression tends to peak in onset at mid-to-late adolescence, i.e. the age of 9–10 years old is thought to be generally young for experience of depression and might suggest more serious illness with poorer prognosis (Thapar *et al.*, 2012). Future studies could investigate an older sample within the adolescent age range.

Second, while the ABCD cognitive battery was comprehensive, only 'cold' emotional-independent tasks were administered. Given that mood dysregulation is a key feature of depression, future work could seek to dissociate vulnerability, state and trait features using 'hot' emotion-laden cognitive paradigms (Roiser and Sahakian, 2013). Third, we have adopted a cross-sectional design comparing healthy children at high- and low-familial risk to examine the vulnerability hypothesis. However, not all high-risk individuals will go on to develop depression and it is unlikely that everyone at low familial risk will remain free of depression. In other words, vulnerability markers derived from family studies are limited in predicting risk for developing the disorder. One way to overcome this weakness in the future is to use prospective longitudinal designs instead, which acquire repeated neurocognitive assessments before and after the onset of depression (Zammit *et al.*, 2004; Airaksinen *et al.*, 2007; Koenen *et al.*, 2009; Simons *et al.*, 2009). Fourth, although the focus of this study was on depression, it should be noted that we also found a relationship between language and reasoning with ADHD (Supplementary material). This suggests that the findings might also be generalizable to ADHD. Fifth, despite the very large sample size in the ABCD study, it was still insufficient to distinguish between some models, such as whether or not deficits in language and reasoning and their associated brain volumes represented unique state and trait features of depression. One possible explanation might be that tasks used in the ABCD study were not sensitive enough, and future studies could consider investigating different tasks (including 'hot' cognition tasks) instead.

To conclude, this is the first empirical study to disentangle pre-existing vulnerability, state and trait features of neurocognitive impairments in depression. By adopting a Bayesian inferential approach in the ABCD study, we showed that state and trait impairments in language and reasoning—as well as state (and possibly trait) abnormalities in brain structures involved in this cognitive domain—were characteristic of depression in early adolescence. However, there is no compelling evidence for the existence of vulnerability, state and trait deficits in other domains of cognition, at least as assessed by the tasks administered in the ABCD study. These findings have important clinical implications, suggesting that cognitive dysfunction may not be useful targets of preventive interventions—although it should be noted that they are cross-sectional in nature and specific to a narrow group of 9–10-year-old children. Depressed patients, even after remission, might also derive benefit from less commonly used treatment strategies such as cognitive remediation therapy (Kim *et al.*, 2018).

Acknowledgements

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10 000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD

Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA 041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA 041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA 041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/scientists/workgroups/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in 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.

Funding

Y.S.A. was supported by the Kaplen Fellowship in Depression from Harvard Medical School as well as the A*STAR National Science Scholarship. E.B. is supported by a Klingenstein Third Generation Foundation Postdoctoral Fellowship and an Adam J. Corneel Young Investigator Award from McLean Hospital. D.A.P. was partially supported by National Institute of Mental Health grant R37 MH068376 and R01 MH101521 as well as the Tommy Fuss Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Competing interests

D.A.P. reports the following financial disclosures for activities unrelated to the current research: funding from NIMH, Brain and Behavior Research Foundation, the Dana Foundation, and Millennium Pharmaceuticals; consulting fees from Akili Interactive Labs, BlackThorn Therapeutics, Boehreinger Ingelheim, Posit Science, Takeda Pharmaceuticals, Compass Pathway and Otsuka Pharmaceuticals; one honorarium from Alkermes; stock options from BlackThorn Therapeutics.

Supplementary material

Supplementary material is available at *Brain* online.

References

Achenbach TM, Rescorla LA, Manual of the ASEBA school-aged forms and profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families; 2001.

Acker W, Acker C. Bexley Maudsley automated processing screening and Bexley Maudsley category sorting test manual. Windsor, England: NFER-Nelson; 1982.

Airaksinen E, Wahlin Å, Forsell Y, Larsson M. Low episodic memory performance as a premorbid marker of depression: evidence from a 3-year follow-up. Acta Psychiatr Scand 2007; 115: 458–65.

Allott K, Fisher CA, Amminger GP, Goodall J, Hetrick S. Characterizing neurocognitive impairment in young people with major depression: state, trait, or scar? Brain Behav 2016; 6: e00527.

Amico F. Structural MRI correlates for vulnerability and resilience to major depressive disorder. J Psychiatry Neurosci 2011; 36: 15–22.

Austin MP, Mitchell P, Goodwin GM. Cognitive deficits in depression: possible implications for functional neuropathology. Br J Psychiatry 2001; 178: 200–6.

Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: rationale and description. Dev Cogn Neurosci 2018; 32: 55–66.

Baune BT, Fuhr M, Air T, Hering C. Neuropsychological functioning in adolescents and young adults with major depressive disorder–a review. Psychiatry Res 2014; 218: 261–71.

Beaujean AA, Parker S, Qiu X. The relationship between cognitive ability and depression: a longitudinal data analysis. Soc Psychiatry Psychiatr Epidemiol 2013; 48: 1983–92.

Belleau EL, Phillips ML, Birmaher B, Axelson DA, Ladouceur CD. Aberrant executive attention in unaffected youth at familial risk for mood disorders. J Affect Disord 2013; 147: 397–400.

Birmaher B, Williamson DE, Dahl RE, Axelson DA, Kaufman J, Dorn LD, et al. Clinical presentation and course of depression in youth: does onset in childhood differ from onset in adolescence?. J Am Acad Child Adolesc Psychiatry 2004; 43: 63–70.

Bloch Y, Aviram S, Braw Y, Gvirts HZ, Rabany L, Walter G. Attention improves after clinical improvement in acutely depressed adolescents. J Nat Prod 2015; 27: 153–6.

Bora E, Harrison BJ, Davey CG, Yücel M, Pantelis C. Meta-analysis of volumetric abnormalities in cortico-striatal-pallidal-thalamic circuits in major depressive disorder. Psychol Med 2012; 42: 671–81.

Bora E, Harrison BJ, Yücel M, Pantelis C. Cognitive impairment in euthymic major depressive disorder: a meta-analysis. Psychol Med 2013; 43: 2017–26.

Bürkner P-C. **brms**: An *R* Package for Bayesian Multilevel Models Using *Stan* [Internet]. J Stat Softw 2017; 80. Available from: http://www.jstatsoft.org/v80/i01/ (11 July 2020, date last accessed).

Carlozzi NE, Tulsky DS, Kail RV, Beaumont JL. VI. NIH toolbox cognition battery (CB): measuring processing speed: NIH toolbox cognition battery (CB). Monogr Soc Res Child 2013; 78: 88–102.

Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. *Stan*: a probabilistic programming language. J Stat Softw 2017; 76: 1–32. Available from: http://www.jstatsoft.org/v76/i01/.

Christensen MV, Kyvik KO, Kessing LV. Cognitive function in unaffected twins discordant for affective disorder. Psychol Med 2006; 36: 1119–29.

Christensen MV, Kyvik KO, Kessing LV. Subclinical psychopathology and socio-economic status in unaffected twins discordant for affective disorder. J Psychiatr Res 2007; 41: 229–38.

Clark L, Kempton MJ, Scarnà A, Grasby PM, Goodwin GM. Sustained attention-deficit confirmed in euthymic bipolar disorder but not in first-degree relatives of bipolar patients or euthymic unipolar depression. Biol Psychiatry 2005; 57: 183–7.

Cromer JA, Schembri AJ, Harel BT, Maruff P. The nature and rate of cognitive maturation from late childhood to adulthood. Front Psychol 2015; 6: 704.

Daniel MH, Wahlstrom D, Zhang O, Equivalence of Q-Interactive™ and Paper Administrations of Cognitive Tasks: WISC–V. Q-Interactive Technical Report 8. 2014.

Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, et al. Measuring episodic memory across the lifespan: NIH

toolbox picture sequence memory test. J Int Neuropsychol Soc 2014; 20: 611–9.
Dinga R, Schmaal L, Penninx BWJH, van Tol MJ, Veltman DJ, van Velzen L, et al. Evaluating the evidence for biotypes of depression: methodological replication and extension of Drysdale et al. (2017). Neuroimage Clin 2019; 22: 101796.
Favre T, Hughes C, Emslie G, Stavinoha P, Kennard B, Carmody T. Executive functioning in children and adolescents with major depressive disorder. Child Neuropsychol J Norm Abnorm Dev Child Adolesc 2009; 15: 85–98.
Friederici AD. White-matter pathways for speech and language processing. Handb Clin Neurol 2015; 129: 177–86.
Friendly M, Fox J. Candisc: Visualizing Generalized Canonical Discriminant and Canonical Correlation Analysis. R package version 0.8-0 [Internet]. 2017. https://CRAN.R-project.org/package=candisc
Garavan H, Bartsch H, Conway K, Decastro A, Goldstein RZ, Heeringa S, et al. Recruiting the ABCD sample: design considerations and procedures. Dev Cogn Neurosci 2018; 32: 16–22.
Gershon RC, Cook KF, Mungas D, Manly JJ, Slotkin J, Beaumont JL, et al. Language measures of the NIH toolbox cognition battery. J Int Neuropsychol Soc 2014; 20: 642–51.
Goel V, Dolan RJ. Functional neuroanatomy of three-term relational reasoning. Neuropsychologia 2001; 39: 901–9.
Goodall J, Fisher C, Hetrick S, Phillips L, Parrish EM, Allott K. Neurocognitive functioning in depressed young people: a systematic review and meta-analysis. Neuropsychol Rev 2018; 28: 216–31.
Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. J Abnorm Psychol 1998; 107: 128–40.
Hasselbalch BJ, Knorr U, Kessing LV. Cognitive impairment in the remitted state of unipolar depressive disorder: a systematic review. J Affect Disord 2011; 134: 20–31.
Hoffman MD, Gelman A. The No-U-Turn Sampler: adaptively Setting Path Lengths in Hamiltonian Monte Carlo. J Mach Learn Res 2014; 15: 1593–623.
Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrika 1965; 30: 179–85.
Hsu KJ, Young-Wolff KC, Kendler KS, Halberstadt LJ, Prescott CA. Neuropsychological deficits in major depression reflect genetic/familial risk more than clinical history: a monozygotic discordant twin-pair study. Psychiatry Res 2014; 215: 87–94.
Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model Multidiscip J 1999; 6: 1–55.
Hughes C, Roman G, Hart MJ, Ensor R. Does maternal depression predict young children's executive function? A 4-year longitudinal study. J Child Psychol Psychiatry 2013; 54: 169–77.
Juárez MA, Steel MFJ. Model-based clustering of non-Gaussian panel data based on skew- *t* distributions. J Bus Econ Stat 2010; 28: 52–66.
Jung J, Kang J, Won E, Nam K, Lee M-S, Tae WS, et al. Impact of lingual gyrus volume on antidepressant response and neurocognitive functions in major depressive disorder: a voxel-based morphometry study. J Affect Disord 2014; 169: 179–87.
Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 1997; 36: 980–8.
Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005; 62: 593–602.
Kim EJ, Bahk Y-C, Oh H, Lee W-H, Lee J-S, Choi K-H. Current status of cognitive remediation for psychiatric disorders: a review. Front Psychiatry 2018; 9: 461.
Klimes-Dougan B, Ronsaville D, Wiggs EA, Martinez PE. Neuropsychological functioning in adolescent children of mothers with a history of bipolar or major depressive disorders. Biol Psychiatry 2006; 60: 957–65.
Klimkeit EI, Tonge B, Bradshaw JL, Melvin GA, Gould K. Neuropsychological deficits in adolescent unipolar depression. Arch Clin Neuropsychol 2011; 26: 662–76.
Knauff M, Fangmeier T, Ruff CC, Johnson-Laird PN. Reasoning, models, and images: behavioral measures and cortical activity. J Cogn Neurosci 2003; 15: 559–73.
Knauff M, Mulack T, Kassubek J, Salih HR, Greenlee MW. Spatial imagery in deductive reasoning: a functional MRI study. Brain Res Cogn Brain Res 2002; 13: 203–12.
Koenen KC, Moffitt TE, Roberts AL, Martin LT, Kubzansky L, Harrington H, et al. Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. Am J Psychiatry 2009; 166: 50–7.
Koolschijn PCMP, van Haren NEM, Lensvelt-Mulders GJLM, Hulshoff Pol HE, Kahn RS. Brain volume abnormalities in major depressive disorder: a meta-analysis of magnetic resonance imaging studies. Hum Brain Mapp 2009; 30: 3719–35.
Korhonen V, Laukkanen E, Antikainen R, Peiponen S, Lehtonen J, Viinamäki H. Effect of major depression on cognitive performance among treatment-seeking adolescents. Nord J Psychiatry 2002; 56: 187–93.
Lee RSC, Hermens DF, Porter MA, Redoblado-Hodge MA. A meta-analysis of cognitive deficits in first-episode major depressive disorder. J Affect Disord 2012; 140: 113–24.
Leung K-K, Lee TMC, Wong MMC, Li LSW, Yip PSF, Khong P-L. Neural correlates of attention biases of people with major depressive disorder: a voxel-based morphometric study. Psychol Med 2009; 39: 1097–106.
Lewinsohn PM, Rohde P, Klein DN, Seeley JR. Natural course of adolescent major depressive disorder: I. Continuity into young adulthood. J Am Acad Child Adolesc Psychiatry 1999; 38: 56–63.
Li C-T, Lin C-P, Chou K-H, Chen I-Y, Hsieh J-C, Wu C-L, et al. Structural and cognitive deficits in remitting and non-remitting recurrent depression: a voxel-based morphometric study. Neuroimage 2010; 50: 347–56.
Liu SK, Chiu C-H, Chang C-J, Hwang T-J, Hwu H-G, Chen WJ. Deficits in sustained attention in schizophrenia and affective disorders: stable versus state-dependent markers. Am J Psychiatry 2002; 159: 975–82.
Lorenzetti V, Allen NB, Fornito A, Yücel M. Structural brain abnormalities in major depressive disorder: a selective review of recent MRI studies. J Affect Disord 2009; 117: 1–17.
Luciana M, Bjork JM, Nagel BJ, Barch DM, Gonzalez R, Nixon SJ, et al. Adolescent neurocognitive development and impacts of substance use: overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. Dev Cogn Neurosci 2018; 32: 67–79.
Luna B, Garver KE, Urban TA, Lazar NA, Sweeney JA. Maturation of cognitive processes from late childhood to adulthood. Child Dev 2004; 75: 1357–72.
Lundy SM, Silva GE, Kaemingk KL, Goodwin JL, Quan SF. Cognitive Functioning and Academic Performance in Elementary School Children with Anxious/Depressed and Withdrawn Symptoms. Open Pediatr Med J 2010; 4: 1–9.
Maalouf FT, Brent D, Clark L, Tavitian L, McHugh RM, Sahakian BJ, et al. Neurocognitive impairment in adolescent major depressive disorder: state vs. trait illness markers. J Affect Disord 2011; 133: 625–32.
Maalouf FT, Klein C, Clark L, Sahakian BJ, Labarbara EJ, Versace A, et al. Impaired sustained attention and executive dysfunction: bipolar disorder versus depression-specific markers of affective disorders. Neuropsychologia 2010; 48: 1862–8.
MacKenzie LE, Uher R, Pavlova B. Cognitive performance in first-degree relatives of individuals with vs without major depressive disorder: a meta-analysis. JAMA Psychiatry 2019; 76: 297.

McDermott LM, Ebmeier KP. A meta-analysis of depression severity and cognitive function. J Affect Disord 2009; 119: 1–8.
McIntyre RS, Cha DS, Soczynska JK, Woldeyohannes HO, Gallaugher LA, Kudlow P, et al. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. Depress Anxiety 2013; 30: 515–27.
Micco JA, Henin A, Biederman J, Rosenbaum JF, Petty C, Rindlaub LA, et al. Executive functioning in offspring at risk for depression and anxiety. Depress Anxiety 2009; 26: 780–90.
Morey RD, Rouder JN. BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. [Internet]. 2018 Available from: https://CRAN.R-project.org/package=BayesFactor
Perfetti B, Saggino A, Ferretti A, Caulo M, Romani GL, Onofrj M. Differential patterns of cortical activation as a function of fluid reasoning complexity. Hum Brain Mapp 2009; 30: 497–510.
Peters AT, Jacobs RH, Crane NA, Ryan KA, Weisenbach SL, Ajilore O, et al. Domain-specific impairment in cognitive control among remitted youth with a history of major depression. Early Interv Psychiatry 2017; 11: 383–92.
Peterson BS, Warner V, Bansal R, Zhu H, Hao X, Liu J, et al. Cortical thinning in persons at increased familial risk for major depression. Proc Natl Acad Sci USA 2009; 106: 6273–8.
Petrides M, Neuroanatomy of language regions of the human brain. 1st edn. Amsterdam: Elsevier/AP, Academic Press is an imprint of Elsevier; 2014.
Phillips L. Do 'frontal tests' measure executive function? Issues of assessment and evidence from fluency tests. In: Rabbit P, editor. Methodology of frontal and executive function. Hove, UK: Psychology Press; 1997. p. 191–213.
Prabhakaran V, Smith JA, Desmond JE, Glover GH, Gabrieli JD. Neural substrates of fluid reasoning: an fMRI study of neocortical activation during performance of the Raven's. Progressive Matrices Test. Cognit Psychol 1997; 33: 43–63.
Rasic D, Hajek T, Alda M, Uher R. Risk of mental illness in offspring of parents with schizophrenia, bipolar disorder, and major depressive disorder: a meta-analysis of family high-risk studies. Schizophr Bull 2014; 40: 28–38.
Revelle W. psych: Procedures for Personality and Psychological Research. 2018. https://CRAN.R-project.org/package=psych Version = 1.8.12.
Rice F, Harold G, Thapar A. The genetic aetiology of childhood depression: a review. J Child Psychol Psychiatry 2002; 43: 65–79.
Rice JP, Reich T, Bucholz KK, Neuman RJ, Fishman R, Rochberg N, et al. Comparison of direct interview and family history diagnoses of alcohol dependence. Alcoholism Clin Exp Res 1995; 19: 1018–23.
Robertson HA, Kutcher SP, Lagace DC. No evidence of attentional deficits in stabilized bipolar youth relative to unipolar and control comparators. Bipolar Disord 2003; 5: 330–9.
Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med 2014; 44: 2029–40.
Roiser JP, Sahakian BJ. Hot and cold cognition in depression. CNS Spectr 2013; 18: 139–49.
Rosselli M, Ardila A, Matute E, Vélez-Uribe I. Language development across the life span: a neuropsychological/neuroimaging perspective. Neurosci J 2014; 2014: 1–21.
Rouder JN, Morey RD, Speckman PL, Province JM. Default Bayes factors for ANOVA designs. J Math Psychol 2012; 56: 356–74.
Santucci AK, Singer LT, Wisniewski SR, Luther JF, Eng HF, Dills JL, et al. Impact of prenatal exposure to serotonin reuptake inhibitors or maternal major depressive disorder on infant developmental outcomes. J Clin Psychiatry 2014; 75: 1088–95.
Schilling C, Kühn S, Paus T, Romanowski A, Banaschewski T, Barbot A, et al. Cortical thickness of superior frontal cortex predicts impulsiveness and perceptual reasoning in adolescence. Mol Psychiatry 2013; 18: 624–30.
Schmid M, Hammar A. A follow-up study of first episode major depressive disorder. Impairment in inhibition and semantic fluency-potential predictors for relapse? Front Psychol 2013; 4: 633.
Semkovska M, Quinlivan L, O'Grady T, Johnson R, Collins A, O'Connor J, et al. Cognitive function following a major depressive episode: a systematic review and meta-analysis. Lancet Psychiatry 2019; 6: 851–61.
Simons CJP, Jacobs N, Derom C, Thiery E, Jolles J, van Os J, et al. Cognition as predictor of current and follow-up depressive symptoms in the general population. Acta Psychiatr Scand 2009; 120: 45–52.
Smith DJ, Muir WJ, Blackwood DH. Neurocognitive impairment in euthymic young adults with bipolar spectrum disorder and recurrent major depressive disorder. Bipolar Disord 2006; 8: 40–6.
Smith K. Mental health: a world of depression. Nature 2014; 515: 181.
Snyder HR. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychol Bull 2013; 139: 81–132.
Sullivan PF, Neale MC, Kendler KS. Genetic epidemiology of major depression: review and meta-analysis. Am J Psychiatry 2000; 157: 1552–62.
Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. Lancet 2012; 379: 1056–67.
Thompson WK, Barch DM, Bjork JM, Gonzalez R, Nagel BJ, Nixon SJ, et al. The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: findings from the ABCD study's baseline neurocognitive battery. Dev Cogn Neurosci 2019; 36: 100606.
Tulsky DS, Carlozzi N, Chiaravalloti ND, Beaumont JL, Kisala PA, Mungas D, et al. NIH Toolbox Cognition Battery (NIHTB-CB): list sorting test to measure working memory. J Int Neuropsychol Soc 2014; 20: 599–610.
Vakil E, Greenstein Y, Blachstein H. Normative data for composite scores for children and adults derived from the rey auditory verbal learning test. Clin Neuropsychol 2010; 24: 662–77.
Vasic N, Walter H, Höse A, Wolf RC. Gray matter reduction associated with psychopathology and cognitive dysfunction in unipolar depression: a voxel-based morphometry study. J Affect Disord 2008; 109: 107–16.
Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, et al. The conception of the ABCD study: from substance use to a broad NIH collaboration. Dev Cogn Neurosci 2018; 32: 4–7.
Wagenmakers E-J, Marsman M, Jamil T, Ly A, Verhagen J, Love J, et al. Bayesian inference for psychology. Part I: theoretical advantages and practical ramifications. Psychon Bull Rev 2018; 25: 35–57.
Wendelken C. Meta-analysis: how does posterior parietal cortex contribute to reasoning? Front Hum Neurosci 2014; 8: 1042.
Winters KC, Stone AA, Weintraub S, Neale JM. Cognitive and attentional deficits in children vulnerable to psychopathology. J Abnorm Child Psychol 1981; 9: 435–53.
Zammit S, Allebeck P, David AS, Dalman C, Hemmingsson T, Lundberg I, et al. A longitudinal study of premorbid IQ score and risk of developing schizophrenia, bipolar disorder, severe depression, and other nonaffective psychoses. Arch Gen Psychiatry 2004; 61: 354.
Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Weintraub S. II. NIH toolbox cognition battery (CB): Measuring executive function and attention. Monogr Soc Res Child 2013; 78: 16–33.
Stemmer B, Whitaker HA. Handbook of the neuroscience of language. 1st edn. Amsterdam: Elsevier Academic Press; 2008.

Supplemental Data for Ang et al.

Table of Contents

Supplemental Methods

Kiddie Schedule for Affective Disorders and Schizophrenia for DSM-5 (KSADS-5)

Participants were evaluated using the KSADS-5, which is a well-established, reliable and valid assessment of lifetime mental disorders in youths (Kaufman *et al.*, 1997; Barch *et al.*, 2018). Children provided self-reports in selected domains including mood and anxiety disorders, sleep and suicidality. Although parents/guardians also completed the KSADS-5 about their children, the ABCD team suggested that children's self-report may provide more accurate assessments due to three main reasons (Barch *et al.*, 2018). First, several studies have shown that parent and youth report start to diverge in early adolescence. Second, some evidence suggests that parents' report may be biased by their own mental health experiences. That is, parents' experience of depression and anxiety can color judgement of their child's level of depression and anxiety. Third, there is evidence for greater predictive utility in youth report over and above parent report. Therefore, we have chosen to use children self-report when determining children's current and past mental health.

Flanker Task (FT)

The FT is a measure of executive attention and inhibitory control in the NIH Toolbox (Zelazo *et al.*, 2013). Participants had to indicate the orientation (left or right) of a stimulus presented in the center of a screen. The central stimulus was flanked by two stimuli on each side, which were either in a congruent or incongruent orientation with the center stimulus. Participants completed a four-trial practice block, followed by an easier 25-trial test block that used fish as stimuli and a harder 25-trial test block that

used arrows at the stimuli. Each test block consisted of 16 congruent and 9 incongruent trials. Participants could only advance to the test block with arrow stimuli if they correctly answered at least 5 of 9 incongruent fish stimuli trials. Scoring procedures utilized a two-vector method, which incorporated accuracy as well as reaction time (RT) for participants who maintained a greater than 80% level of accuracy. Scores were based on the first 20 trials of each block due to fatigue effects found in the last five trials. Accuracy scores ranged from 0 to 5 and were derived from the equation: Accuracy = (5/40) × (Number of correct responses). With regards to RT, trials lower than 100ms or greater than 3SDs from the mean RT were discarded. Median RTs were computed based on correct incongruent trials and were log-transformed. A minimum 500ms and maximum 3000ms RT scoring range was established and RTs falling outside this range were reset to the minimum or maximum as appropriate. RT scores ranged from 0 to 5 and were added to the accuracy scores for participants that met the 80% accuracy criteria. Age-corrected standard scores were analyzed.

Dimension Change Card Sort Task (DCCST)

The DCCST is a measure of cognitive flexibility in the NIH Toolbox (Zelazo *et al.*, 2013). Participants were required to sort the target stimuli (i.e., white boat and green rabbit) by shape or color to match the target stimuli (i.e., a green boat or white rabbit). On each trial, participants were presented with the test stimulus for 100ms along with the instruction of whether to sort by “shape” or “color”. Following that, the target stimuli appeared along with the test stimulus for up to 10,000ms and participants had to select the target stimulus matching the test stimulus. There were 4 practice trials each of sorting by color and shape that contained performance feedback. The test blocks

consisted of five trials sorting on one dimension (pre-switch block), followed by five trials sorting on the other dimension (post-switch block). Participants who correctly responded to four out of the five trials on the post-switch block completed a 50-trial "mixed" block, with 40 "frequent" trials and 10 "infrequent" trials presented in a pseudorandom order. The frequent trials corresponded to the sorting dimension used in the post-switch block. Scores were derived using the same two-vector method described above for FT. Only the first 30 mixed-trial blocks were incorporated into the scoring algorithm due to fatigue effects. Age-corrected standard scores were analyzed.

List Sorting Test (LST)

The LST is part of the NIH Toolbox and participants were presented with a series of stimuli, each for 2 seconds (Tulsky *et al.*, 2014). They had to remember each stimulus, re-order the stimuli in terms of size and recite the names of the stimulus in that order. The task consisted of a "1-list" and "2-list" block, with "1-list" requiring the sequencing of items from a single category (animals or foods) and "2-list" demanding the ordering of items from two categories (animals and foods) simultaneously. Participants began with a two-item string, which increased by one item with each item. In the event of an incorrect response, they were presented with a second trial containing the same number of items. The task was discontinued when participants provided incorrect responses on two trials with the same number of items or correctly sequenced seven items. Scores were derived by summing the total number of correct responses across both lists. Age-corrected standard scores were analyzed.

Supplemental Results

Group differences in internalizing and externalizing syndromes

In terms of CBCL internalizing, Model 6 (LRH≠HRH=RD≠CD) is the most preferred model with BF=2.39×10^{77} relative to null (Model 1: LRH=HRH=RD=CD). It is also 6.36 times more likely than the second-best model (Model 5: LRH≠HRH≠RD=CD) to have produced the observed data (**Table S3** and **S4**). Hence, LRH had the least internalizing problems while CD had the most symptoms. The severity for HRH and RD were similar and ranked in-between the other two groups (**Fig. S2A)**.

Similarly, for CBCL externalizing, Model 6 (LRH≠HRH=RD≠CD) was the best model with a BF=1.53×10^{70} compared to null (Model 1: LRH=HRH=RD=CD). However, it only marginally outperformed (BF=1.72) the second-best model (Model 8: LRH≠HRH≠RD≠CD). Models 6 and 8 were 5.31 and 3.08 times better, respectively, than the third-best model (Model 5: LRH≠HRH≠RD=CD), suggesting that they provided the most parsimonious account of the data (**Table S5** and **S6**). Hence, the severity of externalizing problems was lowest for LRH, highest for CD, and intermediate for HRH and RD (**Fig. S2B**).

Results of Bayesian model comparison in language and reasoning after accounting for CBCL internalizing and externalizing syndromes

One might argue that that our analysis in the main text did not conduct accurate vulnerability or trait comparison since potential subclinical psychiatric symptoms were not considered. Indeed, even though HRH and RD were not clinically depressed, they reported greater internalizing and externalizing problems than LRH (**Fig. S2**). To

address this interpretative conundrum, we investigated four new models that only included observations in LRH, HRH and RD (CD was excluded as it would be unusual to remove effects of psychiatric symptoms from a psychiatric group). Each model also included a random effect of site and covariates of CBCL internalizing and externalizing subscales:

9. **LRH=HRH=RD.** Fully constrained model and null hypothesis.
10. **LRH≠HRH=RD.** Non-specific vulnerability/trait effect.
11. **LRH=HRH≠RD.** Specific trait, but no vulnerability, effect.
12. **LRH≠HRH≠RD.** Specific vulnerability and trait effect.

In line with results in the main text, Model 11 (LRH=HRH≠RD) was most preferred (BF=10.8) relative to null (Model 9: LRH=HRH=RD), and performed 3.61 times better than the second-best model (Model 12: LRH≠HRH≠RD) (**Table S7**). Hence, even after correcting for differences in internalizing and externalizing syndromes, evidence still points to the presence of a trait, but not vulnerability, impairment in language and reasoning (**Table S8**).

Results of Bayesian correlational analyses between language and reasoning and CBCL DSM diagnoses

Might impairments in language and reasoning also be associated with other childhood psychiatric disorders besides depression? We performed Bayesian correlations to examine how various CBCL DSM-oriented scales – specifically depressive problems, anxiety problems, somatic problems, attention-deficit hyperactive disorder (ADHD), oppositional defiant problems and conduct problems – were related to

language and reasoning ability. Only the RD and CD groups were included in these analyses. LRH and HRH were excluded because these children fulfilled no diagnostic criteria on the KSADS-5. Thus, adding such a large number of them (*N*=4123) along with RD and CD (*N*=503) would cause an extreme skew of the CBCL DSM scores towards the low end and result in correlations that are not meaningful. Interrogations of posterior distributions revealed that the correlation between depression and language and reasoning was between -0.20 and -0.01 (median=-0.11) with 95% probability, which was in line with our main findings and suggested that children who were more depressed were worse in this cognitive domain. Interestingly, there was also a negative association between ADHD and language reasoning (between -0.20 and -0.01 [median=-0.10] with 95% probability). See **Supplemental Fig. S5** for details. Using robust linear regression with *t*-distribution, we compared one model that predicted language and reasoning with only ADHD to another that uses both ADHD and depression. The difference in expected log predictive density (ELPD_diff) indicated that the model including depression did not differ in predictive performance to the one with only ADHD (ELPD_diff=-0.6, SE=2.0), which suggested that the findings for language and reasoning might also be generalizable to ADHD.

Results of Bayesian model comparison for other cognitive factors

In terms of cognitive flexibility, the null model (Model 1: LRH=HRH=RD=CD) was comparable to Model 3 (LRH=HRH≠RD=CD) with BF=1.01, but outperformed all other alternative models (Models 2, 4–8) by a BF ranging from 4.22–62.5 (**Table S9**). This suggests that more data will be required to examine whether a non-specific state/trait

impairment was present. Interestingly, after excluding CD and considering CBCL internalizing and externalizing subscale scores, the null model (Model 9: LRH=HRH=RD) was most strongly favored and was 3.46 times more likely to explain the observed data than the best-performing alternative model (Model 11: LRH=HRH≠RD; **Table S10**). Therefore, current evidence supports the lack of any vulnerability, state or trait impairment in cognitive flexibility (**Fig. 2B**).

When examining memory recall, Models 4 (LRH=HRH=RD≠CD), 6 (LRH≠HRH=RD≠CD) and 3 (LRH=HRH≠RD=CD) performed better relative to null (Model 1: LRH=HRH=RD=CD), albeit only anecdotally by 1.45–2.18 times. The fit of all other models were 1.15–4.57 times worse than the null hypothesis (**Table S11** and **S12**). However, after accounting for CBCL internalizing and externalizing by selecting only children in LRH, HRH and RD, the null model (Model 9: LRH=HRH=RD) was best fitting and outperformed the next best alternative model (Model 11: LRH=HRH≠RD) by BF=10.8 (**Table S13**). Together, these results suggest there is no compelling evidence for vulnerability, state or trait impairment in memory recall (**Fig. 2C**).

Results for first cognition canonical variate scores were highly similar to that for the language and reasoning factor

In terms of the first canonical variate (CV1) cognition scores, the results were unsurprisingly highly similar to that found earlier for the language and reasoning factor. Model 7 (LRH=HRH≠RD≠CD) was most preferred (BF=5.42×10^3) relative to null (Model 1: LRH=HRH=RD=CD), but only slightly outperformed the second-best model (Model 3: LRH=HRH≠RD=CD) with BF=1.4. However, when compared to the third-best model (Model 8: LRH≠HRH≠RD≠CD), Models 7 and 3 were 7.45 and 5.41 times better

respectively at explaining the observed data (**Tables S16** and **S17**). Therefore, LRH and HRH showed no difference in CV1 cognition scores, which indicated the absence of a pre-existing vulnerability. On the other hand, RD and CD performed worse than the other two groups and reflected either a non-specific state/trait impairment, or unique state and trait effects (**Fig. 3B**).

Supplemental Discussion

Interestingly, there was no compelling evidence for state or trait impairments in cognitive flexibility and memory recall. We found that a model theorizing the presence of a non-specific state/trait deficit in cognitive flexibility was comparable in likelihood to the null hypothesis; whereas for memory recall, three alternative models that all postulated a state impairment performed better than the null model, albeit only marginally. After excluding CD children and accounting for internalizing and externalizing syndromes, the null model for both cognitive flexibility and memory recall provided the most parsimonious account of the data. Prior investigations of these domains in small samples of acute and/or remitted depression have yielded conflicting findings, with some reporting impairments (Smith *et al.*, 2006; McDermott and Ebmeier, 2009; Lee *et al.*, 2012; Snyder, 2013; Baune *et al.*, 2014; Rock *et al.*, 2014) and others finding no difference (Robertson *et al.*, 2003; Favre *et al.*, 2009; Klimkeit *et al.*, 2011; Peters *et al.*, 2017). Differences in tasks aside, our findings in a relatively larger group (*N*=503) indicate support for the latter.

Supplemental Tables

Table S1: Bayesian model comparison for Language and Reasoning factor.

Model no.	Equality constraint	Bayes Factor
1	LRH = HRH = RD = CD	1
2	LRH ≠ HRH = RD = CD	1.33×10^{0} ± 0.40%
3	LRH = HRH ≠ RD = CD	9.78×10^{2} ± 0.49%
4	LRH = HRH = RD ≠ CD	6.79×10^{1} ± 0.58%
5	LRH ≠ HRH ≠ RD = CD	9.32×10^{1} ± 0.35%
6	LRH ≠ HRH = RD ≠ CD	8.32×10^{0} ± 0.73%
7	LRH = HRH ≠ RD ≠ CD	1.06×10^{3} ± 0.59%
8	LRH ≠ HRH ≠ RD ≠ CD	2.14×10^{2} ± 0.33%

Note: All models include random effect of site.

Table S2: Posterior distributions of the parameter estimates of best model for Language and Reasoning factor (Model 7).

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.234	-0.152	-0.111	-0.070	0.009
group – LRH/HRH	0.080	0.120	0.140	0.159	0.197
group - RD	-0.063	-0.017	0.007	0.031	0.077
group - CD	-0.248	-0.181	-0.147	-0.112	-0.044

Note: Model includes random effect of site. Mu represents the grand mean.

Table S3: Bayesian model comparison for CBCL internalizing subscale.

Model no.	Equality constraint	Bayes Factor
1	LRH = HRH = RD = CD	1
2	LRH ≠ HRH = RD = CD	1.59×10^{76} ± 0.22%
3	LRH = HRH ≠ RD = CD	2.50×10^{10} ± 0.47%
4	LRH = HRH = RD ≠ CD	5.20×10^{3} ± 0.51%
5	LRH ≠ HRH ≠ RD = CD	3.76×10^{76} ± 0.31%
6	LRH ≠ HRH = RD ≠ CD	2.39×10^{77} ± 0.78%
7	LRH = HRH ≠ RD ≠ CD	1.75×10^{10} ± 0.33%
8	LRH ≠ HRH ≠ RD ≠ CD	2.67×10^{76} ± 0.39%

Note: All models include random effect of site.

Table S4: Posterior distributions of the parameter estimates of best model for CBCL internalizing subscale (Model 6).

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.339	0.614	1.090	1.593	2.540
group - LRH	-5.371	-4.879	-4.630	-4.376	-3.899
group - HRH/RD	0.542	1.007	1.251	1.501	1.981
group - CD	2.038	2.920	3.374	3.826	4.703

Note: Model includes random effect of site. Mu represents the grand mean.

Table S5: Bayesian model comparison for CBCL externalizing subscale.

Model no.	Equality constraint	Bayes Factor
1	LRH = HRH = RD = CD	1
2	LRH ≠ HRH = RD = CD	5.71×10^{67} ± 0.33%
3	LRH = HRH ≠ RD = CD	9.75×10^{14} ± 0.47%
4	LRH = HRH = RD ≠ CD	3.41×10^{6} ± 0.66%
5	LRH ≠ HRH ≠ RD = CD	2.88×10^{69} ± 0.84%
6	LRH ≠ HRH = RD ≠ CD	1.53×10^{70} ± 0.34%
7	LRH = HRH ≠ RD ≠ CD	2.87×10^{15} ± 0.37%
8	LRH ≠ HRH ≠ RD ≠ CD	8.86×10^{69} ± 0.36%

Note: All models include random effect of site.

Table S6: Posterior distributions of the parameter estimates of best model for CBCL externalizing subscale (Model 6).

Parameter	2.5%	25%	50%	75%	97.5%
mu	0.136	1.040	1.493	1.943	2.853
group - LRH	-5.333	-4.867	-4.625	-4.382	-3.914
group - HRH/RD	-0.087	0.362	0.602	0.842	1.306
group - CD	2.756	3.592	4.019	4.458	5.293

Note: Model includes random effect of site. Mu represents the grand mean.

Table S7: Bayesian model comparison for Language and Reasoning factor when selecting only LRH, HRH and RD.

Model no.	Equality constraint	Bayes Factor
9	LRH = HRH = RD	1
10	LRH ≠ HRH = RD	0.24 ± 0.57%
11	LRH = HRH ≠ RD	10.8 ± 0.45%
12	LRH ≠ HRH ≠ RD	2.99 ± 0.81%

Note: All models include random effect of site and covariates of CBCL internalizing and externalizing subscales.

Table S8: Posterior distributions of the parameter estimates of best model for Language and Reasoning factor when selecting only LRH, HRH and RD (Model 11).

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.142	-0.067	-0.028	0.009	0.084
group – LRH/HRH	0.017	0.043	0.057	0.071	0.097
group – RD	-0.097	-0.071	-0.057	-0.043	-0.017

Note: Model includes random effect of site. Mu represents the grand mean.

Table S9: Bayesian model comparison for Cognitive Flexibility factor.

Model no.	Equality constraint	Bayes Factor
1	LRH = HRH = RD = CD	1
2	LRH ≠ HRH = RD = CD	0.028 ± 0.38%
3	LRH = HRH ≠ RD = CD	0.991 ± 0.58%
4	LRH = HRH = RD ≠ CD	0.237 ± 0.87%
5	LRH ≠ HRH ≠ RD = CD	0.064 ± 0.43%
6	LRH ≠ HRH = RD ≠ CD	0.016 ± 0.28%
7	LRH = HRH ≠ RD ≠ CD	0.214 ± 0.40%
8	LRH ≠ HRH ≠ RD ≠ CD	0.018 ± 0.56%

Note: All models include random effect of site.

Table S10: Bayesian model comparison for Cognitive Flexibility factor when selecting only LRH, HRH and RD.

Model no.	Equality constraint	Bayes Factor
9	LRH = HRH = RD	1
10	LRH ≠ HRH = RD	3.78×10^{-2} ± 0.42%
11	LRH = HRH ≠ RD	2.89×10^{-1} ± 0.34%
12	LRH ≠ HRH ≠ RD	2.71×10^{-2} ± 0.72%

Note: All models include random effect of site and covariates of CBCL internalizing and externalizing subscales.

Table S11: Bayesian model comparison for Memory Recall factor.

Model no.	Equality constraint	Bayes Factor
1	LRH = HRH = RD = CD	1
2	LRH ≠ HRH = RD = CD	0.219 ± 0.44%
3	LRH = HRH ≠ RD = CD	1.445 ± 0.55%
4	LRH = HRH = RD ≠ CD	2.181 ± 0.38%
5	LRH ≠ HRH ≠ RD = CD	0.717 ± 0.61%
6	LRH ≠ HRH = RD ≠ CD	2.056 ± 0.71%
7	LRH = HRH ≠ RD ≠ CD	0.868 ± 0.79%
8	LRH ≠ HRH ≠ RD ≠ CD	0.511 ± 0.45%

Note: All models include random effect of site

Table S12: Posterior distributions of the parameter estimates of best model for Memory Recall factor (Model 4).

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.247	-0.156	-0.111	-0.066	0.025
group – LRH/HRH/RD	0.020	0.076	0.106	0.136	0.192
group - CD	-0.192	-0.136	-0.106	-0.076	-0.020

Note: Model includes random effect of site. Mu represents the grand mean.

Table S13: Bayesian model comparison for Memory Recall factor when selecting only LRH, HRH and RD.

Model no.	Equality constraint	Bayes Factor
9	LRH = HRH = RD	1
10	LRH ≠ HRH = RD	0.046 ± 0.41%
11	LRH = HRH ≠ RD	0.093 ± 0.63%
12	LRH ≠ HRH ≠ RD	0.007 ± 0.84%

Note: All models include random effect of site and covariates of CBCL internalizing and externalizing subscales.

Table S14: Bayesian model comparison for sMRI cv1.

Model no.	Equality constraint	Bayes Factor
1	LRH = HRH = RD = CD	1
2	LRH ≠ HRH = RD = CD	0.02 ± 0.82%
3	LRH = HRH ≠ RD = CD	3.04 ± 0.47%
4	LRH = HRH = RD ≠ CD	5.94 ± 0.33%
5	LRH ≠ HRH ≠ RD = CD	0.28 ± 0.56%
6	LRH ≠ HRH = RD ≠ CD	0.35 ± 0.34%
7	LRH = HRH ≠ RD ≠ CD	3.00 ± 0.25%
8	LRH ≠ HRH ≠ RD ≠ CD	0.33 ± 0.44%

Note: All models include random effect of site

Table S15: Posterior distributions of the parameter estimates of best model for sMRI CV1 (Model 4).

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.330	-0.172	-0.093	-0.010	0.149
group – LRH/HRH/RD	0.038	0.100	0.131	0.163	0.222
group – CD	-0.222	-0.163	-0.131	-0.100	-0.038

Note: Model includes random effect of site. Mu represents the grand mean.

Table S16: Bayesian model comparison for cognition CV1.

Model no.	Equality constraint	Bayes Factor	
1	LRH = HRH = RD = CD	1	
2	LRH ≠ HRH = RD = CD	0.87×10^{0}	± 0.25%
3	LRH = HRH ≠ RD = CD	4.54×10^{3}	± 0.43%
4	LRH = HRH = RD ≠ CD	1.54×10^{2}	± 0.62%
5	LRH ≠ HRH ≠ RD = CD	6.58×10^{2}	± 0.38%
6	LRH ≠ HRH = RD ≠ CD	8.25×10^{0}	± 0.36%
7	LRH = HRH ≠ RD ≠ CD	5.83×10^{3}	± 0.32%
8	LRH ≠ HRH ≠ RD ≠ CD	9.34×10^{2}	± 0.27%

Note: All models include random effect of site.

Table S17: Posterior distributions of the parameter estimates of best model for cognition CV1 (Model 7).

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.307	-0.199	-0.146	-0.092	0.018
group – LRH/HRH	0.120	0.169	0.196	0.221	0.271
group - RD	-0.083	-0.023	0.009	0.040	0.102
group - CD	-0.340	-0.250	-0.203	-0.158	-0.071

Note: Model includes random effect of site. Mu represents the grand mean.

Table S18: Bayesian model comparison accounting for parental history of depression in remitted and currently depressed children.

Term	Bayes Factor
Language and reasoning	
Group	1
Group + Parental Depression	0.462 ± 0.88%
Group + Parental Depression + Group*Parental Depression	0.457 ± 1.74%
Cognitive Flexibility	
Group	1
Group + Parental Depression	0.137 ± 0.74%
Group + Parental Depression + Group*Parental Depression	0.029 ± 1.45%
Memory Recall	
Group	1
Group + Parental Depression	0.353 ± 1.07%
Group + Parental Depression + Group*Parental Depression	0.063 ± 1.61%
Brain CV1	
Group	1
Group + Parental Depression	0.506 ± 1.14%
Group + Parental Depression + Group*Parental Depression	0.091 ± 0.75%
Cognition CV1	
Group	1
Group + Parental Depression	0.136 ± 0.94%
Group + Parental Depression + Group*Parental Depression	0.035 ± 2.25%

Note: All models include random effect of site.

Table S19: Bayesian model comparison testing for potential gender effects in cognitive factors and canonical variates.

Term	Bayes Factor
Language and reasoning	
Group (LRH = HRH ≠ RD ≠ CD; Model 7 in Table S1)	1
Group + Gender	0.236 ± 1.87%
Group + Gender + Group*Gender	0.009 ± 1.98%
Cognitive Flexibility	
Group (LRH = HRH = RD = CD; Model 1 in Table S9)	1
Group + Gender	0.035 ± 0.92%
Memory Recall	
Group (LRH = HRH = RD ≠ CD; Model 4 in Table S11)	1
Group + Gender	6.09×10^{12} ± 1.92%
Group + Gender + Group*Gender	9.38×10^{11} ± 1.79%
Brain CV1	
Group (LRH = HRH = RD ≠ CD; Model 4 in Table S14)	1
Group + Gender	0.047 ± 1.91%
Group + Gender + Group*Gender	0.008 ± 1.93%
Cognition CV1	
Group (LRH = HRH ≠ RD ≠ CD; Model 7 in Table S16)	1
Group + Gender	0.129 ± 1.05%
Group + Gender + Group*Gender	0.003 ± 0.95%

Note: All models include random effect of site.

Table S20: Posterior distributions of the parameter estimates for gender differences in memory recall.

Parameter	2.5%	25%	50%	75%	97.5%
mu	-0.231	-0.140	-0.096	-0.050	0.040
gender – F	0.079	0.096	0.104	0.113	0.129
gender – M	-0.129	-0.113	-0.104	-0.096	-0.079
group – LRH/HRH/RD	0.008	0.063	0.091	0.120	0.176
group - CD	-0.176	-0.120	-0.091	-0.063	-0.008

Note: Model includes random effect of site. Mu represents the grand mean.

Supplemental Figures

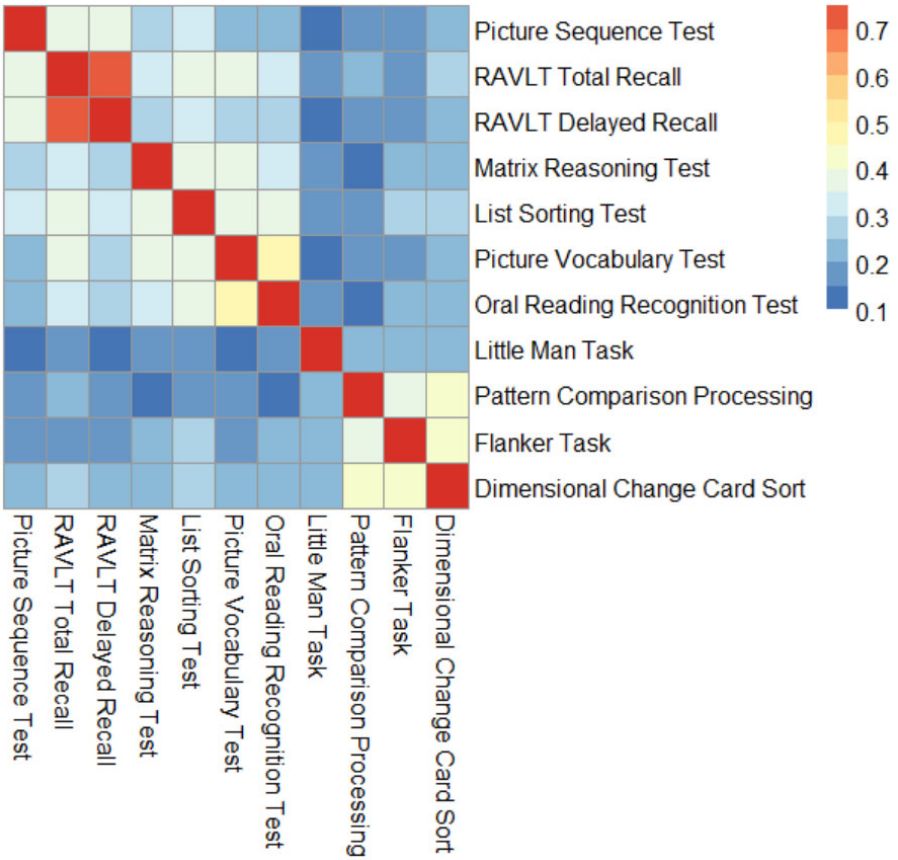

Figure S1. Intercorrelations between cognitive variables. Evidently, some cognitive variables are correlated with a group of other metrics, but not with measures outside that group. This suggests that more general, underlying cognitive factors might be present. RAVLT = Rey Auditory Verbal Learning Test. Refer to Methods for descriptions of tasks and associated variables of interest.

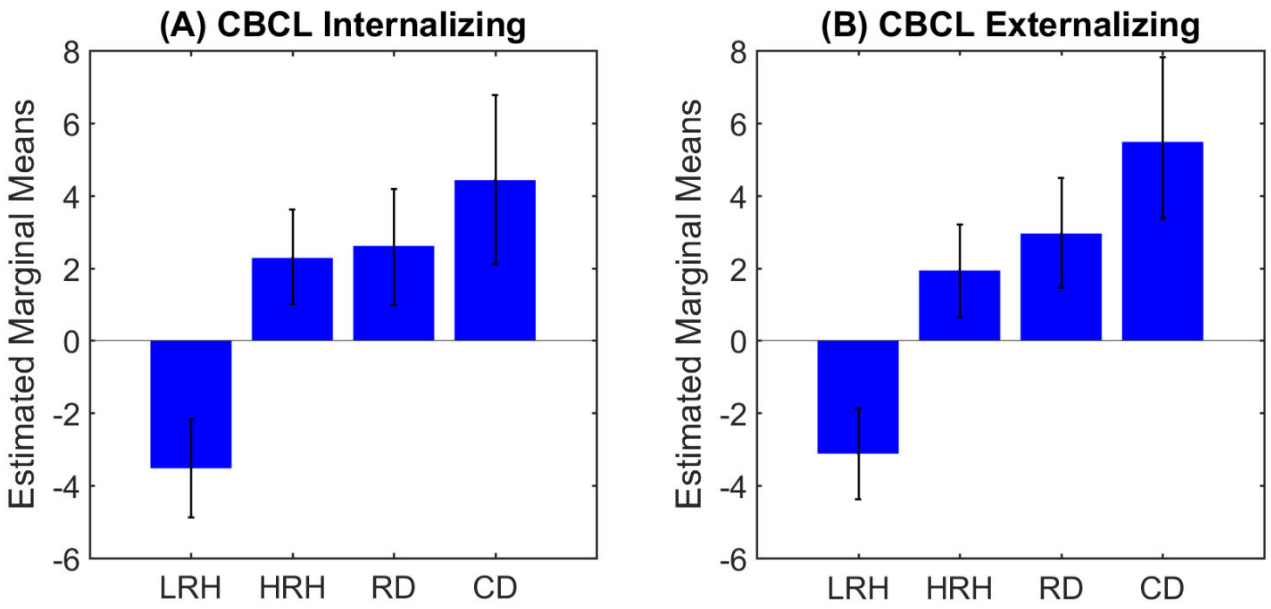

Figure S2. (A) Group comparison of the Achenbach Child Behavior Checklist (CBCL) internalizing subscale. Bayesian model comparison found that model 6 (LRH≠HRH=RD≠CD) outperformed all other hypotheses, indicating the presence of a non-specific vulnerability/trait, as well as a specific state, effect. **(B)** For CBCL externalizing, Models 6 (LRH≠HRH=RD≠CD) and 8 (LRH≠HRH≠RD≠CD) provided the most parsimonious account of the data. Thus, a specific state effect exists, but more data are required to determine whether these reflect a non-specific vulnerability/trait impairment, or unique vulnerability and trait effects. Models included a random effect of site. Vertical lines indicate lower and upper bound of the 95% highest posterior density. Abbreviations: LRH = low-risk healthy, HRH = high-risk healthy, RD = remitted depressed, CD = currently depressed.

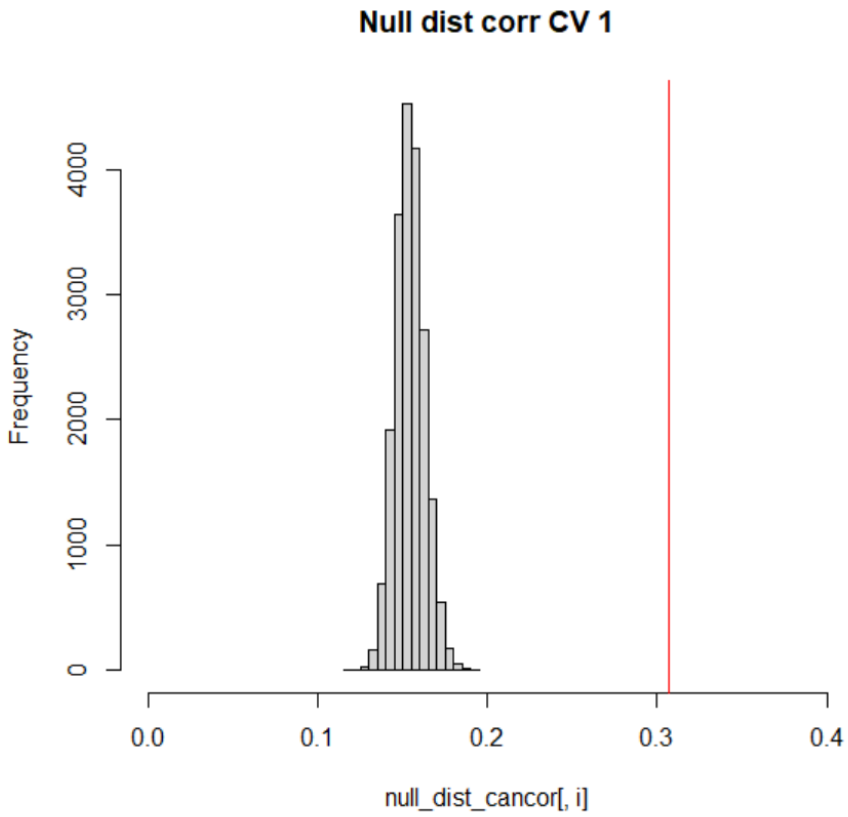

Figure S3. Null distribution of the first canonical pair from canonical correlation analysis on 20,000 permutations of the dataset. The red vertical line indicates the original correlation value.

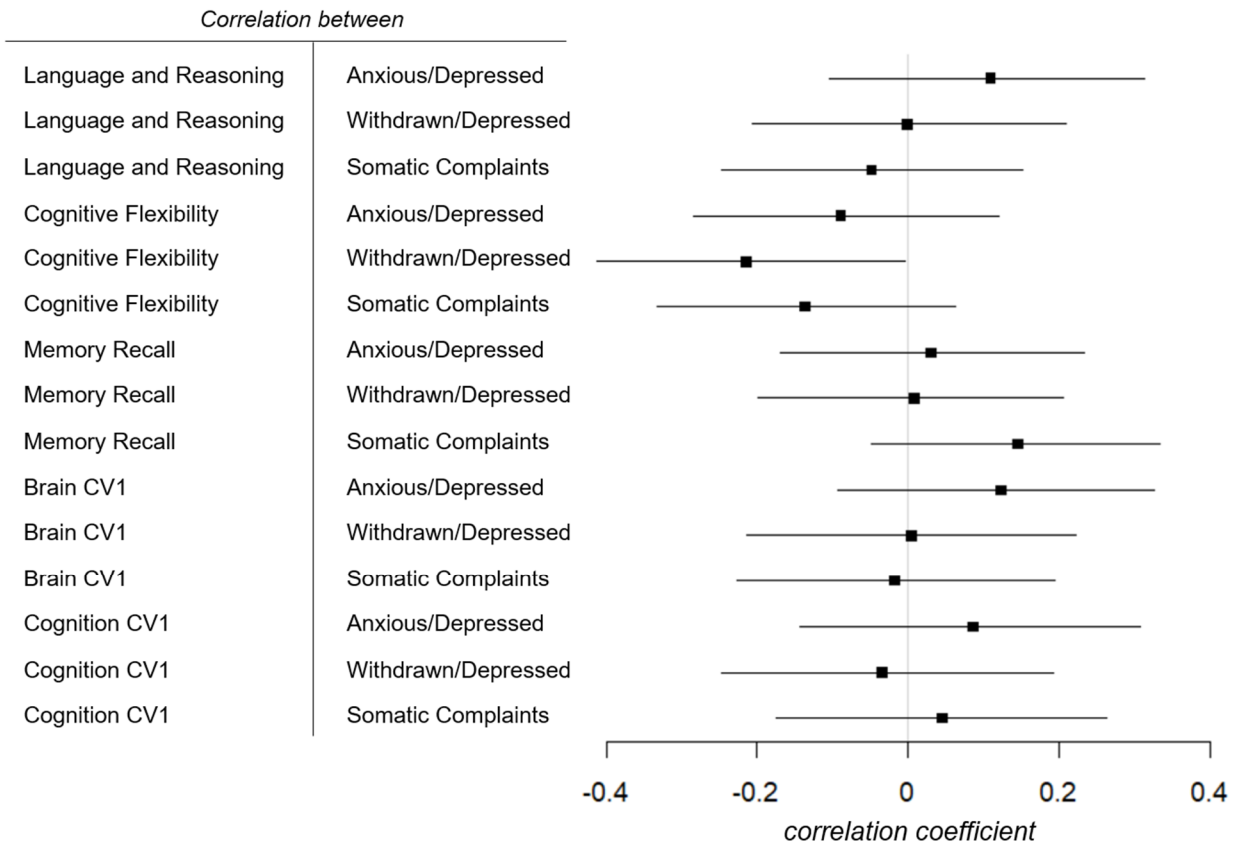

Figure S4. Bayesian correlation coefficients between CBCL depressive syndromes and cognitive factors.

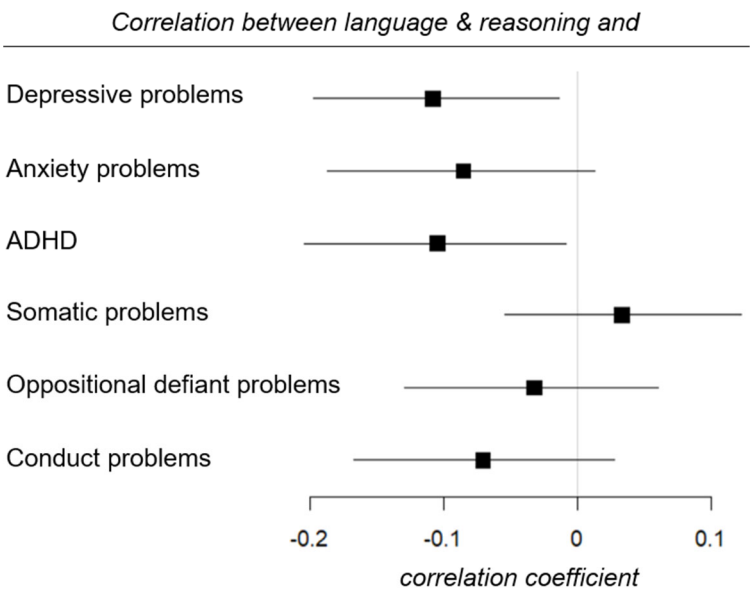

Figure S5. Bayesian correlation coefficients between CBCL DSM-oriented scales and language and reasoning.

Supplemental References

Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. Dev Cogn Neurosci 2018; 32: 55–66.

Baune BT, Fuhr M, Air T, Hering C. Neuropsychological functioning in adolescents and young adults with major depressive disorder--a review. Psychiatry Res 2014; 218: 261–71.

Favre T, Hughes C, Emslie G, Stavinoha P, Kennard B, Carmody T. Executive functioning in children and adolescents with Major Depressive Disorder. Child Neuropsychol J Norm Abnorm Dev Child Adolesc 2009; 15: 85–98.

Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 1997; 36: 980–8.

Klimkeit EI, Tonge B, Bradshaw JL, Melvin GA, Gould K. Neuropsychological deficits in adolescent unipolar depression. Arch Clin Neuropsychol Off J Natl Acad Neuropsychol 2011; 26: 662–76.

Lee RSC, Hermens DF, Porter MA, Redoblado-Hodge MA. A meta-analysis of cognitive deficits in first-episode Major Depressive Disorder. J Affect Disord 2012; 140: 113–24.

McDermott LM, Ebmeier KP. A meta-analysis of depression severity and cognitive function. J Affect Disord 2009; 119: 1–8.

Peters AT, Jacobs RH, Crane NA, Ryan KA, Weisenbach SL, Ajilore O, et al. Domain-specific impairment in cognitive control among remitted youth with a history of major depression. Early Interv Psychiatry 2017; 11: 383–92.

Robertson HA, Kutcher SP, Lagace DC. No evidence of attentional deficits in stabilized bipolar youth relative to unipolar and control comparators. Bipolar Disord 2003; 5: 330–9.

Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med 2014; 44: 2029–40.

Smith DJ, Muir WJ, Blackwood DH. Neurocognitive impairment in euthymic young adults with bipolar spectrum disorder and recurrent major depressive disorder. Bipolar Disord 2006; 8: 40–6.

Snyder HR. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychol Bull 2013; 139: 81–132.

Tulsky DS, Carlozzi N, Chiaravalloti ND, Beaumont JL, Kisala PA, Mungas D, et al. NIH Toolbox Cognition Battery (NIHTB-CB): List Sorting Test to Measure Working Memory. J Int Neuropsychol Soc 2014; 20: 599–610.

Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Weintraub S. II. NIH toolbox cognition battery (CB): Measuring executive function and attention. Monogr Soc Res Child Dev 2013; 78: 16–33.