The relationship between apparent brain age and memory-related changes in early adolescence

Ina Drabløs

Submitted as Master´s Thesis at the Department of Psychology

University of Oslo

Spring 2023

Acknowledgements

First and foremost, I would like to thank my incredible supervisors Esten Leonardsen and Andreas Dahl for their patience and support throughout this year. This project would not have been possible without you. To my main supervisor Esten, thank you for introducing me to the deep learning universe of which I had no prior knowledge. I am truly grateful for all your insights and your guidance through technical hurdles. It has been challenging and exiting. And to my co-supervisor Andreas, thank you for all your invaluable advice throughout this project. I am tremendously thankful for your efficient and persistent feedback on this thesis, and for always checking in on me.

I would also like to thank Lars Westlye and NORMENT for granting me access to this data and giving me the opportunity to be involved with a great research community and exiting projects. Thank you also to all ABCD participants for donating their time, and the ABCD consortium for data collection.

Finally, a thank you to my fellow master-students for contributing to a positive and interesting academic environment over the past two years. A special thanks to Linn, as this final stretch would not have been as manageable without being able to share our frustrations and motivations.

Abstract

Author: Ina Drabløs

Title: The relationship between apparent brain age and memory-related changes in early adolescence

Supervisors: Esten H. Leonardsen[1] & Andreas Dahl[1]

Disclaimer: The idea and design for the study was provided by me. All data material was obtained through the ABCD Study® (http://dx.doi.org/10.15154/1523041). Preprocessing of structural MRI data was conducted in collaboration with my supervisor. The brain age model applied in the present thesis is based on the SFCN-reg by Leonardsen et al. (2022) and was implemented on the included neuroimaging material in collaboration with my supervisor. All data handling and cleaning of outcome measures and covariates was performed by me. The conceptualization and interpretation of statistical models was done in collaboration with supervisors but implemented by me. All writing was conducted by me in full.

Background: Changes in both brain morphology and memory abilities during adolescent development are well established, however, the linkage between the two is not fully understood. Recent deep learning models have shown to be effective and accurate in producing subject-specific predictions for healthy biological brain aging, but it is not known if the effects of brain age on different memory substrates are shared or unique. The present study explores the relationship between working memory, episodic memory, and predicted brain age during early adolescent development. **Methods:** Data material was obtained from two timepoints spaced two years apart, with an average age of 10 years at baseline (n = 11,343) and 12 years at follow up (n = 9975). Measures from three separate tasks form the basis of working memory and episodic memory abilities, and a state-of-the art deep learning model was applied to minimally preprocessed T1w MRI data. The objective was to assess the association between brain age and distinct memory abilities during adolescent development. **Results:** Linear mixed effect models revealed no significant main associations of memory and brain age. However, a significant interaction effect of brain age and time was observed on episodic memory. **Conclusions:** The results from my study show an association between memory, aging and maturation, but does not reveal whether brain aging beyond what is captured in other variables plays a significant role.

[1] Department of psychology, University of Oslo, Norway

Table of contents

Introduction

The transition from childhood to adolescence involves significant physical, cognitive, and behavioral changes. During this period the brain undergoes a process of structural and functional alterations lasting until the mid-twenties, increasing its ability to handle complex information (Arain et al., 2013). This is also a period of self-discovery, academic stress, and intricate social relationships, placing greater demands on cognitive capacities. The ability to navigate through, learn from, and remember the vital aspects of these experiences is highly dependent on different memory systems, each with its own developmental trajectory (Bouyeure & Noulhiane, 2020). Despite previous demonstrations of cortical growth patterns being related to intelligence (Shaw et al., 2006) and general cognitive ability (Walhovd et al., 2016), the interactions between brain maturation and specific memory abilities during adolescent development are less clear. The overall aim of the present study is to explore this complex association between early adolescent brain maturation and memory development, to expand our understanding of how this relationship changes over time.

Memory

Types of memory

When information from the surroundings is perceived, attended to, and processed by the brain, memory is involved. Some of the processed information might be available for later recollection, although not all episodes will lead to a lasting memory trace (Sneve et al., 2015). During the last century, several theories have conceptualized memory as the processing of information, and not simply a linear process from perception to storage (Miller, 1956; Baddeley & Hitch, 1974; Atkinson & Shiffrin, 1968). Such perspectives consider the development of the brain to be essential for gradual increases in processing abilities, so that children and adolescents will continue to improve in how they perceive and store information as they mature. These theories also share the perspective of distinguishing between types of memory. Working memory (WM) has limited capacity and comprise the input that can be attended to and manipulated, while the encoding and storing of this information is conceptualized as long-term memory (LTM). Memory of stored information is further distinguished based on type of input, where the recollection of episodes and events specifically is referred to as episodic memory (EM). Multiple systems and processes are involved in memory formation, where different types of memory engage both specific and overlapping brain regions (Lugtmeijer et al., 2019). Generally, WM and EM performance in young adults are highly correlated, and also correlate with performances in other cognitive

domains such as fluid intelligence and language skills. Understanding the development of specific memory abilities and brain maturation during early adolescence is therefore dependent on using tasks that attempts to tease apart a specific memory construct.

Measurements of working memory and episodic memory

A task measuring WM must assess the amount of information that can be actively maintained, thus challenging memory capacity (Jeneson & Squire, 2012). Most such tasks also measure cognitive control, the intentional selection and suppression of information, as well as preservation and retrieval of relevant information, all components that reflect WM abilities (Pelegrina et al., 2015). A well-established task designed to assess WM is the N-back task, first created by Wayne Kirchner (1958). The original version was employed as a measurement of short-term retention, the ability to temporarily store and organize both ingoing and outgoing information. During the task, participants must keep up with presented information increasing in difficulty, forcing them to exert their full memory capacity. Participants respond by pressing a button to indicate whether the current item matches the item displayed "n" trials in advance, continuously having to attend to and update the information presented. Different conditions in the task allow for several measurements of accuracy, suited for all ages. Other tasks commonly applied as measures of WM typically involve sorting material as it is presented, trying to probe the maximum load of information that can be retained and manipulated by the brain. When compared to sorting tasks, outcomes from higher load conditions during the N-back task indicate similar results, demonstrating its validity as a measure of WM (Rosenberg et al., 2020).

To assess EM abilities either attention must be diverted, or WM capacity must be exceeded. This can be accomplished by increasing memory load during the task or inducing retention intervals with delayed recall of information. Tasks measuring EM typically utilize several aspects of processing and information storage and may differ in the amount of time before recall. Depending on the task, one might have to account for effects of stimuli position and recency effect (Gavett & Horwitz, 2012), as well as the effect of strategy on memory performance by repeated trials (Waris et al., 2021). A widely used and accepted task for EM measurement is the Rey Auditory Verbal Learning Test (RAVLT), where participants attempt to remember and recall a list consisting of 15 items across several trials and with delayed recall (Karpouzian-Rogers et al., 2023). This task was performed by a group of older adults with exceeding EM abilities, who also completed several other neurocognitive tasks from the National Institute of Health Toolbox (NIHTB), including the Picture Sequence Memory task

for EM. Their results on both EM tasks were compared to a group of average performers to assess task validity. No differences were found in any other cognitive measures apart from the Picture Sequence task, displaying its validity to uniquely measure differences in EM abilities across individuals.

To assess rate of change in memory capacity and potential differences between memory types, the same task and stimuli can be applied to evaluate both WM and EM. In a study by Lugtmeijer et al. (2019), adult participants were presented with the higher-load 2-back condition of the N-back task and a subsequent surprising task as a measure of EM. The participants had to determine whether the presented object was displayed in the same corner of the screen as previously shown. In younger adults, performance on the 2-back task showed significant correlation with subsequent memory performance, indicating high correlation between WM and EM. However, a correct response during the WM task did not influence memory of the same object in the following task, thus not enhancing encoding of specific information. Furthermore, a difference in response time between the two tasks, and the unawareness of the second task, makes these results difficult to interpret as distinctive to WM and EM. It is thus not clear how much overlap exist between specific memory abilities, nor how they co-develop with brain maturation during early adolescence.

Development

Structural brain development

Structural changes in the brain accompanying development and aging are characterized either as progressive or regressive processes, with differing rates depending on specific regions or tissue types in the brain (Franke & Gaser, 2019). Changes in brain morphology are commonly measured using properties such as volume, surface area, or thickness, which can be estimated from neuroimaging both regionally and globally across the brain. Despite decades of research on brain structure and function, standardized growth charts for typical brain development during childhood and adolescence have been scarce. However, the accumulated number of studies involving magnetic resonance imaging (MRI) across age groups indicates relatively high individual stability in brain morphology across longitudinal assessments (Bethlehem et al., 2022). Furthermore, there is a general pattern of progressive increase in white matter continuing into adult life, while cortical grey matter shows region-specific declines initiated before adolescence (Giedd et al., 1999).

When looking at normative trajectories for brain development, fronto-temporal regions are the latest to reach their peak of total grey matter volume (GMV) (Bethlehem et al., 2022). These

regions are especially important for memory abilities (Balaconi, 2013; Pelegrina et al., 2015). Inter-individual variability of cortical GMV is at its highest at approximately 4 years of age, much earlier than white matter volume (WMV), displaying most variability during the 40s. Additionally, WMV peaks much later at an average of 28.7 years, reflecting continuing myelination and synaptic formation that shapes the functioning of the brain. In the meta-study by Bethlehem et al. (2022), total cerebral volume peaked at 12.5 years (see fig. 1). This is in line with findings in a younger sample from 2012 (Brown et al.), where there was an increase in total cortical area until age 12.3. Findings in the young sample also revealed a lot of inter-individual variability in hippocampal increase, a region highly involved in memory formation. The average peak of hippocampal volume was at age 14.2, followed by minimal decline in subsequent years. Brown et al. (2012) used a multimodal approach to show that certain aspects of brain development are closely tied to chronological age, where the age of each subject could be determined by comparison to smooth nonlinear age trajectories based on the total group mean and covariance. 92% of the variability in individual differences could be explained by this neuroanatomical model, with a mean prediction error of only 1.03 years, but there were still instances where subjects age was over- or under-predicted. Nonetheless, this demonstrates the possibility of using structural brain measurements as a biomarker for development, as the prediction error could reflect meaningful variability between individuals.

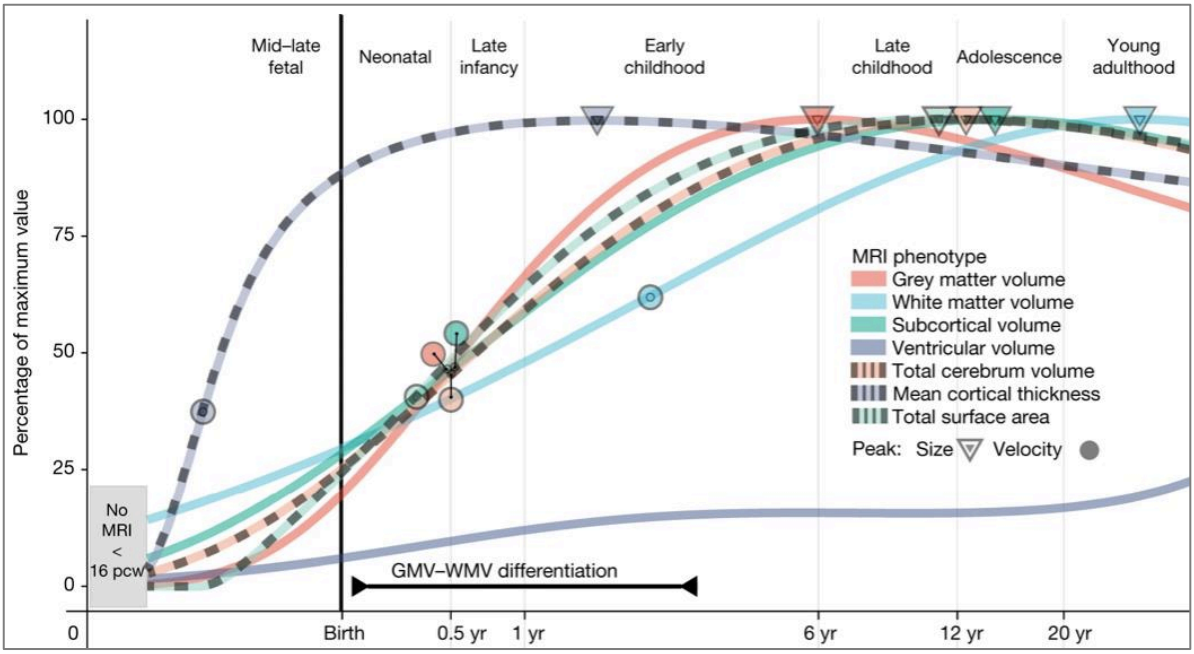

Figure 1. Normative developmental trajectories of average global tissue type.
Circles represent peak rate of growth. Triangles represent peak volume. Retrieved and adapted from Bethlehem et al., 2022.

Changes in cortical thickness is also considered to reflect brain development and is measured by MRI as the distance between the cortical/pial border to the inner gray matter border (Zhou et al., 2015). A general decline is seen from childhood into young adulthood, thought to be caused by cell death of excessive neurons. Rate of change in cortical thickness (mm/year) is defined as the difference of thickness between scans, divided by the age difference between scans. Longitudinal studies indicate an s-shaped curve of continuing reduction in cortical thickness across all brain areas from childhood into young adulthood (Fuhrmann et al., 2022). The rate of thinning is relatively stable from mid childhood to about 9-11 years, before this pattern of thinning is accelerated during adolescence with the most rapid decline seen on average at age 14. However, both regional and individual differences have been observed (Mills et al., 2021).

Most studies display the same pattern of regional cortical thickness across the brain, with the medial frontal lobe, temporal areas and bilateral insula having greater thickness, as compared to a thinner cortex in the occipital and parietal areas. Regardless of these variations, all regions show the same pattern of declining thickness as described above, meaning proportionally greater decreases in frontal and temporal areas (Zhou et al., 2015). There are, however, some differences between the sexes. Overall, males have significantly thicker cortexes, and the pattern and rate of thinning differs slightly between males and females. The sex difference is at its largest during childhood, where females typically display a more continuous rate of thinning from childhood through adolescence, while males show a higher decline of cortical thickness during adolescence (Zhou et al., 2015). Structural brain changes and the pattern of which these tissue types increase and decrease, is likely the result of myelination and synaptic pruning, which improves the efficiency of information processing in the brain. This coincides with other physical- and behavioral changes occurring during this period, but it has not yet been established how the rate of structural changes on an individual level affects cognition and memory development.

Brain and memory co-development

Adolescence is by many considered to be a critical period for the development of higher-order cognition, due to the significant changes seen in both brain and behavior (Larsena & Luna, 2018). Although changes in both brain morphology and memory abilities during this stage are well established, the linkage between the two is not fully understood (Schneider & Ornstein, 2019). Complex mechanisms throughout the brain are implicated in all memory processes, although some regions have been established as more important for

specific memory abilities. For instance, the frontal lobes have a crucial role for memory, and are highly involved in attentional maintenance, planning, organizing, and strategies used in information manipulation (Balaconi, 2013). These functions are especially important for WM, where abilities increase rapidly during the first year of life and continues to develop until late adolescence (Rosenberg et al., 2020).

Previous studies applying the 2-back condition of the N-back task revealed age-related increases in performance, indicating increased attentional control and ability to manage interference as children and adolescents mature (Pelegrina et al., 2015; Rosenberg et al., 2020). The rate of increase diminishes with age, with more individual variance. In addition, a gender difference is commonly observed in this task, with females having a higher percentage of correct responses, but males responding faster. There has also been established a relationship between performance on a WM-targeted list sorting task and brain activation during the 2-load condition of the N-back task (Rosenberg et al., 2020). Children who performed better on list sorting also displayed increased activity in frontoparietal brain regions during the 2-load condition, reflecting its involvement in sustaining and manipulating information. Results also show that the frontoparietal activity displayed during WM-targeted tasks is not significantly associated with activation as measured by other tasks and content, and instead has a domain-specific signature.

Frontal areas of the brain are also crucial for several of the features composing EM, and evidence indicate that the rate of cortical thinning is linked with activity in the prefrontal cortex during memory recall (Mechie, Plaisted-Grant & Cheke, 2021). Retrieving memories require the selection and organization of information and is highly dependent on strategies employed during the initial encoding. The dorsolateral prefrontal cortex is especially important when little to no cues are given. This region also displays more pronounced age-related differences during childhood, and findings indicate a linear increase in performance between ages 8 and 24. Another feature of EM is the association of unique components of information, and to distinguish this from other experiences. Differences in this associative feature of EM has been linked to age-related changes in hippocampal volume, located in the medial temporal lobes (Tamnes et al., 2018). Evidence suggests a non-linear developmental pattern for increases in whole hippocampal volume in early adolescence. Furthermore, the integration between the cortex and hippocampus is crucial for sustained memory (Sneve et al., 2015). To summarize, there is evidence supporting the co-development of memory and brain, particularly highlighting the importance of the prefrontal cortex and the medial temporal lobes. Nevertheless, the intricate structural changes occurring both regionally and globally

across separate tissue types in the brain, and memory abilities partially overlapping when measured by tasks, makes it complicated to establish clear links between the two. Rather than isolating specific brain regions and thus removing potential influences from left-out regions, models utilizing the whole brain to assess brain maturity can be applied.

Brain age

A decade of machine learning algorithms and brain age

The multimodal approach by Brown et al. (2012) previously mentioned shows that certain aspects of structural brain changes during development can be utilized by neuroanatomical models to assess brain maturation based on nonlinear age trajectories. Since then, further technological advancements have prompted the use of machine learning techniques to assess individual brain development and brain health, as these is methods can handle large datasets and produce subject-specific predictions for healthy biological brain aging (Franke & Gaser, 2019). Computational models are trained on large sets of data and learn to recognize patterns of brain tissue in the voxels of MRI data to represent normative reference curves based on age-specific morphological variation, which subsequently allows for predictions on a single subject level. This approach has shown promise as an intuitive and precise biomarker for general brain health and has produced great results in predicting the age of a subject from morphological variation in the brain (Cole et al., 2017). The difference between the visually apparent age of the brain encoded in the prediction and the chronological age allows us to place individuals in relation to a normative brain aging trajectory. In the literature this difference is often referred to as the brain age delta (Peng et al., 2021). A positive delta, where the brain age model predicts a higher age than the subjects chronological age, indicates an older appearing brain.

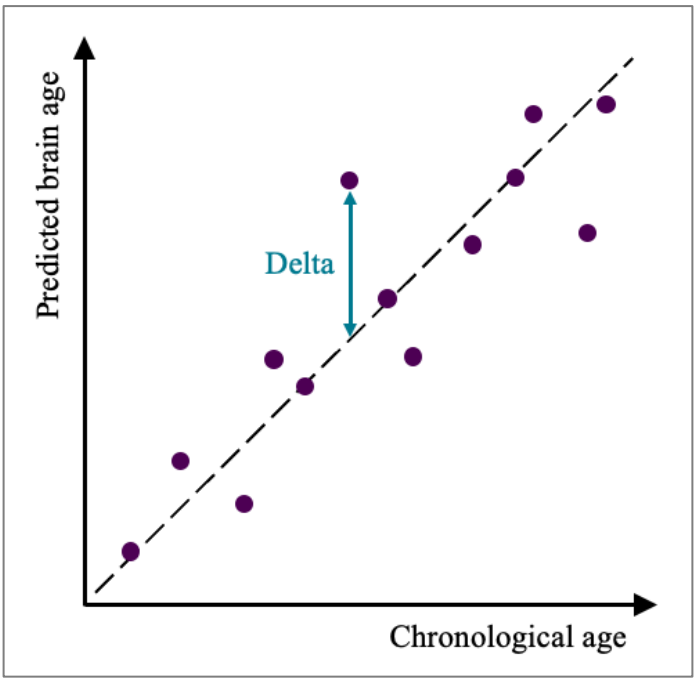

Figure. 2. Delta illustrated as the difference between predicted brain age and chronological age.

Based on their review of all studies utilizing brain age methods over the past decade, Franke & Gazer (2019) trained and cross-validated a brain age model to compare with previous results. The model was generated based on a reference sample consisting of 394 healthy children and adolescents between ages 5 and 18 years. When training the model on data from five MRI scanners they achieved a mean absolute error (MAE) of 1.1-1.3 years and overall correlations between $r = 0.90$-0.95 on data collected from a scanner unseen by the model during training, comparable to previous studies applying brain age models in young samples (Franke & Gazer, 2019.). MAE is the difference between the measured value, in this case the predicted brain age, and the true value, here the chronological age. Introducing their model to two other independent samples with age ranges between 19-86 and 20-59 years yielded a MAE of 5 years and $r = 0.92$, showing decent generalizability compared to previous studies where MAEs range between 4.3-13.5 years and correlations between $r = 0.43$-0.97.

The latter results indicate that machine learning models predicting brain age vary in accuracy. This can be seen both when the same model is applied to different datasets, but also when different models are applied to the same morphological data (Han et al., 2022). The characteristics of the training sample influence the generalizability to novel samples, a potential explanation for the difference in MAE observed between prediction accuracies in the younger and older samples previously described. In addition, the regional features of the brain vary in how they contribute to the brain age prediction error based on the choice of classification algorithm in the models (Valizadeh et al., 2017). Compared to traditional machine learning techniques, more recent deep learning models based on neural networks have shown to be both effective and accurate, yielding more precise results without having to design or select specific brain features for the models in advance (LeCun, Bengio & Hinton, 2015).

Deep learning: Simple Fully Convolutional Networks

Predictive approaches utilizing neural networks have shown success as a method in neuroimaging, for instance contributing to diagnostic decisions and treatment options (Yin, Li & Wu, 2022). Even though several models have yielded highly accurate predictions when comparing training and validation samples, a recurring issue with these increasingly complex models is overfitting (Bejani & Ghatee, 2021). This phenomenon occurs when the model learns overly specific details of the training sample, effectively treating noise as signal, resulting in a low gap between predicted and chronological age within the training sample. Consequently, this reduces the predictive performance when the model is applied to

unfamiliar datasets. To solve this one must employ rigorous validation and testing schemes to determine the optimal model size and complexity, such that the resulting model is able to generalize to other datasets. In addition to boosting performance on unseen data, increased ability to generalize also points towards the model makings its prediction based on informative patterns of brain structure rather than noise or insufficient data quality.

State-of-the art convolutional neural networks for processing 2-dimensional natural images are typically large, with up to hundred million parameters (Shin et al., 2016), requiring datasets consisting of millions of images to train. By reducing the number of model layers, less computational cost is required and the number of parameters to be estimated decreases, reducing the risk of overfitting (Liao & Carneiro, 2016). A model architecture containing fewer layers that has proven successful in realistic neuroimaging datasets limited to tens of thousands of scans is Simple Fully Convolutional Network (SFCN), where minimally preprocessed T1-weighted MRI data is used as input (Peng et al., 2021). The model performs a sequential computation with three main stages. Stage one consists of five blocks, each comprised of a convolutional layer, a batch normalization layer, a max pooling layer, and a rectified linear activation unit (ReLU). The convolutional layer uses a kernel to detect patterns in the input, sliding a window across regions of the image such that the patterns are identified across the full spatial range, producing a four-dimensional activation map (Goodfellow, Bengio. & Courville, 2016, p. 326-335). The last dimension of the activation maps represents the different patterns the model learns to look for. These activations are standardized using batch normalization (Liao & Carneiro, 2016) to stabilize the training process. Next, the ReLU-activation is applied to the normalized activations to add non-linearity, deciding whether values of specific nodes in the layer should be activated or ignored (Goodfellow et al., 2016, p. 164-172), effectively increasing the expressive power of the model. Finally, output from the activation maps is further reduced by the max pooling-operation, identifying important features based on maximum values in a given region and reducing the spatial resolution of the signal (Goodfellow et al., 2016, p. 335-339). The result is that each of these blocks will extract feature maps from the images, gradually capturing more complex patterns spanning larger regions in the image (Peng et al., 2021).

The next block contains a 1x1x1 convolutional layer, a batch normalization layer, and a ReLU activation. The intention of this intermediate block is to further increase the non-linearity of the output with respect to the input, while also reducing the dimensionality and computational load. Together these first two stages represent the process of converting the input images to a feature vector which encodes to what degree various visual patterns are

occurring in the input. In the last block an average pooling-layer is applied to smooth out the activations, and a dropout layer is added to increase generalizability. Finally, the original SFCN model maps the extracted features to a probability distribution of predicted age through a convolutional layer and a softmax activation layer.

The SFCN model was part of a brain age prediction competition in 2019 and was ranked first in the "most accurate age prediction while minimizing bias" category among 79 participating teams (Peng et al., 2021). A MAE of 1.36 was obtained in the training sample, which is slightly higher compared to the other models. However, it showed superior performance both in the test sample with a MAE of 2.14, and when comparing performance in the training and test sample with a difference in MAE of only 0.83. This is indicative of reduced overfitting and better generalization performance, likely due to the smaller model size and being able to extract useful data patterns. Based on the original SFCN model, further generalization and precision and has been achieved by replacing the final soft classification-layer with a pure regression layer predicting a single continuous value (Leonardsen et al., 2022).

Previous applications of brain age models

Significant associations between a high brain age delta and a variety of biomedical and lifestyle factors such as diabetes, alcohol consumption, smoking, high blood pressure and heart disease has been found in adult and older populations (Cole, 2020). In addition, accelerated brain aging is usually seen in conjunction with neurodevelopmental and neurodegenerative disorders such as schizophrenia, bipolar disorder, and multiple sclerosis, highlighting the potential use of brain age as a valuable tool in clinical neuroscience (Kaufmann et al., 2019). There is also evidence of overlapping genes between brain disorders and brain age delta, pointing towards heritability of brain age. Furthermore, brain age has been applied in studies conducted with adolescent, displaying a difference between the sexes likely related to pubertal development, as females typically reach puberty earlier than males (Brouwer et al., 2021; Holm et al., 2023). The developmental pattern is similar between the sexes, but on average females display a higher brain age delta between ages 12 and 19, before this discrepancy decreases and evens out before entering young adulthood.

Brain age has also been employed to predict memory performance in healthy adolescents. By applying a support vector regression (SVR) model to different MRI modalities is a young sample, Ullman et al. (2014) was able to predict developmental change in working memory performance. The SVR model outperformed the initial WM measures in

terms of prediction, suggesting better predictive ability from neuroimaging than cognitive measures alone. However, no studies to date have assessed the effects of adolescent brain age on both WM and EM at the same time. Here, I will employ a SFCN-based brain age model in a large-scale longitudinal sample of early adolescents. Presumably, an older apparent brain in adolescence would be linked to increased memory performance. However, it is not known if the effects of brain age on memory substrates are shared across memory types or unique to each substrate.

The present study

In the present study, I assess the relationship between memory abilities and structural brain aging during childhood and adolescence, as predicted by a deep learning-based brain age model applied to structural neuroimaging material from a large sample of young individuals. Measures from three separate tasks form the basis of working memory and episodic memory abilities. The brain age model predicts the age of the subjects, where the resulting gap between predicted and chronological age, the brain age delta, is used as a proxy for describing brain maturity. We know that frontal lobe function is important for WM and that grey matter increases in this region peaks in early adolescence, and that the medial temporal lobe and the hippocampus change considerably until late adolescence, regions especially important for the storage and retrieval of information necessary for EM (Schneider & Ornstein, 2019). However, there has also been shown a lot of overlap between brain regions activated during memory performance, as well as inter-individual variability in region-specific brain development (Oschwald et al., 2019). In addition, the research on brain age in younger samples is still sparse, especially relating to longitudinal neurocognitive measures. The data material applied in the present study follows the same subjects across two timepoints with measurements of neurocognition and brain structure taken two years apart, giving the opportunity to look at both interindividual- and intraindividual differences in brain- and memory development. The overall aim is to assess to what degree memory abilities are linked to brain structure and development.

Hypotheses

Based on previously established associations between memory- and brain development, I hypothesize that memory performance on both memory tasks will be positively correlated with predicted brain age, indicating a more mature brain in comparison to chronological age. As the prefrontal cortex is especially important for working memory performance and is the last brain area to mature, and both brain age and WM are heritable and

associated with fluid intelligence, I further hypothesize that the strength of the association with the delta will be stronger for working memory than episodic memory. This hypothesis should be considered as exploratory as I am unaware of any literature investigating the relationship between cognition and longitudinal changes in brain age predictions during adolescent development.

Methods and materials

Participants

All participants were originally recruited as part of the Adolescent Brain Cognitive Development (ABCD) study currently ongoing in the US. The review and approval of the ABCD research protocol was handled by a central Institutional Review Board at the University of San Diego, California (Auchter et al., 2018), and the study follows established federal and state regulations regarding biomedical ethics in the U.S. (Clark et al., 2018). Informed consent was given by parents or guardians and assent was given by children before participation. While the initial goal of the ABCD study was to assess substance use, it evolved into a longitudinal cooperative developmental study which tracks children from ages ~10-20 and contains a wide range of data on observable phenotypes (Volkow et al., 2018). A total of 11,877 participants between the ages of 9 and 10.99 years were included at baseline (ABCD 4.0 Data Release, 2022), recruited from 21 data collection sites across the US, representing a wide range of socioeconomic strata. Participants had to be able to complete all baseline measurements, including MRI scanning, to participate. Exclusion criteria from the ABCD study include sensorimotor impairments, persistent major neurological disorders, severely premature birth, birthweight of less than 1200 grams, current diagnosis of psychiatric disorder, and traumatic brain injury. For the present study, participants will have a varying amount of MRI- and neurocognitive data due to exclusions based on quality control and task responses across the two timepoints. An overview of the final sample is displayed in table 1. See data acquisition for further elaborations on participant exclusions.

Table 1

Final Sample Demographic Characteristics at Baseline and Follow-up

Characteristics	Baseline				Follow-up			
	N	%	M	SD	N	%	M	SD
Age (years)			9.92	0.63			11.93	0.64
Sex								
F	5017	44.23			4445	44.56		
M	5465	48.18			4861	48.73		
Ethnicity*								
White	6021	53.08			5473	54.87		
Black	1628	14.35			1300	13.03		
Hispanic	2285	20.14			1981	19.86		
Asian	220	1.94			186	1.86		
Other	1181	10.41			1032	10.35		
PDS			7.86	2.46			10.55	3.69
Birth weight (g)			3183	658			3184	658

Note. N = 11,343 at baseline and N = 9975 at follow-up. M = mean, SD = standard deviation, PDS = pubertal developmental scale. *Based on self-report. Not all participants have complete demographic data.

Data acquisition

Data for the present study was specifically collected from the ABCD release 4.0 (doi: 10.15154/1523041), with the exception of MRI data from the two-year follow up obtained from the ABCD fast-track imaging data release (see https://abcdstudy.org/scientists/data-sharing/fast-track-imaging-data-release/). My access to the data material was granted through Request #7474 (PI: Westlye), and local approval for handling of data is registered as REK 2019/943 (Regional Committees for Medical and Health Research Ethics). All data handling was performed on the TSD (Tjeneste for Sensitive Data) facilities, a secure server environment owned by the University of Oslo.

Structural MRI

Participants were scanned at 21 data collection sites across the US, acquiring structural T1w images from a total of 28 different scanner types from Siemens, General Electric (GE) and Philips (ABCD Data Release 4.0, 2022). Images from Siemens scanners were obtained with 176 slices, TE: 2.88 ms, and acquisition time: 7:12. Philips scanners had the following parameters: 225 slices, TE: 2.9 ms, and acquisition time: 5:38. Finally, images obtained from GE scanners had 208 slices, TE: 2, and acquisition time: 6:09. All scanners shared a

resolution of 1.0 x 1.0 x 1.0. and a TR of 2500 ms for the T1w images. Due to a large amount of missing MRI data from the 2-year follow up, unprocessed images from this timepoint were obtained from the ABCD Fast-track Imaging Data Release. Here, several participants had more than one T1w image, likely due to movement by the participant causing insufficient image quality. Images for each of these participants were matched with - and selected based on timestamps documented in an additional file provided by ABCD regarding subsequent quality control. Participants were excluded from a timepoint if the T1w data did not meet all criteria for inclusion, as recommended by ABCD (ABCD Data Release 4.0, 2022). Further exclusions were made if the data did not pass the initial post processing quality control.

Image preprocessing

All images were preprocessed as described by Leonardsen et al. (2022) to produce inputs equivalent to those used when training the brain age model (fig. 3). Skull-stripping, the removal of non-brain tissue, was accomplished with steps 1-5 from the auto-recon pipeline in Freesurfer version 5.3.0 (Ségonne et al., 2004). Following this, images were converted and reoriented to fit the standard MNI152 orientation from FSL (Jenkinson et al. 2012), and then linearly registered to MNI152 space using the FLIRT tool from FSL (Jenkinson & Smith, 2001). Finally, images were cropped to remove non-brain voxels, and voxel intensities were normalized in the range 0 to 1.

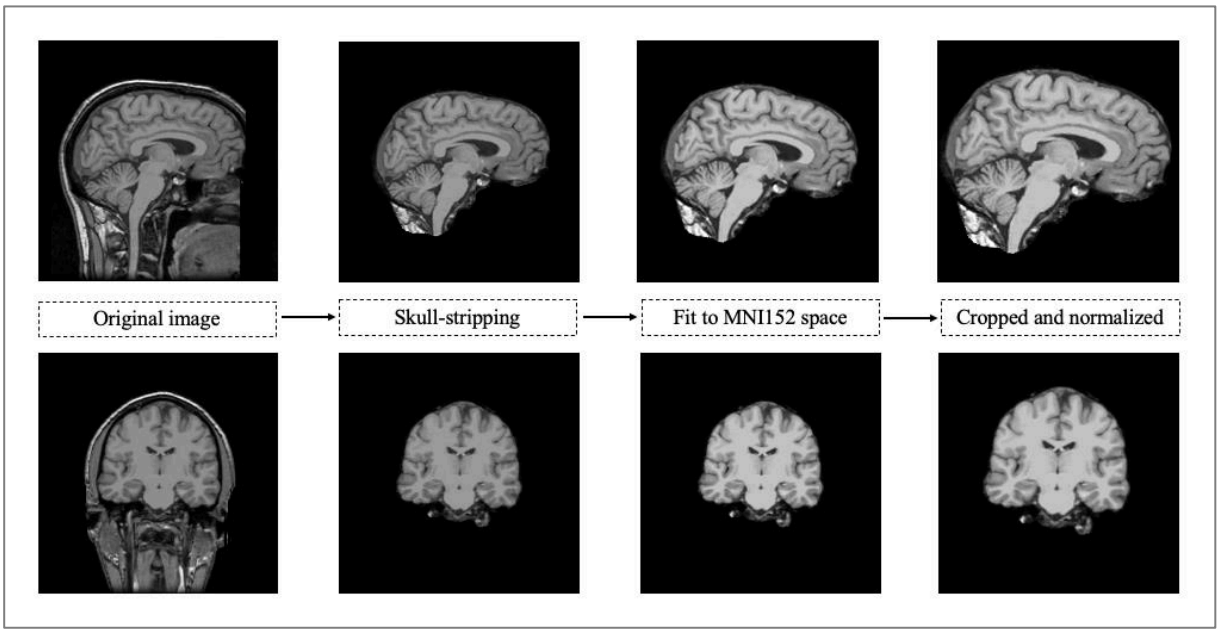

Figure 3. Illustration of the preprocessing steps applied to the T1w structural MRI data.

The brain age model

The SFCN-regression model applied in the present study is a publicly accessible model that was trained, validated, and tested on minimally preprocessed T1w MRI data from

21 different datasets (Leonardsen et al., 2022). A total of 42,829 healthy individuals between ages 3 and 95 years comprised the sample for model building, split into 34,285 (80%) for training and 8544 (20%) for validation. The model architecture is based on the original SFCN model previously described (Peng et al., 2021), with five repeated convolution blocks, each consisting of a convolutional layer with a filter size of 3x3x3, a batch normalization layer, ReLU activation, and a max-pooling layer. Then follows a channel-wise convolutional layer with a filter size of 1x1x1, a batch normalization layer, a bounded ReLU activation, and a global average pooling layer (Leonardsen et al., 2022). Unlike the softmax output in the model by Peng et al. (2021), the SFCN-reg is based on regression, and has a single output node predicting a single continuous value. This output reflects the predicted age for each individual (fig. 4) (Leonardsen et al., 2022).

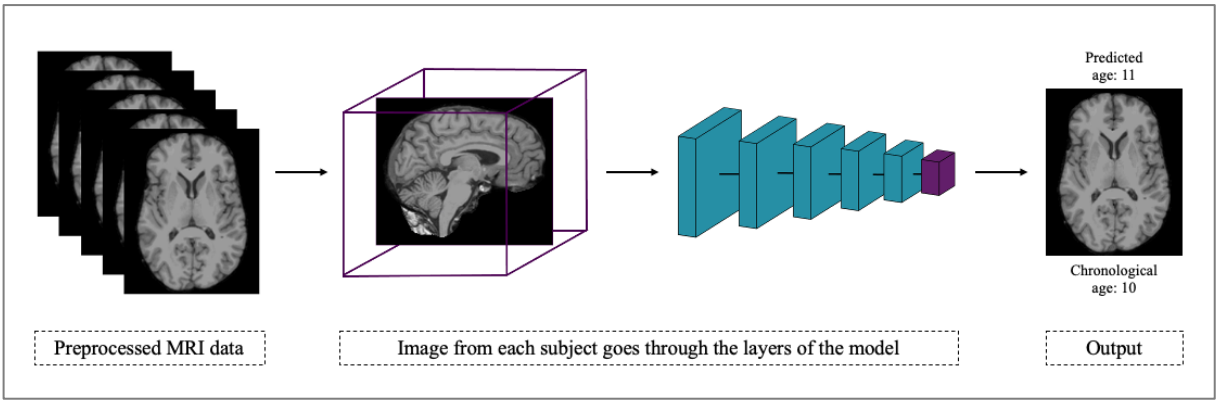

Figure 4. The SFCN-reg model is comprised of a total of six blocks. It is applied to the preprocessed data resulting in a predicted brain age per subject.

The model was applied to the preprocessed data in Python (version 3.9) using the deep learning library TensorFlow (version 2.6) (Abadi et al., 2015). Chronological age was then subtracted from the predicted brain age to calculate the brain age delta for each subject. Due to imaging material being collected from 28 different scanner types, brain age delta values were residualized for scanner-effects (Fortin et al., 2018) using an adapted version of the umx_residualize function from the umx R package, which uses the residuals from a simple linear model with delta as the outcome and scanner ID as a covariate (Bates, Maes & Neale, 2019).

The initial sample size containing brain age prediction data was 11,436 at baseline and 4755 at the two-year follow-up. This marked difference is due to difficulties in data-collection from US COVID-19 restrictions (Saragosa-Harris et al., 2022). In addition to previous exclusions due to QC, brain age delta outliers were detected and removed based on robust z-

scores. It is robust due to the scale being measured as the median of all absolute deviations from the median, resulting in a scale less influenced by extreme outliers (Rousseeuw & Hubert, 2018). Exclusions of participants with a z-score above 3 results in a final sample of n = 10,560 (baseline) and n = 4469 (follow-up), of which 4170 participants have brain age data at both timepoints.

Behavioral measurements

ABCD performs bi-annual collection of neuroimaging material and substantial cognitive testing, employing tasks suitable for participants between age 9 and early adulthood (Luciana et al., 2018). For the present study, three different tasks were selected to assess memory performance.

Episodic memory

The picture sequence memory task is a part of the NIH Toolbox cognition measures (Weintraub et al., 2013). Participants must reproduce a sequence of fifteen thematically related pictures portraying objects and activities, however, the order of the pictures does not follow a specific narrative. They are displayed one at a time, with the content of each picture simultaneously being described by an audio file. After all pictures have been portrayed, they are moved to the center of the screen and the participant attempts to reproduce the sequence previously demonstrated (see fig. 5). The task is administered on an iPad with a completion time of approximately 10 minutes, and the final score is the cumulative number of adjacent pictures remembered across trials. As there is no delayed condition in this task, I computed a composite score attempting to get a more comprehensive measure of episodic memory, with results also including the Rey Auditory Verbal Learning Test (RAVLT). In this task participants both listen to and attempt to recall a list containing 15 unrelated words across a total of five trials (Luciana et al., 2018). Following this they are presented with a separate list of 15 new words as a distraction. There are two retention intervals that follows these trials, one short-term delay after the distractor list, and a long-term delay after 30 minutes. Combined, the short- and long-delay conditions yield a score between 0 and 30. ABCD employed an automated version where responses were checked off on an iPad by the experimenter, and calculations of scores were done by the program. The composite episodic memory score was calculated by adding the means of the two tasks based on their robust z-scores, such that both tasks were weighted equally. Both the picture sequence score and the composite score will be applied in separate models.

Due to the implications of COVID-19 restrictions, remote and hybrid testing was administered at the two-year follow-up, where participants completed the tasks on personal devices (ABCD data release 4.0: Neurocognition, 2022). Research associates monitored the performance using Zoom screen sharing to ensure proficient execution. The variety of personal device could, however, influence the results from the second timepoint.

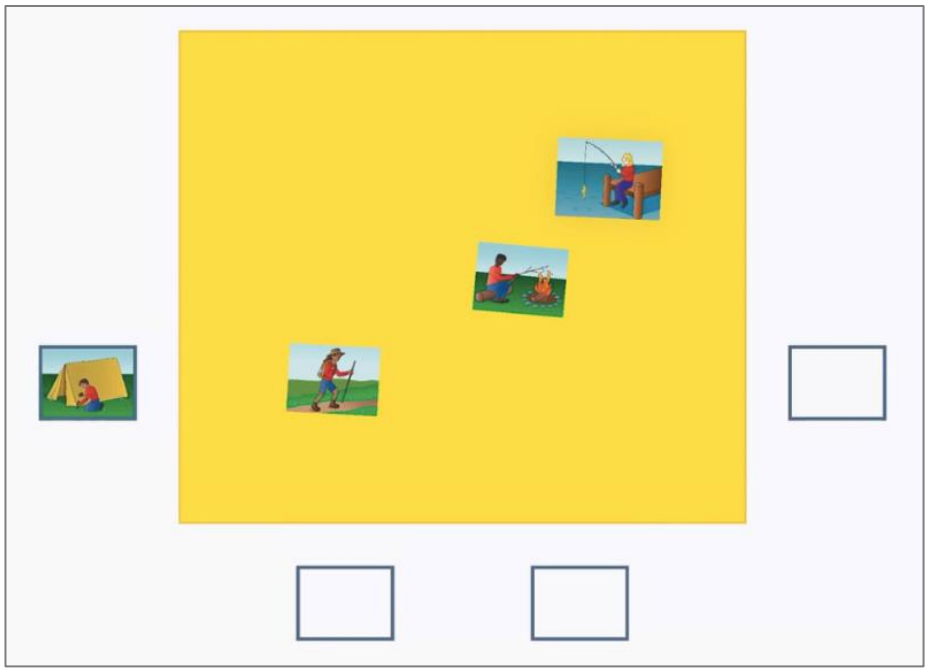

Figure 5. Illustration of the Picture sequence task.
Participants attempt to reproduce a sequence of displayed pictures. The difficulty gradually increases, finally reaching a maximum of 15 pictures. Figure retrieved from © 2022 Toolbox Assessments.

Working memory

The NIH toolbox cognition battery previously described also includes a list sorting task to assess working memory (Tulsky et al., 2014), but unfortunately this task could not be administered in an adequate manner with remote testing at the two-year follow-up. However, the participants completed an emotional n-back task at both timepoints, another valid measure of working memory. The task is administered while participants are inside the MRI scanner, but for the purposes of the present study, only behavioral data is used (Casey et al., 2018). Participants are presented with images of faces or places and must determine whether the current stimulus is a match or not depending on the task condition (fig. 6). There are a total of 160 trials, with 10 trials in each block, and the 16 blocks are completed across two runs. During a low memory load condition (0-back), participants are asked whether the current stimulus matches a target presented at the beginning of the block, and for the high memory load condition (2-back), whether the stimulus matches the one presented two trials back. Participants were excluded if they responded on less than 40 of the 80 trials (50%) of material in each condition. A total score was calculated by subtracting accuracy results for the 0-back

from the 2-back condition to assess the effect of load (Pedersen et al., 2023). To ensure a fair measure that does not yield a better result for low effort on the 0-back compared to the 2-back condition, I made further exclusions if they had a task accuracy below 60% on the 0-back condition. The 2-back score and total n-back score will be applied in separate models.

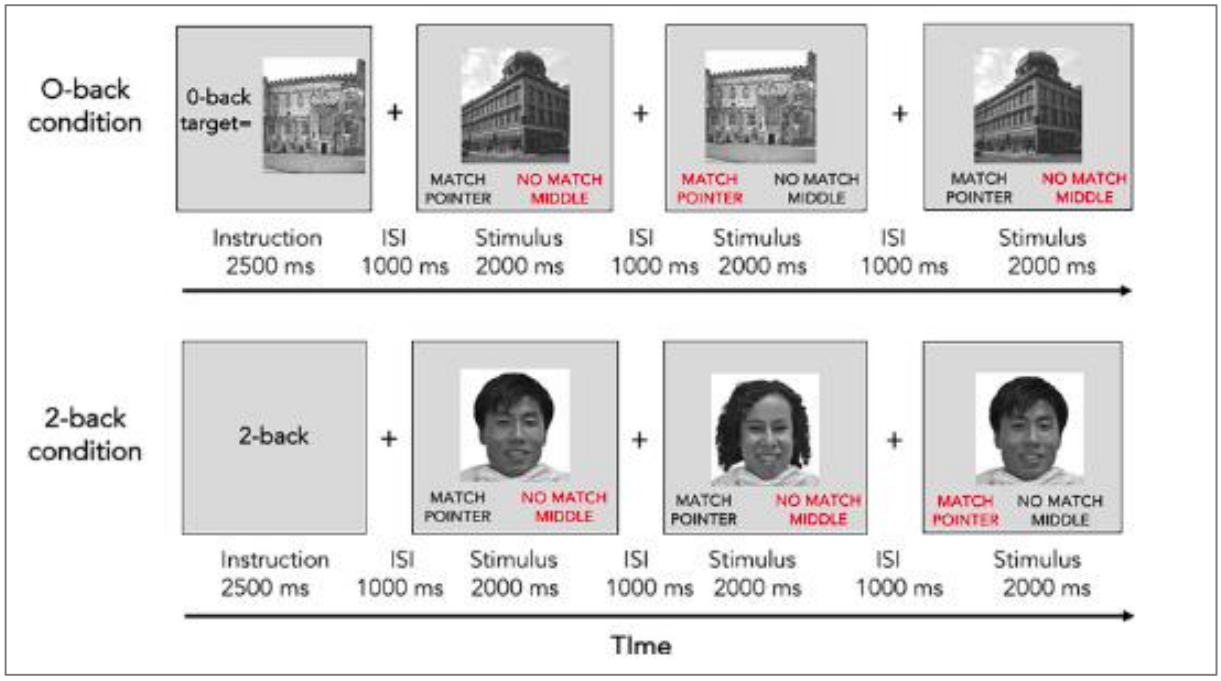

Figure 6. Emotional n-back task.
In the 0-back condition, participants determine whether the image matches a target stimulus. In the 2-back condition participants determine whether the image matches the stimulus presented two trials prior. Figure adapted from Casey et al., 2018.

Statistical analysis

All analyses were performed using R Statistical Software version 4.0.0 (R Core Team, 2020). To assess change over time, linear mixed effect models were computed using the lme4 package (Bates, Mächler, Bolker & Walker, 2015). This approach was selected as mixed effects models yields the possibility of modelling both fixed and random effects, handle missing data across time and predictors, and can model time-varying covariates and account for patterns of correlations across time (Cheng et al., 2010). To create an interpretable intercept, the predictors applied in the linear mixed models (delta, brain age, and all memory measures) were centered around the grand mean (Enders & Tofghi, 2007). For each predictor, this entails a subtraction of the variable mean of the full sample from the variable value of each subject. As a result, the intercept reflects the value of the dependent variable when the independent centered predictors are at their mean of 0 (Hoffman & Walters, 2022).

Covariates were specified a priori based on hypotheses, derived from factors thought to influence structural brain development and cognitive abilities, or as recommended by the ABCD study. As previously mentioned, there are certain differences between males and females regarding structural brain maturation (Mills et al., 2021; Fuhrman et al., 2022), which is also influenced by pubertal maturation (Holm et al., 2023). Age and puberty appear to have interactive as well as independent effects on brain development relevant to cognitive improvements (Goddings et al., 2014). The Pubertal developmental scale (PDS) is an assessment on perceived physical pubertal markers, where results from primary caregiver(s) reports are applied as a covariate in the present study (Herting et al., 2021). In addition, birth weight is included as it has shown to be significantly associated with cortical morphology across the lifespan (Walhovd et al., 2016), and has been shown to interact with brain age estimates (Vidal-Pineiro et al., 2021). Finally, self-reported ethnicity is included, where participants are grouped in one of the following categories: White, Black, Hispanic, Asian, or Other (Saragosa-Harris et al., 2022). These categories are broad, potentially reflecting biological, cultural, as well as socioeconomic influences.

Linear mixed effects analyses

To test the first hypothesis anticipating positive associations between memory performance and predicted brain age, four models were computed. To evaluate potential differences between memory measures, each model contains a different memory score as the dependent variable, and brain age as the key independent variable. As brain age is expected to increase with time, an interaction effect between time and brain age was added. I also expected individual differences to be present, thus including a random intercept for each subject. All models include the covariates age, sex, pubertal development, ethnicity, and birth weight. The first two models assess WM, with 2-back accuracy as the dependent variable in model 1, and total n-back score in model 2. Models 3 and 4 assess EM, with Picture sequence task scores as the dependent variable in model 3, and the composite EM score in model 4. Apart from the dependent variable, all other covariates are kept consistent in the models

$$Y_{ij} = \beta_{0j} + brainage + time + brainage * time + age + PDS + sex + ethnicity + birthweight,$$

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

where Y_{ij} is the memory performance for subject *j* at timepoint *i*, and β_{0j} denotes the random intercept for subject *j*, where γ_{00} is the intercept and U_{0j} is the random effect on the intercept. Note that the β is removed from the fixed effects in the formulas to simplify interpretation.

As formulated by the second hypothesis, a stronger association between delta and WM was expected. This was tested by a single model with delta as the dependent variable, and WM and EM performance as independent variables. Here, the n-back score and composite EM score are used as proxies for WM and EM respectively. Interaction effects between memory performances and time was added to the model

$$Y_{ij} = \beta_{0j} + workingMem + episodicMem + time + workingMem * time + episodicMem * time + age + PDS + sex + ethnicity + birthweight$$

$$\beta_{0j} = \gamma_{00} + U_{0j}$$

where Y_{ij} is the delta for subject *j* at timepoint *i*, and β_{0j} denotes the random intercept for subject *j*.

Note that p-values are included for the predictors in all models and was calculated based on Satterthwaite´s method for approximating degrees of freedom by the lmerTest package in R (Kuznetsova, Brockhoff & Christensen, 2017). As the degrees of freedom is an estimation, resulting p-values should only be considered as an approximation.

Results

All plots were made with the ggplot2 package in R (Wickham, 2016) unless otherwise stated.

Brain age prediction accuracy

Predicted brain age ranged between 7.2-13.6 years at baseline (mean = 9.88, SD = 0.85) and 7.8-16.3 years at follow up (mean = 11.18, SD = 1.50), with a correlation between chronological age and predicted brain age of $r = 0.31$ and $r = 0.37$ respectively. Fig. 7 displays the distribution of chronological and predicted age at both timepoints.

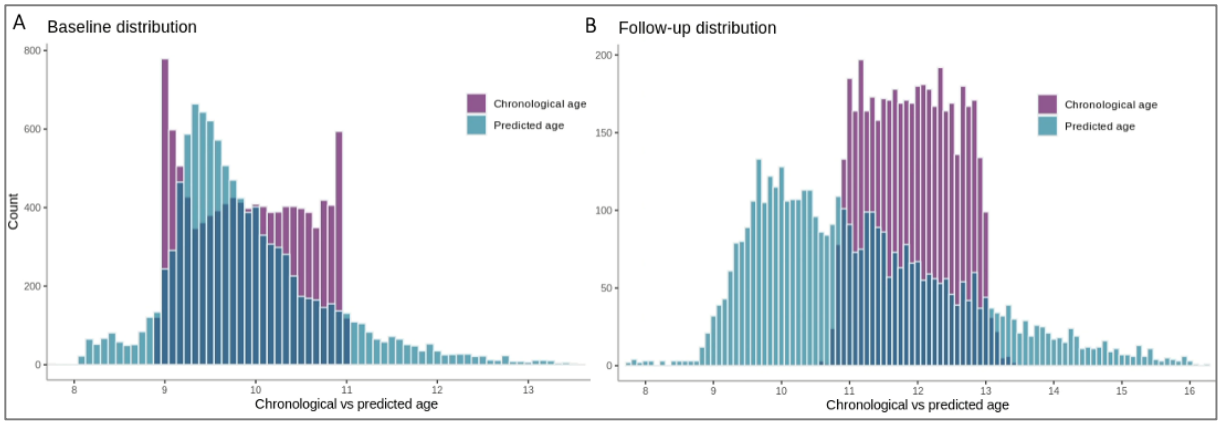

Figure 7. Double histograms of chronological age and predicted brain age**.**

Distribution at (A) baseline (N = 10,560) and (B) follow up (N = 4469). Purple illustrates chronological age, and blue illustrates predicted brain age.

Delta, the subtraction of chronological age from predicted brain age, showed a range between -2.66-2.65 at baseline (SD = 0.88) and -4.87-3.35 (SD = 1.40) at follow up. The model predicted with an accuracy of MAE = 0.70 at baseline and MAE = 1.34 at follow up. The correlation between delta at baseline and follow up was $r = 0.68$ (fig. 8).

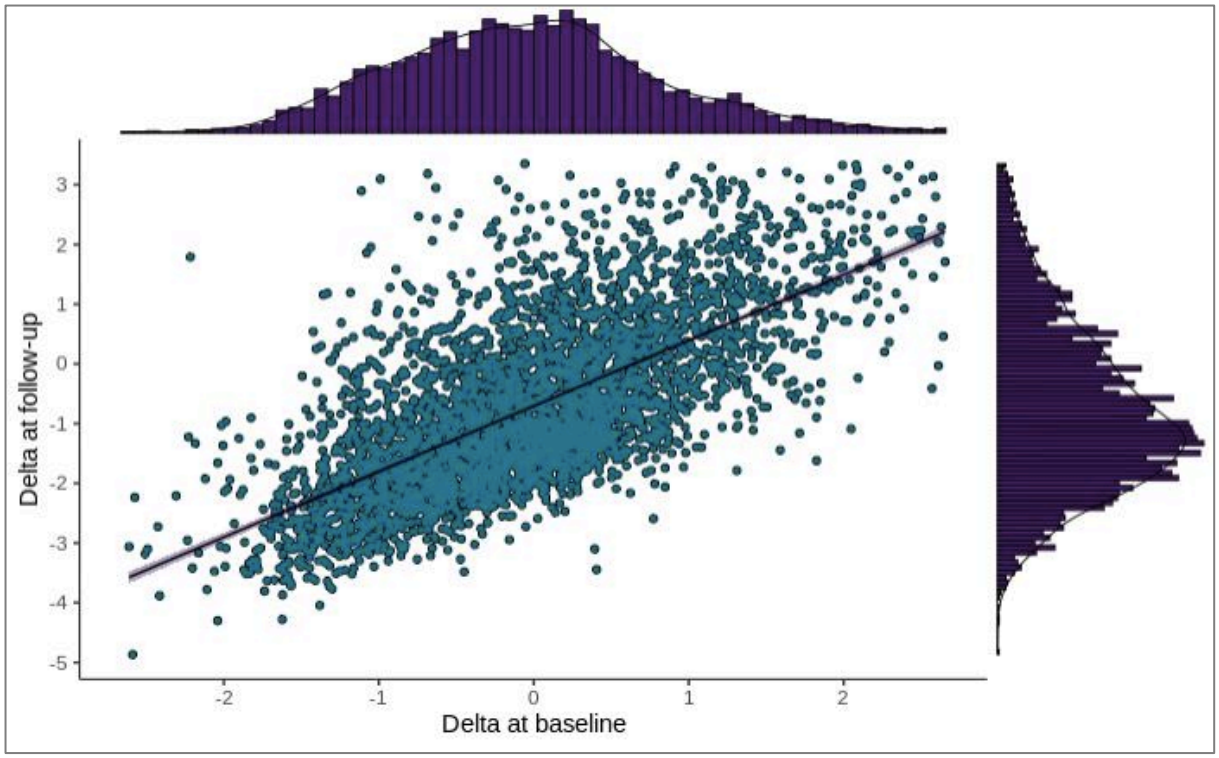

Figure 8. Delta at both timepoints.

Scatterplot of delta at baseline and follow up, and marginal histograms displaying the distribution of delta at baseline (top) and follow up (right). N = 4170.

Memory performance

There was an overall increase in memory performance between baseline and follow up (see Table 2 for complete overview). WM performance assessed by the total n-back scores (2-back accuracy – 0-back accuracy) had a mean = -6.3 at baseline and mean = 6.45 two years later. Looking at the 2-back scores in particular, the results also show an increase in performance with a mean = 62 at baseline and mean = 68.3 at follow up. EM results from the picture sequence task had a mean = 102.8 at baseline and a mean = 108.6 at follow up, while the increase was more modest when looking at the composite EM score across time, with a mean = 98.7 and mean = 101.8 respectively. As the composite score is based on both the picture sequence task and RAVLT, this indicates less variation in performance on the RAVLT across the two timepoints.

Table 2

Overview of Memory Tasks and Scores

	Baseline			Follow-up		
	Range	M	SD	Range	M	SD
Working memory						
2-back scores	5-79	62.0	8.76	2-80	68.3	7.72
Total n-back scores	-59.0-28.0	-6.3	8.23	-78.0-23.0	-4.2	6.45
Episodic memory						
Picture sequence task scores	76-136	102.8	12.07	76-136	108.6	12.58
Composite scores	62.2-131.3	98.7	11.9	62.1-131.3	101.8	11.6

Note. M = mean, SD = standard deviation.

For visual comparisons of memory performance across tasks and timepoints (fig. 9), robust z-scores were applied to all memory scores (Rousseeuw & Hubert, 2018).

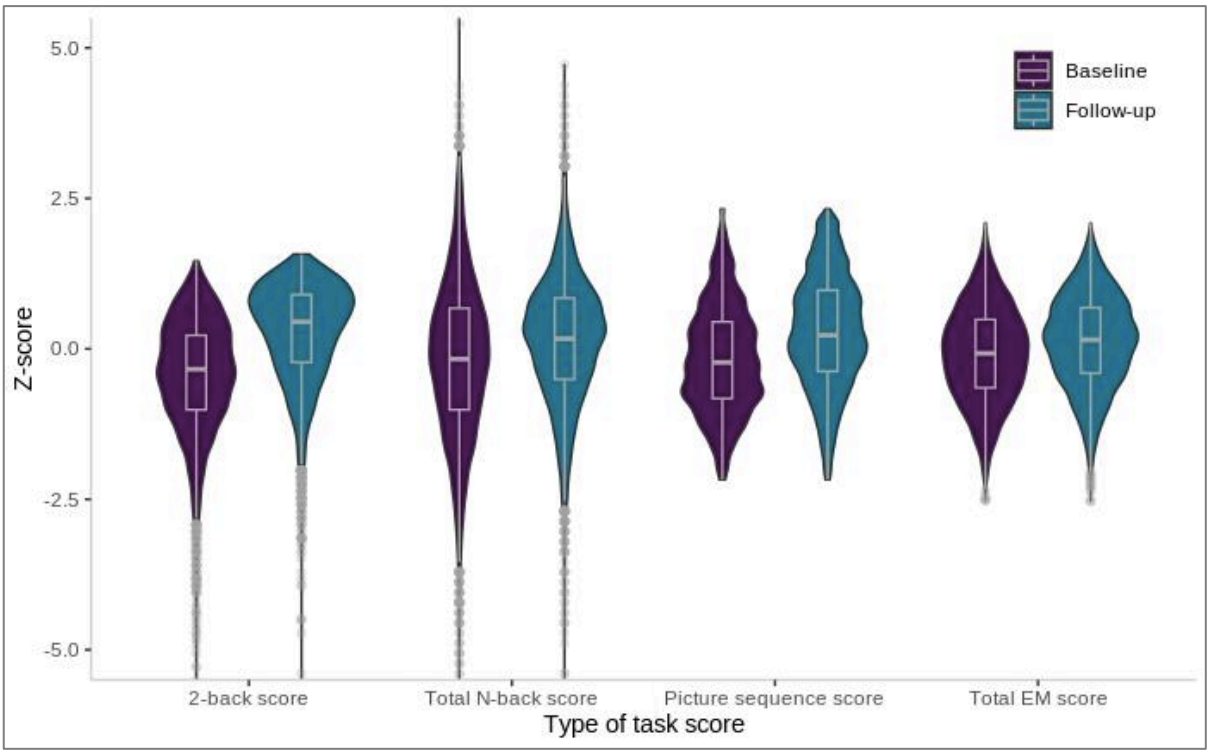

Figure 9. Violin plots of memory performance.

Memory measures are displayed along the x-axis, and robust z-scores along the y-axis. Boxplots display the interquartile range (1-3) with the line displaying the median. Dots represent scores identified as outliers.

Main results

Effect of brain age on memory performance

The statistical results of main effects from models 1-4 assessing parameter effects on memory performance are summarized in table 3 (see appendix for an overview of all fixed effects in each model). There was an apparent positive linear relationship between brain age and memory performance when all covariates were removed, as displayed in figure 10. However, looking at model outcomes there was no significant effect of brain age on WM in either model ($p = 0.812$ in model 1; $p = 0.628$ in model 2), and no significant interaction effect between brain age and time ($p = 0.450$ in model 1; $p = 0.379$ in model 2) on WM. Collinearity between predictors, in this instance between age and time, can in some cases result in sign flipping, as seen with brain age in table 3, model 1 (Elman et al., 2022). The random intercept of subjects explained approximately half of the variance in 2-back memory performance (model 1) across timepoints (subject = 30.16, residual = 31.26), and approximately 25% of the variance in n-back performance (model 2) across timepoints (subject = 13.20, residual = 40.89).

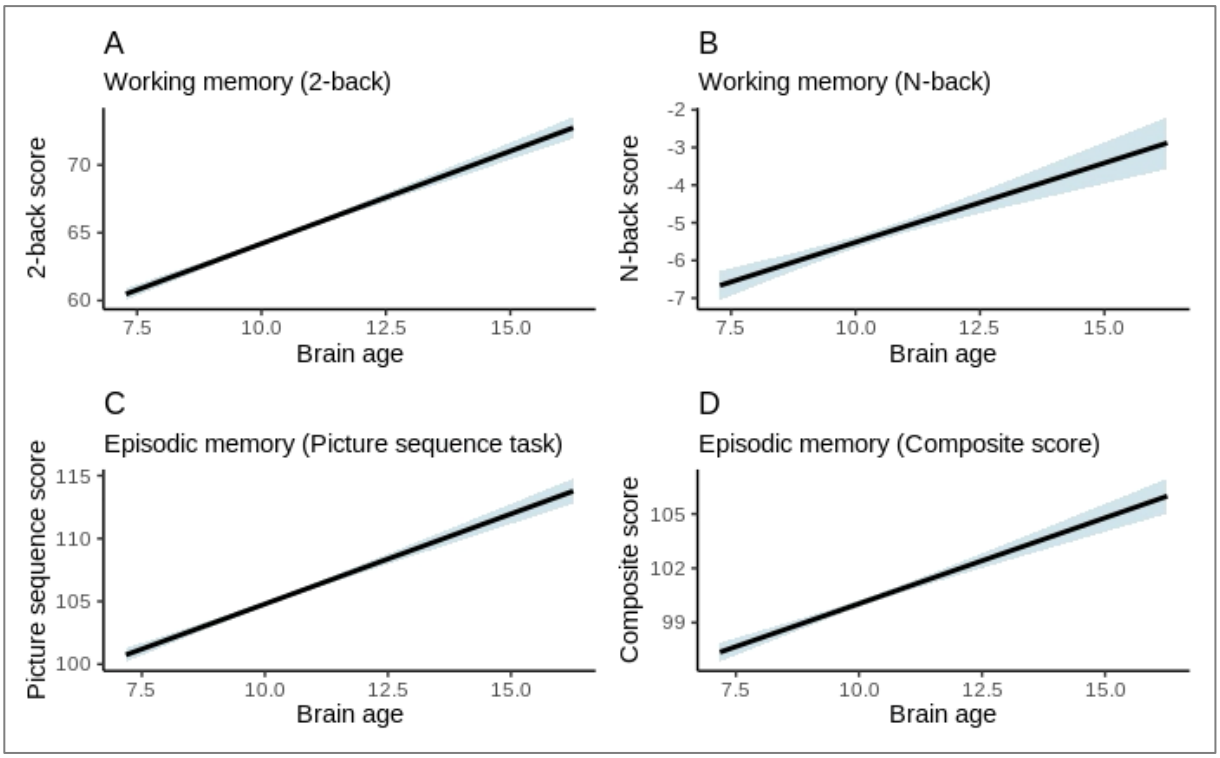

Figure 10. Displaying positive linear relationships between memory and brain age as measured by (A) 2-back performance, (B) overall N-back performance, (C) picture sequence task performance, and (D) composite EM performance. Confidence interval = 95%. Note the difference in intercept.

Models 3 and 4 assessed parameter effects on EM performance. No significant main effect of brain age was found in either model (model 3: $p = 0.099$; model 4: $p = 0.134$). A significant negative interaction effect of brain age and time was seen in both models (model 3: $p = 0.035$; model 4: $p = 0.003$). The random intercept of subjects explained approximately 43% of the variance in picture sequence task memory performance (model 3) across timepoints (subject = 57.91, residual = 31.26), and approximately 60% of the variance in the composite score performance (model 4) across timepoints (subject = 75.76, residual = 47.86).

Table 3

Results from models 1-4 assessing memory and brain age

Working memory	Model 1: 2-back score				Model 2: N-back score			
Fixed effects	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	3.749	1.585	23.647	<0.001	-5.970	4.812	-12.407	<0.001
Timepoint	1.810	3.093	5.851	<0.001	1.465	2.907	5.040	<0.001
Brain age	-2.927	1.237	-0.237	0.812	5.765	1.189	0.485	0.628
Time:Brain age	-9.969	1.319	-0.756	0.450	-1.186	1.349	-0.880	0.379
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	30.16	5.49			13.20	3.63		
Residual	31.26	5.59			40.89	6.40		
Episodic memory	Model 3: Picture sequence task score				Model 4: Composite score			
Fixed effects	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	1.040	6.364	163.377	<0.001	9.983	6.239	160.023	<0.001
Timepoint	2.449	3.845	6.369	<0.001	-7.175	3.681	-1.949	0.051
Brain age	2.309	1.398	1.652	0.099	1.986	1.326	1.497	0.134
Time:Brain age	-3.297	1.563	-2.110	0.035	-3.993	1.357	-2.941	0.003
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	57.91	7.61			75.76	8.70		
Residual	76.43	8.74			47.86	6.92		

Note. *Approximated p-values.

To summarize, there was an overall positive relationship between brain age and memory performance, though this effect decreased with time. However, this effect was only significant for the relationship between brain age across time and EM performance. As for the covariates, age, sex, and birthweight showed a significant effect on WM performance, and all covariates showed a significant effect on EM performance (see appendix).

Effect of memory performance on delta

Similar trends as described above applies to model 5 assessing the relationship between memory and delta, of which main results are summarized in table 4 (see appendix for all covariates). The total n-back score and the composite EM score were applied in the model to represent WM and EM. Both WM and EM show non-significant main effects on delta ($p = 0.723$ and $p = 0.270$ respectively), as well as non-significant interaction effects with time on delta ($p = 0.419$ for WM; $p = 0.784$ for EM). Allowing subjects to have a random intercept explains 58% of the variance seen in delta (subject = 0.66, residual = 0.48).

Table 4

Results from model 5 assessing delta and memory

	Model 5: Delta			
Fixed effects	**Coef. (B)**	**SE (B)**	**t-value**	**p-value***
Intercept	4.134	2.116	19.532	<0.001
Timepoint	8.170	4.302	1.899	0.058
WM	5.886	1.662	0.354	0.723
EM	-1.332	1.208	-1.102	0.270
Time:WM	-2.408	2.980	-0.808	0.419
Time:EM	4.754	1.736	0.274	0.784
Random intercept	**Variance**	**SD**		
Subject	0.66	0.81		
Residual	0.48	0.70		

Note. WM = Working memory. EM = Episodic memory. *Approximated p-values.

After closer inspection, delta shows an inverse relationship to both age and memory performance, as illustrated in figure 11 (A, D, E). This entails that an increase in age reduces the average delta, as does an increase in memory performance. Age by itself, however, has a positive relationship with memory performance. Thus, while memory performance increase, delta shows a linear decrease (figure 11D and E).

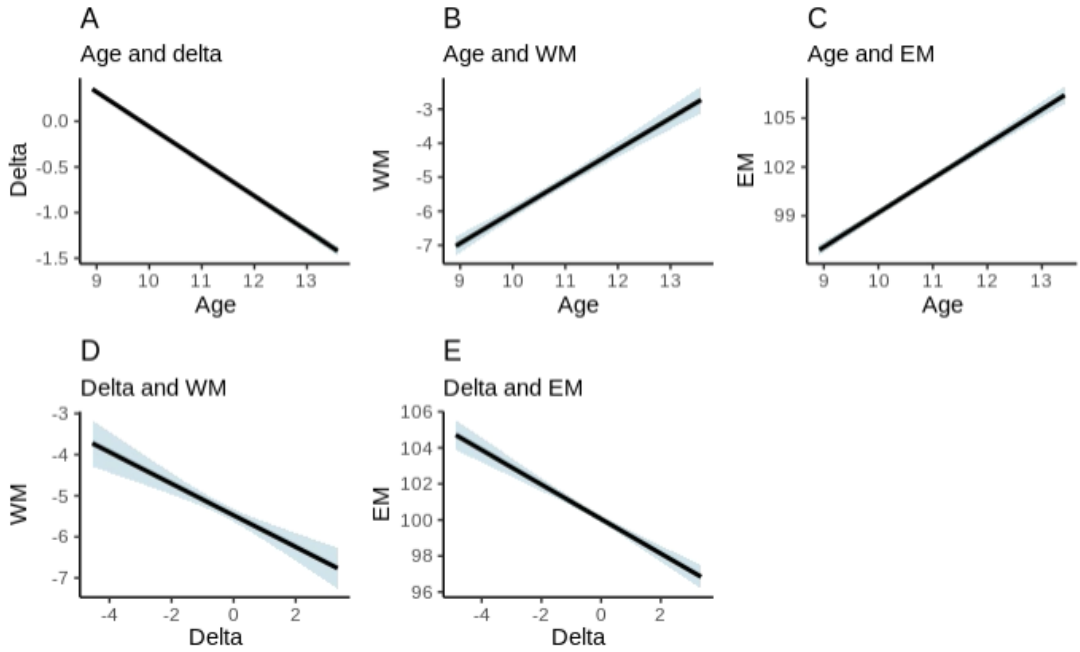

Figure 11. Displaying the linear relationship between (A) age and delta, (B) age and WM, (C) age and EM, (D) delta and WM and (E) delta and EM. Although memory performance increase with age, the mean decrease in delta with increasing age results in the negative relationship between delta and memory. Confidence interval = 95%.

Results of post-hoc analyses

Observing that there was no significant association, post-hoc analyses were performed with a selected set of subjects having the most deviating brain age deltas, to investigate whether brain-memory associations were significant in those whose brain appeared very mature or young in comparison to chronological age. Based on the data from the two-year follow-up, subjects were sorted from highest to lowest in terms of delta-values, and the top 10% and lowest 10% were selected, amounting to a total of 866 subjects. All parameters were kept consistent with the models described above, so that models 6 and 7 assess WM-brain age associations, models 8 and 9 assess EM-brain age associations, and model 10 assess delta and memory associations. In this group we say stronger evidence of an effect of brain age on memory performance, however, only significant on WM. The interaction effect of brain age and time was significant for all models.

Table 5

Results from models 6-9 with deviating delta participants, assessing memory and brain age

Working memory	Model 6: 2-back score				Model 7: N-back score			
Fixed effects	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	4.258	4.625	9.265	<0.001	-4.246	1.184	-3.587	<0.001
Timepoint	1.698	9.322	1.822	0.069	1.882	8.211	-3.587	0.022
Brain age	5.532	2.886	1.917	0.055	7.835	2.682	2.992	0.004
Time:Brain age	-5.834	2.776	-2.101	0.036	-8.567	-8.567	-3.163	0.002
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	25.47	5.047			11.33	3.367		
Residual	32.09	5.665			34.43	5.868		
Episodic memory	Model 8: Picture sequence task score				Model 9: Composite score			
Fixed effects	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	1.032	1.856	55.624	<0.001	9.950	1.790	55.589	<0.001
Timepoint	2.900	1.254	2.313	0.021	5.572	1.202	0.464	0.643
Brain age	5.383	3.385	1.590	0.112	4.610	3.057	1.508	0.132
Time:Brain age	-6.360	3.171	-2.006	0.045	-5.767	2.653	-2.174	0.030
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	56.55	7.520			68.71	8.289		
Residual	74.86	8.652			45.90	6.775		

Note. *Approximated p-values.

There were no significant effects of memory on delta in this group, and no significant interaction effects, as displayed in table 6.

Table 6

Results from model 10 with deviating delta participants, assessing delta and memory

	Model 10: Delta			
Fixed effects	**Coef. (B)**	**SE (B)**	**t-value**	**p-value***
Intercept	4.134	2.116	19.532	<0.001
Timepoint	8.170	4.302	1.899	0.058
WM	5.886	1.662	0.354	0.723
EM	-1.332	1.208	-1.102	0.270
Time:WM	-2.408	2.980	-0.808	0.419
Time:EM	4.754	1.736	0.274	0.784
Random intercept	**Variance**	**SD**		
Subject	0.66	0.81		
Residual	0.48	0.70		

Note. WM = Working memory. EM = Episodic memory. *Approximated p-values.

Overall, our analyses indicate that associations between memory and brain age primarily are non-significant in this sample. The potential effects of brain age on memory are only observable in participants estimated as substantially more or less mature.

Discussion

The overall aim of this study was to explore the links between memory- and brain development from childhood to early adolescence. To the best of my knowledge, this is the first study to specifically investigate different memory types in conjunction with brain age during this key transitional period. Overall increases in both WM and EM performance across the two timepoints was seen, as well as an expected increase in predicted brain age over time. Based on previous knowledge I hypothesized there would be a positive relationship between memory- and brain development, but observed that all association between memory and brain age in the main analyses were non-significant. I did find a significant negative interaction effect of brain age and time on EM, potentially reflecting lower impact of brain age on EM at follow up. However, I interpret this effect cautiously due to the opposite sign of the non-significant main effect of brain age. Furthermore, it was hypothesized that the association between brain age delta and memory would be stronger for WM, but no significant associations between brain age delta and either memory type was found. Finally, my post-hoc analyses of participants with the highest and lowest predicted brain age in comparison to chronological age revealed significant effects of brain age on WM, as well as significant negative interaction effects of brain age and time on all memory measures in the extreme

groups. Surprisingly, the associations between memory and delta were non-significant also in these.

Validity and reliability of neurocognitive measures

The composite EM score applied in the analyses was calculated to capture a wider set of EM-components, including the storage and retrieval of information after a delayed period. The picture sequence task only measures immediate recall, so the delayed conditions of the RAVLT adds measures of storage and retrieval. Contrary to our expectations, the delayed task scores from the RAVLT showed no variance across time. This is, however, in line with findings from other studies conducted with the ABCD data, that additionally show inverse learning effects for this task (Anokhin et al., 2022; Smith et al., 2023). RAVLT is widely considered to be a valid and reliable measurement of EM, though differences in test administration and instructions have been mentioned as possible influences on non-systematic variations in test-performance (Magalhães, Malloy-Diniz & Hamdan, 2012). The lack of overall improvement as measured by the RAVLT in the ABCD sample could thus be a result of differences in task measurements across time, possibly due to changes in data collection procedures resulting from US COVID-19 restrictions. As a result, the composite EM score displays little variance across time, complicating subsequent analyses of memory-brain associations.

Covid-19 restrictions also influenced collection of WM data, as the original list sorting task to assess WM could not be performed at the two-year follow up. Because of previously mentioned associations between the list sorting task and the N-back task (Rosenberg et al., 2020), the latter was considered as the best available proxy for WM abilities in the present study. However, the correlation between the 2-back condition and the list sorting task was only $r = 0.32$ at baseline. Similar results were displayed in the study by Rosenberg et al. (2020). Here, the correlation between the two WM tasks were $r = 0.36$, and moreover, children with superior WM performance at baseline also achieved higher results in other neurocognitive tasks, specifically those measuring fluid intelligence, language skills and other aspects of memory. As WM is a complex system comprising processes such as maintenance, attention, manipulation, and inhibition of information, a specific task measurement might feature some of these processes more than others, and potentially be influences by other higher-cognition abilities (Lugtmeijer et al., 2019). Hence, N-back might capture certain WM features, such as attention an updating of information, while information manipulation is not assessed to the same degree. Consequently, there is a chance that N-back performance reflect

task-specific rather than WM-specific behavior. Regardless of this, the task measured and captured change over time, displaying an increase in task-performance with development, despite non-significant associations with brain age.

Interpretability of brain age predictions

Brain age predictions in the present study were on average accurate, with MAEs of 0.70 and 1.34 at baseline and follow up, but the within-sample correlation with age was relatively low, namely $r = 0.31$ and 0.37 respectively. I believe this could be indicative that the prediction procured by the specific model we used could lack the specificity needed to disentangle individual differences in the subsequent analyses. The model was trained using a training set with ages ranging from 3 to 95 years, whereas I apply it to a dataset with a span of only two years. Consequently, the model could have learned to recognize broader signs of aging, and not have the expressive power to discern participants which are months apart. This suspicion is strengthened by the high correlation of delta values within subjects ($r = 0.68$) which indicate a level of stability. Continuing, it is worth mentioning that a general challenge of deep learning is the lack of insight into what underlies the model predictions, reflecting the established trade-off between predictive performance and interpretability (Tejavibulya et al., 2022). While we know that the average brain age delta, representing prediction error, is low, we are not able to determine what the inter-individual variation represents. Ideally, the delta should reflect biological variation, but imprecisions due to biases and noise in the data, such as head movement, is also probable sources of variance. And importantly, the effect of these potential sources of noise are magnified since we are trying to discern differences of aging in a narrow age range.

Stability of delta from adolescence to later life

The relatively high correlation between delta at baseline and follow up confirms previous findings that indicate stability in individual brain age relative to chronological aging (Vidal-Pineiro et al., 2021). This is likely influenced by perinatal, early life- and genetic factors on individual brain development and health (Walhovd et al., 2016). Despite this, the lack of longitudinal neuroimaging studies across the whole lifespan makes it difficult to determine how estimates of brain age delta during childhood and adolescent development translates to later life. A higher delta indicates a more mature brain compared to same-age peers, and while this is considered as positive during adolescence, the reverse is the case as one reach adulthood and old age, complicating the interpretation of brain age in this critical period. In a longitudinal study by Shaw et al. (2006) findings suggested a negative correlation

between cortical thickness and intelligence in early childhood, followed by a developmental shift where the correlation turns positive from late childhood and onwards. As delta serves as a proxy for brain health, one could assume that a developmental shift would be reflected as a positive correlation between delta and cognitive abilities in childhood and adolescence, that is reversed after reaching adulthood when the brain is fully developed. Assuming this is the case, delta as a measure would be difficult to interpret, warranting caution when conducting longitudinal measurements during this stage. Another study suggested a link between poor cognitive functioning in childhood and poor brain health at age 45 (Elliot et al., 2021). Participants with a higher delta as adults, reflecting poorer brain health, also displayed lower cognitive ability both in childhood and adulthood. Naturally, no childhood neuroimaging data exists for the same participants due to the age of the sample, but these results do indicate that later stability in brain health retroactively correlates with earlier cognitive functioning. My results indicate a stability of the delta in the years covered in the dataset and could help disentangling the role of brain aging during adolescence, but longitudinal studies following subjects from childhood to adulthood are imperative to assess the developmental pattern of delta throughout life.

Capturing the co-developing changes in memory and the brain

In addition to stability in brain age, individuals also show high stability in general cognitive ability through life when compared to same-age peers, with cortical developmental trajectories following the same pattern (Walhovd et al., 2016). This implies that someone with exceeding cognitive abilities and a healthy brain in childhood likely will display the same qualities in adulthood compared to the average individual. It is probable that a similarly stable relationship exists between the brain and memory specifically, although difficult to capture in the present study for several reasons. As data exists from only two timepoints, potential inconsistencies in data acquisition, model prediction and subjective task performance will likely have a greater impact, making it difficult to capture individual differences over time. To assess whether the potential association between brain age and memory could be captured in more extreme cases, I performed post-hoc analyses (models 6-10, table 5 & 6) including only participants with the largest deviations between predicted and chronological age measured at follow up. This resulted in significant associations between brain age and WM, implying that individuals whose brain appears considerably more mature than expected by their chronological age tend to show increased WM abilities. Furthermore, as an interaction effect with a negative coefficient between brain age and time was observed on all memory

measures, which might suggest a lower impact of brain age on memory as time passes, also in these more extreme cases.

Non-significant associations between brain age and memory could presumably also result from collinearity between predictors (Elman et al., 2022). Despite recent studies suggesting a relation between interindividual variability in brain structure and behavior, the effect of individual brain structure on cognitive processes has proven more difficult to observe as sample sizes increase (Genon, Eickhoff & Kharabian, 2022). With larger and more diverse samples, mutual dependencies between brain and behavioral measures such as memory could be influenced by covariate factors such as age, pubertal development, and ethnicity. Post-hoc analyses revealed correlations between brain age and PDS of $r = 0.33$, and age and PDS of $r = 0.42$. While correcting for these covariates is the recommended practice from the ABCD study, effect sizes of other associations have shown to diminish in the ABCD sample when they are added (Owens et al., 2021), and several distinctive measures of development could disguise some of the effects specific to brain age on memory.

The choice of dependent variables and main predictors

Most studies utilizing brain age models to assess linear relationships with other phenotypes use the predicted brain age rather than the delta as the dependent variable or main predictor. I followed this approach to assess hypothesis 1, where brain age was a predictor for memory measures. As age is included as a covariate, I assumed the combined inclusion of brain age and age would yield an effect of the discrepancy between the two on the memory measures. When assessing hypothesis 2, I was specifically interested in the strength of the association between memory types and delta, thus placing delta as the dependent variable while controlling for age. Although neither approach can be considered as superior based on existing literature, these subjective choices might influence the results and should be taken into consideration.

Limitations and future directions

The ABCD data contains a large sample with several measurements which should yield sufficient power to detect associations. However, in addition to the complications outlined above, the differences in sample size across time and varying amount of specific data measurements across participants could diminish the power and lead to null findings. The discrepancies between measurements were in part due to COVID-19 restrictions, in particular regarding availability of MRI data at follow up. Furthermore, even though the applied brain age model was trained on a large and diverse dataset, most of the subjects were of white

background, potentially reducing generalizability to the more demographically diverse sample in the ABCD study (Leonardsen et al., 2022). This, in combination with the discrepancies in age ranges described above, could lead to reduction in the signal to noise-ratio, affecting the subsequent analyses.

It has been observed that brain changes are underestimated in brain charts from cross-sectional data (Di Biase et al., 2023), highlighting the importance of longitudinal assessment of brain measurement to properly assess brain-memory changes. As more data is released by the ABCD study there will be increased opportunities to study memory-brain relationships in this sample. By following the same individuals from childhood to late adolescence with repeated measurements of memory performance and brain structure, co-developing changes can be properly assessed. Furthermore, continuing technological advancements will likely yield deep learning models with reduced influences from noise and bias, yielding even more precise predictions. Finally, utilizing nonlinear statistical models might capture significant associations, as developmental trajectories are not necessarily linear.

Conclusion

To the best of my knowledge, this study is the first to assess the development of distinct memory types in conjunction with predicted brain age during adolescence. The state-of-the-art SFCN-reg model yielded accurate predictions, but results were insufficient in establishing a clear relationship between memory and brain age. Nonetheless, overall increases were observed in memory performance and predicted brain age across time, and post-hoc analyses suggest an association between these measures in individuals with more extreme brain age deviations. As the present study was limited by the narrow age range and insufficient amounts of measurements across time, future longitudinal studies are necessary to properly assess how brain morphology and memory abilities co-develop during the critical developmental period of adolescence.

Acknowledgements: The ABCD consortium

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development[SM] (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from NIMH Data Archive Digital Object Identifier 10.15154/1523041. DOIs can be found at http://dx.doi.org/10.15154/1523041.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C. …, Zheng, X. (2015). Tensorflow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

ABCD 4.0 Data Release. (2022). The National Institute of Mental Health Data Archive. http://dx.doi.org/10.15154/1523041

ABCD Study®. (n.d.) ABCD Fast-track Imaging Data Release. Retrieved April 24, 2022 from https://abcdstudy.org/scientists/data-sharing/fast-track-imaging-data-release/

Anokhin, A. P., Luciana, M., Banich, M., Barch, D., Bjork, J. M., Gonzalez, M. R., Gonzalez, R., … Wesley Thompson, W. (2022). Age-related changes and longitudinal stability of individual differences in ABCD Neurocognition measures. *Developmental Cognitive Neuroscience, 54*(101078). https://doi.org/10.1016/j.dcn.2022.101078

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R. & Sharma, S. (2013). Maturation of the adolescent brain. *Neuropsychiatric Disease and Treatment, 9*, 449-469. doi: 10.2147/NDT.S39776.

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation, 2*, 89-195. https://doi.org/10.1016/S0079-7421(08)60422-3

Auchter, A. M., Mejia, M. H., Heyser, C. J., Shilling, P. D., Jernigan, T. L., Brown, S. A., … Dowling, G. J. (2018). A description of the ABCD organizational structure and communication framework. *Developmental Cognitive Neuroscience, 32*, 8-15. https://doi.org/10.1016/j.dcn.2018.04.003

Baddeley, A. D. & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8,* 47-89. https://doi.org/10.1016/S0079-7421(08)60452-1

Balaconi, M. (2013). Dorsolateral prefrontal cortex, working memory and episodic memory processes: Insight through transcranial magnetic stimulation techniques. *Neuroscience Bulletin, 29*(3), 381-389. doi: 10.1007/s12264-013-1309-z

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, T. C., Maes, H. & Neale, M. C. (2019). umx: Twin and path-based structural equation modeling in R. *Twin Research and Human Genetics , 22*(1), 27-41. doi: https://doi.org/10.1017/thg.2019.2

Bejani, M. M. & Ghatee, M. (2021). A systematic review on overfitting control in shallow and deep neural networks. *Artificiall Intelligence Review, 54*, 6391–6438. https://doi.org/10.1007/s10462-021-09975-1

Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., … Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. *Nature, 604*, 525-533. https://doi.org/10.1038/s41586-022-04554-y

Bouyeure, A. & Noulhiane, M. (2020). Memory: Normative development of memory systems. *Handbook of Clinical Neurology, 173*, 201-213. doi: 10.1016/B978-0-444-64150-2.00018-6.

Brouwer, R. M., Schutte, J., Janssen, R., Boomsma, D. I., Pol, H. E. H. & Schnack, H. G. (2021). The Speed of Development of Adolescent Brain Age Depends on Sex and Is Genetically Determined. *Cerebral Cortex, 31*(2), 1296-1306. https://doi.org/10.1093/cercor/bhaa296

Brown, T. T., Kuperman, J. M., Chung, Y., Erhart, M., McCabe, C., Hagler Jr., D. J. … Dale, A. M. (2012). Neuroanatomical assessment of biological maturity. *Current Biology, 22*(18), 1693-1698. https://doi.org/10.1016/j.cub.2012.07.002

Casey, B. J., Cannonier, T., Conley. M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E. … ABCD Imaging Acquisition Workgroup. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience, 32*, 43-54. 10.1016/j.dcn.2018.03.001

Cheng, J., Edwards, L. J., Maldonado-Molina, M. M., Komro, K. A. & Muller, K. E. (2010). Real Longitudinal Data Analysis for Real People: Building a Good Enough Mixed Model. *Statistics in Medicine, 29*(4), 504-520. doi: 10.1002/sim.3775

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

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Steves, C., Spector, T. D. & Montana, G. (2017). Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage, 163*, 115-124. https://doi.org/10.1016/j.neuroimage.2017.07.059

Cole, J. H. (2020). Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors. Neurobiology of Aging, 92, 34-42. https://doi.org/10.1016/j.neurobiolaging.2020.03.014

Di Biase, M. A., Tian, Y. E., Bethlehem, R. A. I. & Zalensky, A. (2023). Mapping human brain charts cross-sectionally and longitudinally. *Neuroscience, 120*(20). https://doi.org/10.1073/pnas.221679812

Elliott, M. L., Belsky, D. W., Knodt, A. R., Ireland, D., Melzer, T. R., Poulton, R., … Hariri, A. R. (2021). Brain-age in midlife is associated with accelerated biological aging and

cognitive decline in a longitudinal birth cohort. *Molecular Psychiatry, 26*, 3829–3838. https://doi.org/10.1038/s41380-019-0626-7

Elman, J. A., Vogel, J. W., Bocancea, D. I., Ossenkoppele, R., van Loenhoud, A. C., Tu, X. M., Kremen, W. S. & the Alzheimer's Disease Neuroimaging Initiative. (2022). Issues and recommendations for the residual approach to quantifying cognitive resilience and reserve. *Alzheimer's Research & Therapy, 14*(102). https://doi.org/10.1186/s13195-022-01049-w

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A New Look at an Old Issue. *Psychological Methods, 12*(2), 121-138. Doi: 0.1037/1082-989X.12.2.121

Fortin, J. P., Cullen, N., Sheline, Y., Taylor, W. D., Aselcioglu, I., Cook, P. A., … Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage, 167*, 104-120. https://doi.org/10.1016/j.neuroimage.2017.11.024

Franke, K. & Gaser, C. (2019). Ten years of brainAGE as a neuroimaging biomarker of brain aging: What insights have we gained? *Frontiers in Neurology, 10.* https://doi.org/10.3389/fneur.2019.00789

Fuhrmann, D., Madsen, K. S., Johansen, L. B., Baaré, W. F. C. & Kievit, R. A. (2022). The midpoint of cortical thinning between late childhood and early adulthood differs between individuals and brain regions: Evidence from longitudinal modelling in a 12-wave neuroimaging sample. *NeuroImage, 261*(119507). 10.1016/j.neuroimage.2022.119507

Gavett, B. E. & Horwitz, J.E. (2012). Immediate list recall as a measure of short-term episodic memory: Insights from the serial position effect and item response theory. *Archives of Clinical Neuropsychology, 27*, 125-135. doi: 10.1093/arclin/acr104.

Genon, S., Eickhoff, S. B. & Kharabian, S. (2022). Linking interindividual variability in brain structure to behaviour. *Nature Reviews Neuroscience, 23*, 307–318 (2022). https://doi.org/10.1038/s41583-022-00584-7

Giedd, J., Blumenthal, J., Jeffries, N., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C. & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience, 2*, 861–863. https://doi.org/10.1038/13158

Goddings, A. L., Mills, K. L., Clasen, L. S., Giedd, J. N., Viner, R. M. & Blakemore, S. J. (2014). The influence of puberty on subcortical brain development. *Neuroimage, 88*, 242–251. doi: 10.1016/j.neuroimage.2013.09.073

Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep learning. MIT Press, p. 164-172, 326-339. http://www.deeplearningbook.org

Han, J., Kim, S. Y., Lee, J. & Lee, W. H. (2022). Brain Age Prediction: A Comparison between Machine Learning Models Using Brain Morphometric Data. *Sensors, 22*(8077). https://doi.org/10.3390/s22208077

Herting, M. M., Uban, K. A., Gonzalez, M. R., Baker, F. C., Kan, E. C., Thompson, W. K., … Sowell, E. R. (2021). Correspondence between perceived pubertal development and hormone levels in 9-10 year-olds from the adolescent brain cognitive development Study. *Frontiers in Endocrinology, 11*. 10.3389/fendo.2020.549928

Hoffman, L. & Walters, R. W. (2022). Catching Up on Multilevel Modeling. Annual Review of Psychology, 73, 659-689. https://doi.org/10.1146/annurev-psych-020821-103525

Holm, M. C., Leonardsen, E. H., Beck, D., Dahl, A., Kjelkenes, R, de Lange, A. M. & Westlye, L. (2023). Linking brain maturation and puberty during early adolescence using longitudinal brain age prediction in the ABCD cohort. *Developmental Cognitive Neuroscience, 60*(101220). https://doi.org/10.1016/j.dcn.2023.101220

Jeneson, A. & Squire, L. R. (2012). Working memory, long-term memory, and medial temporal lobe function. *Learn Mem., 19*(1), 15-25. doi: 10.1101/lm.024018.111

Jenkinson, M. & Smith, S. (2001). A global optimization method for robust affine registration of brain images. *Medical Image Analysis, 5*(2), 143-156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich. M. W. & Smith, S. M. (2012). FSL. *NeuroImage, 62*(2), 782-790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Karpouzian-Rogers, T., Makowski-Woidan, B., Kuang, A., Zhang, H., Fought, A., Engelmeyer, J., … Rogalski, E. (2023). NIH Toolbox® Episodic Memory Measure Differentiates Older Adults with Exceptional Memory Capacity from those with Average-for-Age Cognition. *Journal of the International Neuropsychological Society, 29*(2), 230-234. doi: 10.1017/S135561772200008X

Kaufmann, K., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I. & Westlye, L. T. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience, 22*, 1617-1623. https://doi.org/10.1038/s41593-019-0471-7

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352-358. https://doi.org/10.1037/h0043688

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Larsena, B. & Luna, B. (2018). Adolescence as a neurobiological critical period for the development of higher-order cognition. *Neuroscience & Biobehavioral Reviews, 94*, 179-195. doi: 10.1016/j.neubiorev.2018.09.005

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature, 521*, 436–444. https://doi.org/10.1038/nature14539

Leonardsen, E. H., Peng, H., Kaufmann, T., Agartz, I., Andreassen, O. A., Celius, E. G., … Wang, Y. (2022). Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage, 256*(119210). https://doi.org/10.1016/j.neuroimage.2022.119210

Liao, S. & Carneiro, G. (2016). On the importance of normalisation layers in deep learning with piecewise linear activation units. *Institute of Electrical and Electronics Engineers, Winter Conference on Applications of Computer Vision (WACV),* 1-8. doi: 10.1109/WACV.2016.7477624.

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

Lugtmeijer, S., de Haan, E. H. F. & Kessels, R. P. C. (2019). A comparison of visual working memory and episodic memory performance in younger and older adults. *Aging, Neuropsychology, and Cognition, 26*(3), 387-406. https://doi.org/10.1080/13825585.2018.1451480

Magalhães, S. S., Malloy-Diniz, L. F. & Hamdan, A. (2012). Validity convergent and reliability test-retest of the rey auditory verbal learning test. *Clinical Neuropsychiatry, 9*(3), 129-137. Retrieved from https://www.researchgate.net/publication/287784953_Validity_convergent_and_reliability_test-retest_of_the_rey_auditory_verbal_learning_test

Mechie, I. R., Plaisted-Grant, K. & Cheke L. G. (2021). How does episodic memory develop in adolescence? *Learn Mem., 28*(6), 204–217. doi: 10.1101/lm.053264.120.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97. https://doi.org/10.1037/h0043158

Mills, K. L., Siegmund, K. D., Tamnes, C. K., Ferschmann, L., Wierenga, L. M., Bos, M. G. N., … Herting, M. M. (2021). Inter-individual variability in structural brain development from late childhood to young adulthood. *NeuroImage, 242*(118450). https://doi.org/10.1016/j.neuroimage.2021.118450

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., … Mérillat, S. (2019). Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Reviews in the Neurosciences, 31*(1), 1–57. https://doi.org/10.1515/revneuro-2018-0096

Owens, M. M., Potter, A., Hyatt, C. S., Albaugh, M., Thompson, W. K., Jernigan, T., ... Garavan, H. (2021). Recalibrating expectations about effect size: A multi-method survey of effect sizes in the ABCD study. *PloS one, 16*(9). https://doi.org/10.1371/journal.pone.0257535

Pedersen, M. L., Alnæs, D., van der Meer, D., Fernandez-Cabello, S., Berthet, Dahl, A., … Westlye, L. T. (2023). Computational modeling of the n-back task in the ABCD Study: Associations of drift diffusion model parameters to polygenic scores of mental disorders and cardiometabolic diseases. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 8*(3), 290-299. https://doi.org/10.1016/j.bpsc.2022.03.012

Pelegrina, S, Lechuga, M. T., Garcia-Madruga, J. A., Elosúa, M. R., Macizo, P., Carreiras,

M., … Bajo, M. T. (2015). Normative data on the n-back task for children and young adolescents. *Frontiers in Psychology, 6*(1544). https://doi.org/10.3389/fpsyg.2015.01544

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A. & Smith, S. M. (2021). Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis, 68*(101871). https://doi.org/10.1016/j.media.2020.101871

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rosenberg, M. D., Martinez, S. A., Rapuano, K. M., Conley, M. I., Cohen, A. O., Cornejo, M. D., … Wager, T. D. (2020). Behavioral and neural signatures of working memory in childhood. *Journal of Neuroscience, 40*(26) 5090-5104. https://doi.org/10.1523/JNEUROSCI.2841-19.2020

Rousseeuw, P. J. & Hubert, M. (2018). Anomaly detection by robust statistics. *WIREs Data Mining and Knowledge Discovery, 8*(2). https://doi.org/10.1002/widm.1236

Saragosa-Harris, N. M., Chaku, N., MacSweeney, N., Williamson, V. G., Scheuplein, M., Feola, B., … Mills, K. L. (2022). A practical guide for researchers and reviewers using the ABCD Study and other large longitudinal datasets. *Developmental Cognitive Neuroscience, 55*(101115). https://doi.org/10.1016/j.dcn.2022.101115

Schneider, W. & Ornstein, P.A. (2019), Determinants of memory development in childhood and adolescence. *International Journal of Psychology, 54*(3), 307-315. https://doi-org.ezproxy.uio.no/10.1002/ijop.12503

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K. & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. NeuroImage, 22(3), 1060-1075. https://doi.org/10.1016/j.neuroimage.2004.03.032

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., … Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. Nature, 440, 676-679. https://doi.org/10.1038/nature04513

Shin, H. C., Holger, R., Roth, H. R., Gao, M., Lu, L., Xu, Z., … Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *Institute of Electrical and Electronics Engineers, 35*(5), 1285-1298. doi: 10.1109/TMI.2016.2528162.

Smith, D. M., Loughnan, R., Friedman, N. P., Parekh, P., Frei, O., Thompson, W. K., … Dale, A. M. (2023). Heritability Estimation of Cognitive Phenotypes in the ABCD Study® Using Mixed Models. *Behavior Genetics, 53*(3), 169-188. doi: 10.1007/s10519-023-10141-2

Sneve, M. H., Grydeland, H., Nyberg, L., Bowles, B., Amlien, I. K., Langnes, E., … Fjell, A. M. (2015). Mechanisms underlying encoding of short-lived versus durable episodic memories. *Journal of Neuroscience, 35*(13), 5202-5212. doi: 10.1523/JNEUROSCI.4434-14.2015

Tamnes, C. K., Bos, M. G. N., van de Kamp, F., Peters, S. & Crone, E. A. (2018). Longitudinal development of hippocampal subregions from childhood to adulthood. *Developmental Cognitive Neuroscience, 30*, 212-222. https://doi.org/10.1016/j.dcn.2018.03.009

Tejavibulya, L., Rolison, M., Gao, S., Liang, O., Peterson, H., Dadashkarimi, J., … Scheinost, D. (2022). Predicting the future of neuroimaging predictive models in mental health. *Molecular Psychiatry, 27*, 3129–3137. https://doi.org/10.1038/s41380-022-01635-2

Tulsky, D. S, Carlozzi, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. A., Mungas, D., … Gershon, R. (2014). NIH toolbox cognition battery (NIHTB-CB): List sorting test

to measure working memory. *Journal of the International Neuropsychological Society, 20*(6), 599-610. 10.1017/S135561771400040X

Valizadeh, S. A., Hänggi, J., Mérillat, S. & Jäncke, L. (2017). Age prediction on the basis of brain anatomical measures. *Human Brain Mapping, 38*(2). Doi: 10.1002/hbm.23434

Vidal-Pineiro, D., Wang, Y., Krogsrud, S. K., Amlien, I. K., Baaré, W. F., Bartres-Faz, D., … Fjell, A. (2021). Individual variations in 'brain age' relate to early-life factors more than to longitudinal brain change. *Elife, 10*. Retrieved from https://elifesciences.org/articles/69995

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., … Weiss, S. R. B. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience, 32,* 4-7. https://doi.org/10.1016/j.dcn.2017.10.002

Walhovd, K. B., Krogsrud, S. K., Amlien, I. K., Bartsch, H., Bjørnerud, A., Due-Tønnesen, P., … Fjell, A. M. (2016). Neurodevelopmental origins of lifespan changes in brain and cognition. *Proceedings of the National Academy of Sciences, 133*(33), 9357-9362. https://doi.org/10.1073/pnas.1524259113

Waris, O., Fellman, D., Jylkkä, J. & Laine, M. (2021). Stimulus novelty, task demands, and strategy use in episodic memory. *Quarterly Journal of Experimental Psychology, 74*(5), 872-888. doi: 10.1177/1747021820980301

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology, 80*(11 Suppl 3), 54–64. https://doi.org/10.1212/WNL.0b013e3182872ded

Zhou, D., Lebel, C., Treit, S., Evans, A. & Beaulieu, C. (2015). Accelerated longitudinal cortical thinning in adolescence. *NeuroImage, 104*, 138-145. https://doi.org/10.1016/j.neuroimage.2014.10.005

Wickham, H. (2016). Ggplot2: Elegant graphics for data analysis (2nd ed.). *Springer International Publishing.*

Yin, W., Li, L. & Wu, F-X. (2022). Deep learning for brain disorder diagnosis based on fMRI images. *Neurocomputing, 469*, 332-345. https://doi.org/10.1016/j.neucom.2020.05.113

© 2022 Toolbox Assessments. (n. d.). Picture Sequence Memory Test. Retrieved April, 2023, from https://www.nihtoolbox.org/test/picture-sequence-memory-test/

Appendix

Table A1

Models 1 and 2 assessing working memory and brain age

	Model 1: 2-back score				Model 2: N-back score			
Fixed effects	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	3.749	1.585	23.647	<0.001	-5.970	4.812	-12.407	<0.001
Timepoint	1.810	3.093	5.851	<0.001	1.465	2.907	5.040	<0.001
Brain age	-2.927	1.237	-0.237	0.812	5.765	1.189	0.485	0.628
Age	2.407	1.500	16.043	<0.001	3.848	1.372	2.805	0.005
Sex †	-1.281	1.900	-6.741	<0.001	-1.444	1.740	-8.296	<0.001
PDS	-4.484	3.255	-1.377	0.168	1.481	3.123	0.047	0.962
Ethnicity2	-4.547	3.043	-14.942	<0.001	-7.192	2.790	-2.578	<0.01
Ethnicity3	-3.113	2.295	-13.565	<0.001	-1.559	2.084	-7.482	<0.001
Ethnicity4	1.213	6.582	1.842	0.065	4.867	5.968	0.815	0.414
Ethnicity5	-9.837	2.966	-3.316	<0.001	5.481	2.689	0.002	0.998
Birthweight	8.757	1.357	6.456	<0.001	3.229	1.227	2.630	<0.01
Time:Brain age	-9.969	1.319	-0.756	0.450	-1.186	1.349	-0.880	0.379
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	30.16	5.49			13.20	3.63		
Residual	31.26	5.59			40.89	6.40		

Note. PDS = Pubertal developmental scale. Ethnicity based on self-report (1/intercept = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = other). * Approximated p-values. † Sex coded with male = 0, female = 1.

Table A2

Models 3 and 4 assessing episodic memory and brain age

	Model 3: Picture sequence task score				Model 4: Composite score			
Fixed effects	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	1.040	6.364	163.377	<0.001	9.983	6.239	160.023	<0.001
Timepoint	2.449	3.845	6.369	<0.001	-7.175	3.681	-1.949	0.051
Brain age	2.309	1.398	1.652	0.099	1.986	1.326	1.497	0.134
Age	2.207	1.814	12.168	<0.001	2.548	1.787	14.256	<0.001
Sex †	2.146	2.287	9.385	<0.001	2.889	2.254	12.817	<0.001
PDS	-1.701	3.999	-4.252	<0.001	-2.372	3.693	-6.422	<0.001
Ethnicity2	-7.508	3.389	-22.154	<0.001	-8.477	3.350	-25.308	<0.001
Ethnicity3	-2.547	2.771	-9.191	<0.001	-3.009	2.765	-10.881	<0.001
Ethnicity4	2.212	7.861	2.814	<0.01	2.113	7.738	2.731	<0.01
Ethnicity5	-2.182	3.616	-6.033	<0.001	-2.628	3.604	-7.291	<0.001
Birthweight	8.048	1.637	4.917	<0.001	1.017	1.631	6.237	<0.001

Time:Brain age	-3.297	1.563	-2.110	0.035	-3.993	1.357	-2.941	0.003
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	57.91	7.61			75.76	8.70		
Residual	76.43	8.74			47.86	6.92		

Note. PDS = Pubertal developmental scale. Ethnicity based on self-report (1/intercept = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = other). * Approximated p-values. † Sex coded with male = 0, female = 1.

Table A3

Model 5 assessing delta and memory

	Model 5: Delta			
Fixed effects	**Coef. (B)**	**SE (B)**	**t-value**	**p-value***
Intercept	4.134	2.116	19.532	<0.001
Timepoint	8.170	4.302	1.899	0.058
WM	5.886	1.662	0.354	0.723
EM	-1.332	1.208	-1.102	0.270
Age	-4.552	2.014	-22.605	<0.001
Sex †	-2.746	2.672	-1.027	0.304
PDS	5.049	4.419	11.426	<0.001
Ethnicity2	4.928	4.351	1.133	0.257
Ethnicity3	2.312	3.237	0.714	0.475
Ethnicity4	-6.526	9.103	-0.072	0.943
Ethnicity5	9.261	4.149	2.232	0.026
Birthweight	-1.803	1.900	-0.949	0.343
Time:WM	-2.408	2.980	-0.808	0.419
Time:EM	4.754	1.736	0.274	0.784
Random intercept	**Variance**	**SD**		
Subject	0.66	0.81		
Residual	0.48	0.70		

Note. WM = Working memory. EM = Episodic memory. PDS = Pubertal developmental scale. Ethnicity based on self-report (1/intercept = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = other). * Approximated p-values. † Sex coded with male = 0, female = 1.

Table A4

Models 6 and 7 assessing working memory and brain age

	Model 6: 2-back score				Model 7: N-back score			
Fixed effects	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef. (*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	4.258	4.625	9.265	<0.001	-4.246	1.184	-3.587	<0.001
Timepoint	1.698	9.322	1.822	0.069	1.882	8.211	-3.587	0.022
Brain age	5.532	2.886	1.917	0.055	7.835	2.682	2.992	0.004

Age	1.990	4.341	4.584	<0.001	4.895	3.730	0.131	0.900
Sex †	-1.227	5.534	-2.214	0.027	-1.500	4.787	-1.133	0.002
PDS	1.361	8.161	0.167	0.868	-1.889	7.414	-0.255	0.800
Ethnicity2	-5.166	8.165	-6.327	<0.001	-2.369	-2.369	-3.369	<0.001
Ethnicity3	-3.716	6.485	-5.729	<0.001	-2.613	-2.613	-4.717	<0.001
Ethnicity4	2.201	2.156	0.102	0.919	1.951	1.951	1.074	0.283
Ethnicity5	-1.029	8.668	-1.187	0.235	-4.056	-4.056	-0.547	0.585
Birthweight	5.303	3.700	1.433	0.152	3.009	3.009	0.095	0.923
Time:Brain age	-5.834	2.776	-2.101	0.036	-8.567	-8.567	-3.163	0.002
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	25.47	5.047			11.33	3.367		
Residual	32.09	5.665			34.43	5.868		

Note. PDS = Pubertal developmental scale. Ethnicity based on self-report (1/intercept = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = other). * Approximated p-values. † Sex coded with male = 0, female = 1.

Table A5

Models 8 and 9 assessing episodic memory and brain age

	Model 3: Picture sequence task score				Model 4: Composite score			
Fixed effects	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***	**Coef.(*B*)**	**SE (*B*)**	**t-value**	**p-value***
Intercept	1.032	1.856	55.624	<0.001	9.950	1.790	55.589	<0.001
Timepoint	2.900	1.254	2.313	0.021	5.572	1.202	0.464	0.643
Brain age	5.383	3.385	1.590	0.112	4.610	3.057	1.508	0.132
Age	1.667	5.861	2.844	0.005	1.630	5.737	2.840	0.005
Sex †	1.779	7.381	2.410	0.016	1.712	7.179	2.385	0.017
PDS	-1.205	1.072	-1.124	0.261	-1.643	9.481	-1.733	0.083
Ethnicity2	-8.795	1.072	-8.207	<0.001	-9.296	1.051	-8.841	<0.001
Ethnicity3	-3.935	8.679	-4.534	<0.001	-3.949	8.594	-4.595	<0.001
Ethnicity4	7.451	2.905	2.565	0.010	6.759	2.872	2.353	0.019
Ethnicity5	-1.721	1.184	-1.454	0.146	-2.130	1.168	-1.824	0.069
Birthweight	1.249	5.028	2.485	0.013	1.200	4.966	2.416	0.016
Time:Brain age	-6.360	3.171	-2.006	0.045	-5.767	2.653	-2.174	0.030
Random intercept	**Variance**	**SD**			**Variance**	**SD**		
Subject	56.55	7.520			68.71	8.289		
Residual	74.86	8.652			45.90	6.775		

Note. PDS = Pubertal developmental scale. Ethnicity based on self-report (1/intercept = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = other). * Approximated p-values. † Sex coded with male = 0, female = 1.

Table A6

Model 10 assessing delta and memory

	Model 5: Delta			
Fixed effects	**Coef. (B)**	**SE (B)**	**t-value**	**p-value***
Intercept	9.587	1.269	7.555	<0.001
Timepoint	1.475	2.443	6.039	<0.001
WM	1.038	8.989	1.154	0.249
EM	1.911	6.482	0.295	0.768
Age	-1.109	1.186	-9.347	<0.001
Sex †	-2.819	1.501	-0.188	0.851
PDS	1.241	1.936	6.412	<0.001
Ethnicity2	4.737	2.317	2.045	0.041
Ethnicity3	7.193	1.810	0.397	0.691
Ethnicity4	3.035	6.086	0.499	0.618
Ethnicity5	2.613	2.370	1.103	0.271
Birthweight	2.085	1.015	2.055	0.040
Time:WM	-2.205	1.248	-1.766	0.079
Time:EM	-1.140	7.088	-1.609	0.108
Random intercept	**Variance**	**SD**		
Subject	2.723	1.650		
Residual	1.204	1.097		

Note. WM = Working memory. EM = Episodic memory. PDS = Pubertal developmental scale. Ethnicity based on self-report (1/intercept = White, 2 = Black, 3 = Hispanic, 4 = Asian, 5 = other). * Approximated p-values. † Sex coded with male = 0, female = 1.

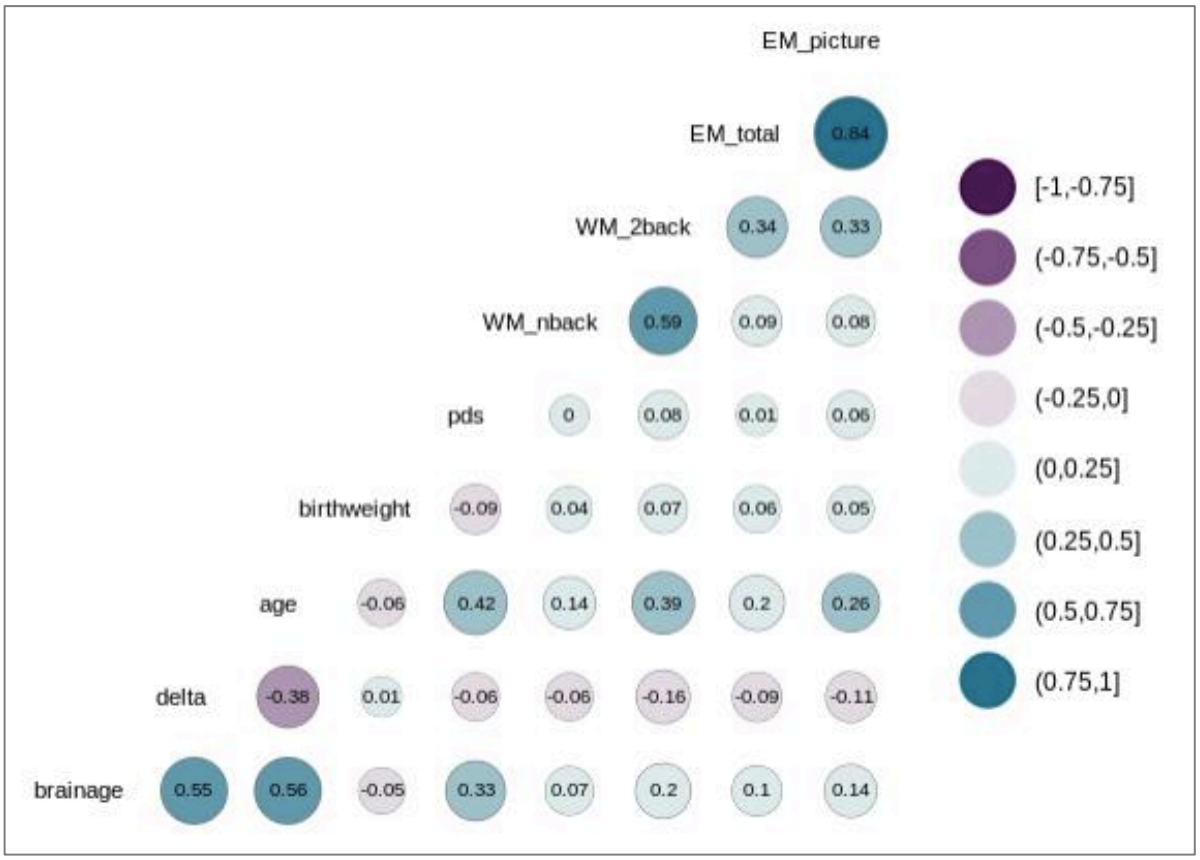

Figure A7. Pearson correlations between main variables and continuous covariates.