Maturation of the brain in adolescents who develop a mood disorder.

Laura de Nooij

University of Edinburgh

A thesis presented for the degree of Master of Science

09-08-2018

University of Amsterdam

Research Master's Psychology

Student Number: 10341757

Supervised by:

Dr. Tim Ziermans, University of Amsterdam

Dr. Heather Whalley, University of Edinburgh

Abstract

Background: Support for the accelerated brain ageing hypothesis for mood disorders originates from cross-sectional studies of adults. It is therefore yet unknown whether differential adolescent brain maturation is associated with illness.

Methods: This study implemented a structural MRI-based brain age prediction model to estimate 'biological brain age' for initially well young individuals (16-25) included in the prospective longitudinal Scottish Bipolar Family Study (SBFS), at baseline and follow-up, two years apart**.** Groups were categorised as controls with no family history who remained well (C-well, n = 98) and who developed a mood disorder (C-MD, n = 13), and those at high familial risk who remained well (HR-well, n = 73) and who developed a mood disorder (HR-MD, n = 38). Each individual's brain age was compared with chronological age to derive a 'brain age gap estimate' (BrainAGE), capturing deviation from typical brain maturation. BrainAGEs and longitudinal changes herein were compared across groups with a mixed effects model.

Results: HR-well showed a delay in brain maturation at baseline compared to C-well (β = -0.38 years, p = .035), which remained constant over time. In contrast, C-MD showed no initial delay but a deceleration in maturation over time (β = -1.33 years, p < .001). HR-MD showed a trend of an initial delay (β = -0.41 years, p=0.06) and of deceleration in brain maturation (β = -0.39 years, p = .07).

Discussion: These findings suggest brain maturation differences may emerge over time on route to illness as a lag in maturational trajectory. Given modest sample sizes however, further replication is required.

Keywords: mood disorder, brain age prediction, brain maturation, structural MRI.

Maturation of the Brain in Adolescents Who Develop a Mood Disorder

Mood disorders are amongst the most prevailing psychiatric disorders, with a life-time prevalence of around 15% (Kessler & Bromet, 2013). Globally, they are the greatest contributor to non-fatal health loss (World Health Organization, 2017). Major Depressive Disorder (MDD) is the most prevalent mood disorder and is characterised by episodes of low mood and/or loss of interest or pleasure, in combination with other psychological, behavioural and cognitive symptoms. Within Bipolar Disorder (BD) episodes of mood disturbances occur at the other side of the spectrum in the form of (hypo)manic episodes, characterised by symptoms such as euphoria, increase in energy, recklessness and risky behaviour; usually one or more depressive episodes occur as well. These often occur before the manic episode, so that BD initially manifests with a depressive episode. MDD and BD are known to have a shared genetic architecture, and as a result, individuals with a family history of BD are at risk of developing either mood disorder, with the absolute risk of developing MDD being at least twice as high as the risk of developing BD (Smoller & Finn, 2003).

Both MDD (Wolkowitz, Reus & Mellon, 2011) and BD (Rizzo et al., 2014) are proposed to be related to accelerated ageing, as biological mechanisms related to ageing, such as inflammation and oxidative stress, overlap with biological pathways implicated in these psychiatric disorders (for a review, see Sibille, 2013). Furthermore, shorter leukocyte telomere length (TL) is regarded as biomarker of ageing based on its association with oxidative stress and cellular senescence (two molecular processes thought to regulate ageing; for a review, see Sanders & Newman, 2013), and TL has been consistently found to be shorter within mood disorders subjects (Simon et al., 2006; Ridout, Ridout, Price, Sen & Tyrka, 2015; Schutte & Malouf, 2015).

Moreover, mood disorders have been associated with age-related diseases, such as type

II diabetes (Mezuk, Eaton, Albrecht & Golden, 2008; Regenold, Thapar, Marano, Gavirneni & Kondapavuluru, 2003), coronary heart disease (for MDD; Pan, Sun, Okereke, Rexrode & Hu, 2011; Whooley et al., 2008) and obesity (for MDD; Luppino et al., 2010), and with increased mortality rate (Ösby, Brandt, Correia, Ekbom & Sparén, 2001; Schulz et al., 2000). Some of these associations remained significant when considering depression as independent risk factor (e.g., Mezuk et al, 2008; Schulz et al, 2000), suggesting the involvement of biological mechanisms in mood disorders that relate to accelerated ageing

Since mood disorders are considered to be disorders of the brain, this support for accelerated ageing raises the question of whether mood disorders are also associated with accelerated ageing of the brain. This question is especially pertinent due to increasing evidence of structural brain alterations associated with mood disorders obtained using structural Magnetic Resonance Imaging (MRI). Meta-analyses summarise the most consistently found structural brain differences associated with MDD and BD: enlarged lateral ventricle volume (Kempton et al., 2011) and reduced grey matter volume (GMV) in various brain areas (e.g., Arnone, McIntosh, Ebmeier, Munafò, & Anderson, 2011; Bora, Fornito, Pantelis, & Yücel, 2012; Kempton et al., 2011; Wise et al., 2017), although Ioannidis (2011) argues that such meta-analyses probably overestimate the structural differences due to publication bias and selective reporting of explorative results. The Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium (http://enigma.ini.usc.edu/) collaborates internationally to combine data of multiple research samples, and is therefore less vulnerable to publication bias. Including 35 research samples with a total of 2148 MDD patients, analysis by the ENIGMA MDD working group showed associations of MDD with thinner cortical GM in the OFC, posterior and anterior cingulate, insula and temporal lobes within adults, and with lower GMV in prefrontal regions and primary and higher-order visual, somatosensory and motor areas within adolescents (Schmaal et al., 2017). Subcortically, a

lower hippocampal GMV was found to be associated with MDD, greatest for recurrent MDD (Schmaal et al., 2016). For BD, ENIGMA results showed thinner cortical GM bilaterally in frontal, temporal and parietal regions (Hibar et al., 2018). Of note, many of these structural brain characteristics implicated in mood disorders (i.e., thinner cortex and lower GMV) are also related to ageing (e.g., Scahill et al., 2003), thus supporting the accelerated ageing hypothesis.

Recently, a new framework for the investigation of brain maturation and ageing in relation to neuropsychiatry has emerged. Such fundamental research gives insight into the relationship between mental disorders and the brain, increasing the understanding of the illnesses, and ultimately contributing to better treatment. In parallel with other measures of biological ageing such as the epigenetic clock (Horvath, 2013), one can study brain maturation within a cross-sectional design by comparing the "biological brain age", derived by application of an MRI-based brain age prediction model, with the chronological age; a large brain age gap estimate (BrainAGE) between those would reflect delayed or advanced ageing. Based on this principle, researchers have developed and validated multiple methodologies to predict brain age, most of which apply a linear or non-linear regression method to grey matter brain maps, although other types of prediction models (e.g., deep learning based on raw T1-weighted images; Cole et al., 2017) have been applied as well (for an overview see: Cole, Marioni, Harris & Deary, 2018). Specifically, Relevance Vector Regression (RVR) has been most often implemented within previous brain age models (Cole et al., 2018). RVR is a sparse Bayesian alternative to Support Vector Regression (SVR) that does not require optimisation of hyperparameters (C, ε) through cross-validation (Tipping, 2001). Previously, application of RVR with linear kernel has indicated the most favourable performance for prediction of BrainAGE when compared to various other regression methods (Franke, Ziegler, Klöppel & Gaser, 2010).

Research that addressed the accelerated ageing hypothesis for mood disorders by associations with TL or with BrainAGE predominantly included adult participants. Koutsouleris and colleagues (2014) implemented a linear SVR model and found higher BrainAGE scores for adults with BD (+3.1 years; *n* = 57) and MDD (+4.0 years; *n* = 104) as compared to healthy control participants (*n* = 437). In contrast, Nenadić and colleagues (2017) found no differing BrainAGE scores for adults with BD (*n* = 22) versus healthy controls (*n* = 70) after application of a linear RVR model.

However, many psychiatric disorders manifest during or shortly after the transition from adolescence to young adulthood, and this period involves great neurodevelopmental change (for a review, see: de Girolamo, Dagani, Purcell, Cocchi, & McGorry, 2012). The temporal origins of the probable association between accelerated brain ageing and mood disorders in adulthood are yet unknown, but could occur during this important developmental period. A second, alternative hypothesis would be that mood disorder onset during adolescence and young adulthood is associated with delayed brain maturation instead. During adolescence, cubic GMV development trajectories reach their peak (e.g., Giorgio et al., 2010; Tamnes et al., 2010). Subsequent decreases in GMV result from “synaptic pruning”, a process that is thought to fine-tune neuronal connectivity and consequently enhance cognitive functioning. Therefore, delayed brain maturation could result in emotional instability, increasing vulnerability to the development of mood disorders – especially if maturation is delayed within brain systems involved in emotion regulation and cognitive control.

Only one previous cross-sectional study investigated the association between brain maturation and mood disorders in adolescents by application of the BrainAGE framework: with the use of RVR with linear kernel, Hajek and colleagues (2017) found comparable BrainAGE scores between adolescent and young adult subjects at high familial risk that had developed a mood disorder (*n* = 48), those who remained well (*n* = 48), and control subjects

(n = 60).

A longitudinal volumetric study by Whittle and colleagues (2014) showed the potential importance of divergence in brain maturation in either direction, as they found a moderation by sex in the association between amygdala GMV and depression onset. This suggests a third possibility for the nature of the association between brain maturation and adolescent onset of depression: arguably, accelerated as well as decelerated maturation may be disadvantageous, so that any atypical form of adolescent brain maturation trajectory could be related to depression onset. If so, the direction of the effect may differ over time in such a way it could not be captured by cross-sectional studies (such as Hajek et al., 2017).

Within cross-sectional studies it additionally remains unclear when the brain age gap emerges, and to what extent it is associated with mood disorders as predisposition, cause, or consequence. For instance, while Koutsouleris et al. (2014) found a more positive BrainAGE indicating an increased brain age for those with early onset of MDD, this does not show causality, nor a direction; as brain maturation may affect mood disorder onset, but mood disorders may also affect further ageing, the relationship between brain maturation (as indicated by BrainAGE) and mood disorder onset remains unknown.

The current study investigated this relationship between brain maturation and the development of mood disorder by application of the BrainAGE framework within a longitudinal design, starting before mood disorder onset. We used data from the Scottish Bipolar Family Study (SBFS), a prospective longitudinal study that included young individuals who were all initially well, and of whom some had a family history of BD. The dynamics of brain maturation trajectories were captured by changes in BrainAGE over two years. Recognising similarities between BD and MDD and the difficulty of differential diagnosis at young age, we predicted that the development of any early-onset mood disorder (defined as having an onset in adolescence or young adulthood) was associated with divergent

brain maturation trajectories showing either (i) accelerated brain maturation, or (ii) decelerated brain maturation, or (iii) increased individual divergence in brain maturation in both directions. This would be reflected by significant differences in BrainAGE over time for participants that developed a mood disorder as compared to those who remained well. Furthermore, familial risk may be expressed by a cross-sectional difference in BrainAGE at baseline between those with family history and those with no family history of mood disorder.

Method

Study Participants

Participants are adolescents and young adults (N = 288, age 16-25 years) recruited as part of the Scottish Bipolar Family Study (SBFS; for more information, see: Papmeyer et al., 2015; Sprooten et al., 2011; Whalley et al., 2011). Participants at high familial risk of mood disorder, subsequently referred to as HR-participants, have at least one first-degree relative or two second-degree relatives with BD type-I, and are thus at increased risk of developing a mood disorder (both BD and MDD; Smoller & Finn, 2003). Unrelated control participants without family history of BD were recruited via HR-participants, and were matched to the HR-group by age and sex. Inclusion criteria ensured that all participants were with no personal history of MDD, mania or hypomania, psychosis, or any other major neurological or psychiatric disorder, substance dependence, learning disability, or head injury that included loss of consciousness, and that they were without contraindications to MRI.

The following exclusion criteria were applied: (i) missing MRI or age data, (ii) baseline scans of insufficient image or segmentation quality that could not be replaced by time2 MRI scans (see supplemental method), (iii) ambiguous mood disorder diagnosis, or diagnosis of another psychiatric disorder, and (iv) high familial risk for mood disorder without follow-up measurement. These criteria excluded 67 participants (see Table S1 for specification), reducing the sample size to a total of 222 participants at time1 (111 HR-participants). For

time2, data was available from 141 of these participants (83 HR-participants). This difference in sample size between time1 and time2 was mainly due to dropout. Table 1 provides the sample sizes and characteristics per group after post-study group division (see 'comparison of brain maturation trajectories').

Procedure

Participants of the SBFS were invited every two years for a total of four assessments over six years. To ensure that all participants were initially well, participants were interviewed

Table 1

Sample sizes and characteristics per group after application of exclusion criteria and group division based on clinical information.

	C-well		HR-well		HR-MD		C-MD	
	Time1	Time2	Time1	Time2	Time1	Time2	Time1	Time2
N	98	51	73	48	38	32	13	7
Age								
M	21.18	23.07	21.46	23.70	21.55	23.26	23.00	24.92
(SD)	(2.44)	(2.43)	(2.80)	(2.78)	(3.38)	(3.04)	(2.20)	(2.22)
Min.	16.3	18.3	15.2	17.6	16.0	18.1	18.9	20.9
Max.	25.6	27.6	26.6	28.1	30.0	28.1	25.8	27.4
% Males	43.8	39.2	50.6	52.1	42.1	40.6	46.1	42.9

M = Mean, SD = Standard deviation, Min = Minimum, Max = Maximum.

Note. The HR-well group consisted of participants at high familial risk for mood disorder who remained well throughout the course of the study (as far as known when considering drop-out), and HR-MD of those at familial risk who developed a mood disorder at any point throughout the course of the study (either BD or MDD). Similarly, C-well refers to control participants (without familial risk) who remained well, while C-MD consists of control participants who developed a mood disorder.

and screened with the Structured Clinical Interview for DSM-IV Axis-I Disorders (SCID; First, Spitzer, Gibbon & Williams, 2002). At later assessments (time2, time3, time4), the SCID was used to determine the presence of any mood disorder meeting diagnostic criteria at any time since the previous assessment. All SCID assessments were completed by trained psychiatrists. The participant's age at the time of each assessment was registered in years with a precision of two decimals. Assessments at time1, time2 and time4 included a MRI session, although only time1 and time2 MRI measurements were considered within this study to restrict to a single scanner and avoid potential bias by the introduction of an additional scanner with differences in strength and acquisition method.

The SBFS was approved by the Research Ethics Committee for Scotland. Written informed consent was acquired from all participants. In addition, most participants provided written consent for acquisition of electronic health record linkage data in case of dropout, but this data was not yet available for analysis.

MRI Data Acquisition and Pre-processing

Time1 and time2 MRI sessions were carried out on a 1.5 T Signa scanner (GE Medical, Milwaukee, USA) at the Brain Research Imaging Centre in Edinburgh. The scan protocol included a structural T1 weighted sequence that yielded 180 contiguous 1.2 mm coronal slices (matrix = 192 x 192; fov = 24 cm; flip angle 8°). MRI sessions at time4 were carried out on a different scanner with 3 T magnetic field strength and were therefore not considered within this study.

Pre-processing of T1 weighted scans was done in Statistical Parametric Mapping (SPM) version 12. Spatial registration to a reference brain and segmentation of grey matter was performed with the Computational Anatomy Toolbox (CAT) toolbox (version CAT12.3 (r1318); Gaser & Dahnke, 2018), which runs on SPM12 software (see supplemental method for details). Insufficiently segmented scans were excluded (see supplemental method). Grey

matter maps (GMM) were subsequently smoothed with a Gaussian kernel (FWHM = 8 mm). After loading the smoothed GMM into Python version 3.5.4, voxels were resampled into voxels of double the original voxel size, i.e. 3 x 3 x 3 mm^3. As images had already been smoothed, this resulted in a reduction of features without much loss of information. To ensure that voxels outside the brain were represented by value zero, GMM were masked with a threshold of 0.01.

BrainAGE model

Training sample. The training sample for the brain age prediction model (n = 171) was based on all control participants and HR-participants that remained well (C-well and HR-well; see 'comparison of brain maturation trajectories'). A model just including control participants was underpowered and did not reached sufficient quality of predictions (see supplemental method, Table S3), and since the focus of the study was to investigate whether brain maturation is associated with the development of mood disorders, we considered that on balance it was best to maximise the training sample by combining the 'well' groups in order to develop a sufficiently accurate prediction model. The training sample was equally balanced across time1 and time2 measurements, but in such way that only one MRI measurement per participant was included. Per group (C-well and HR-well), we included time2 measurements for 50% of the time1 sample. Individuals with the highest age at age time2 were selected in order to maximise the age range, and thus model accuracy; this resulted in an age range of 15.21-28.07 years with a mean age of 22.35 years (SD = 3.00).

Brain age prediction. The brain age prediction model was implemented in Python version 3.5.4 and trained on the GMM of the training sample. Within the brain age prediction model, the number of features (originally all voxels within a GMM) was reduced with Principal Component Analysis (PCA) based on Singular Value Decomposition (SVD) without scaling. Following the Kaiser criterion of retaining principal components with eigenvalue

greater than one, 73 orthogonal components were retained (see supplemental method for details on the components). Subsequently, an RVR model was trained on the rotated training data with the use of the scikit-rvm package developed by James Ritchie (available at: https://github.com/JamesRitchie/scikit-rvm/archive/master.zip). The use of a linear kernel was preferred within our sample as the use of a radial basis function (rbf) kernel did not result in a significant correlation between brain age prediction and chronological age (see supplemental method, Table S3). The PCA rotation and RVR model based on the training sample were applied to all time1 and time2 GMM to predict each participants' brain age per measurement. Leave-one-out training was applied in order to prevent bias in the brain age prediction of participants that were part of the training sample. That is, we trained separate models that excluded the measurement corresponding to that participant from the training sample for whom brain age was being predicted.

Calculation of BrainAGE. Further analyses were implemented in R (version 3.2.3). A residuals approach was used to calculate BrainAGE, i.e. the gap between brain age prediction and chronological age. This approach takes the residuals of regressing brain age prediction on chronological age within the training sample (see supplemental method for details), and is standardly used to derive other measures of accelerated ageing, for example in the field of epigenetic ageing (Chen et al., 2016; Horvath, 2013). To account for plausible sex differences in brain maturation and achieve more accurate BrainAGE predictions, sex was additionally included in our regression model. Figure 1 shows a scatterplot of training sample observations, including the regression lines for either sex. Resulting BrainAGE predictions indicated the gap between the brain age prediction and chronological age for each individual at each timepoint; a positive BrainAGE would reflect a relatively maturated brain, while a negative BrainAGE would reflect a delay in brain maturation. Changes in BrainAGE over time indicate a relative acceleration in brain maturation if BrainAGE becomes more positive

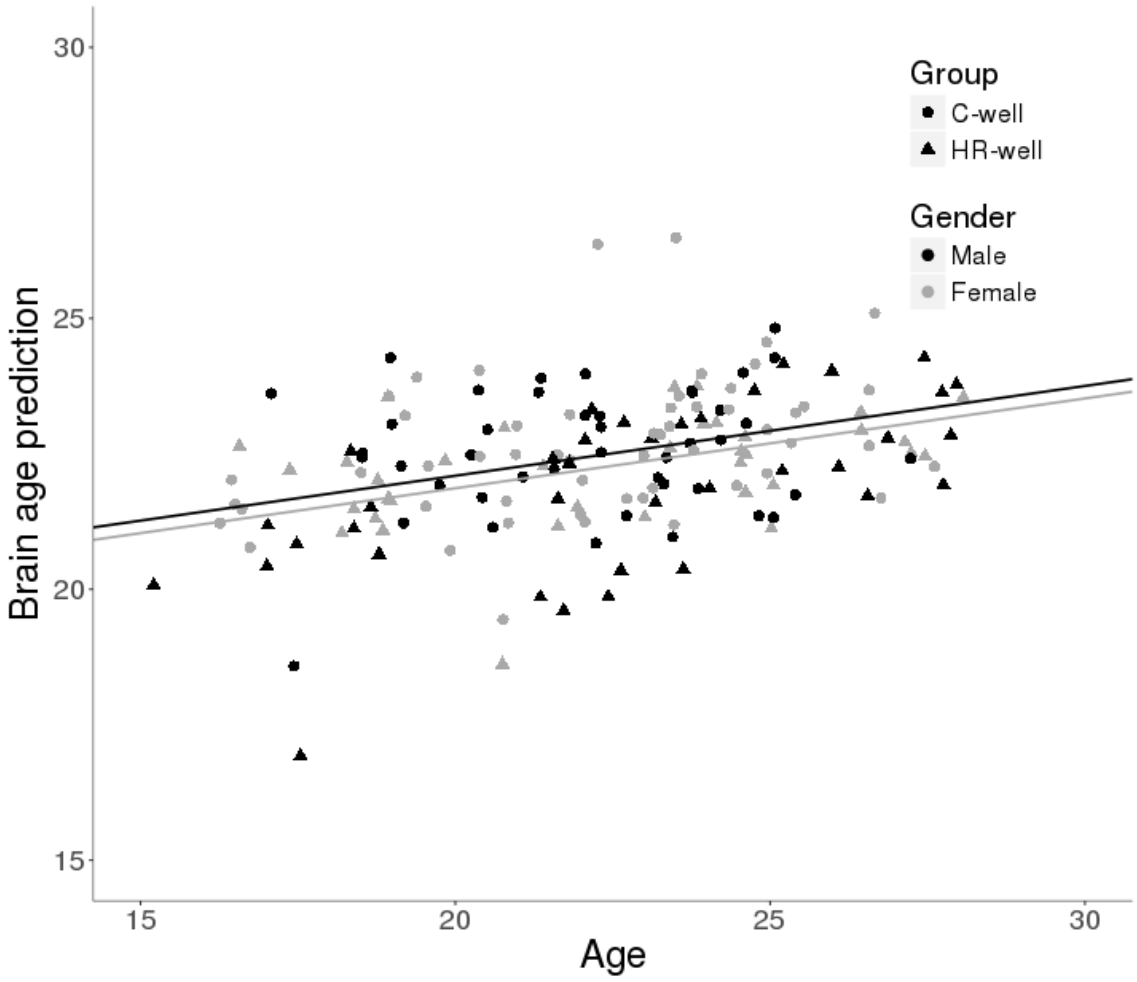


Figure 1. Scatterplot for the training sample that shows brain age prediction against chronological age. The solid black line graph (formula: 18.78 + 0.17 * chronological age) and grey line graph (formula: 18.55 + 0.17 * chronological age) represent the regression lines for respectively males and females that were used to determine the "observed brain age" for each participant. The result was subsequently subtracted from the brain age prediction to calculate brain age gap estimate (BrainAGE).

(or less negative), or a relative deceleration in brain maturation if BrainAGE becomes more negative (or less positive).

Model evaluation. Due to the rarity of such prospective adolescent data, no independent dataset was available for optimal evaluation of the BrainAGE model. Therefore, the BrainAGE model was evaluated based on brain age predictions for the training sample of

individuals who remained well. Ideally, a good BrainAGE model would have three important characteristics: (i) A positive relationship between predicted brain age and chronological age as indicated by a high positive correlation (and thus a high explained variance R^2), (ii) Accurate brain age predictions, indicated by a small Mean Absolute Error (*MAE*) for brain age predictions, and (iii) No systematic over- or underestimation of brain age predictions as indicated by a mean brain age prediction close to the mean of chronological age. As previous BrainAGE research shows a wide range of methodologies, alternative models were explored and compared with the implemented model (see supplemental method, Table S3, Table S4). Results show that only a Simple Linear Regression model was roughly equivalent to the current model (Table S4), and order to establishing the stability of results across a different methodology, this model was additionally implemented as explorative analysis.

Comparison of brain maturation trajectories

As the objective of this study was to investigate deviation of brain maturation trajectories in adolescents that developed a mood disorder, participants were divided in four groups based on clinical information from all four assessments (Table 1). The C-well and HR-well group consisted respectively of control and HR-participants that remained well. Clinical information revealed that 38 HR-participants and 13 control participants developed a mood disorder (either MDD or BD) at some point during the study; these participants constituted respectively the HR-MD and C-MD group. Additional analyses that considered one heterogenous group consisting of all participants who developed a mood disorder – regardless whether at high familial risk or not – were also completed.

A mixed effects model was applied to the BrainAGE results in order to compare brain maturation trajectories between groups, as this type of model is able to effectively deal with missing values as a consequence of dropout (Gueorguieva & Krystal, 2004). The fixed effects were referenced by the control group time1 measurements. In order to choose an appropriate

model, inclusion of a random slope and intercept were both considered, and the best model selected by the Bayesian Information Criterion (BIC; Schwarz, 1978) was the model including a random intercept per subject. Within this model, the random effects on BrainAGE were characterised by an intercept standard deviation of 0.92 and a residual standard deviation of 0.69. Testing of model assumptions revealed that model residuals showed a slight deviation of normality as indicated by a small upward linear trend within the residuals plot. This may be a consequence of the relatively small sample size with restricted age range and our focussed recruitments method, as the current study was not population-based. Furthermore, group differences at time2 were explored with independent t-tests, of which p-values were adjusted using false discovery rate (FDR). These analyses facilitate the comparison of findings with results of previous cross-sectional studies.

Results

Model evaluation

The Pearson correlation between brain age predictions and chronological age within the training sample was significantly positive with $r = 0.38$ ($p < 0.001$), albeit the explained variance (R^2) of 14.4% was low. A mean absolute error (MAE) of 2.27 years indicated reasonably accurate brain age predictions. Furthermore, a comparison of the mean brain age prediction ($M = 22.38$) and mean chronological age ($M = 22.35$) within the training sample shows no systematic under- or overestimation of brain age in our model, but the variation in chronological age ($SD = 2.89$) was higher than the variation in predicted brain age ($SD = 1.23$). In conclusion, the implemented brain age prediction model showed sufficient performance with regard to the set quality criteria (see supplemental results for a more thorough discussion).

Comparison of brain maturation trajectories

Comparison at baseline. Application of the brain age prediction model and BrainAGE residuals approach resulted at baseline in mean BrainAGEs of 0.06 (*SD* = 1.27) for C-well, -0.32 (*SD* = 1.11) for HR-well, 0.41 (*SD* = 0.78) for C-MD, and -0.36 (*SD* = 1.44) for HR-MD (Figure 2). The fixed effects of our mixed effects model (Table 2) showed a significant difference in BrainAGE at baseline for HR-well participants as compared to C-well participants which corresponded to a relative delay in brain maturation of 0.38 years on average for those at familial risk who remained well. In contrast, the delay in brain

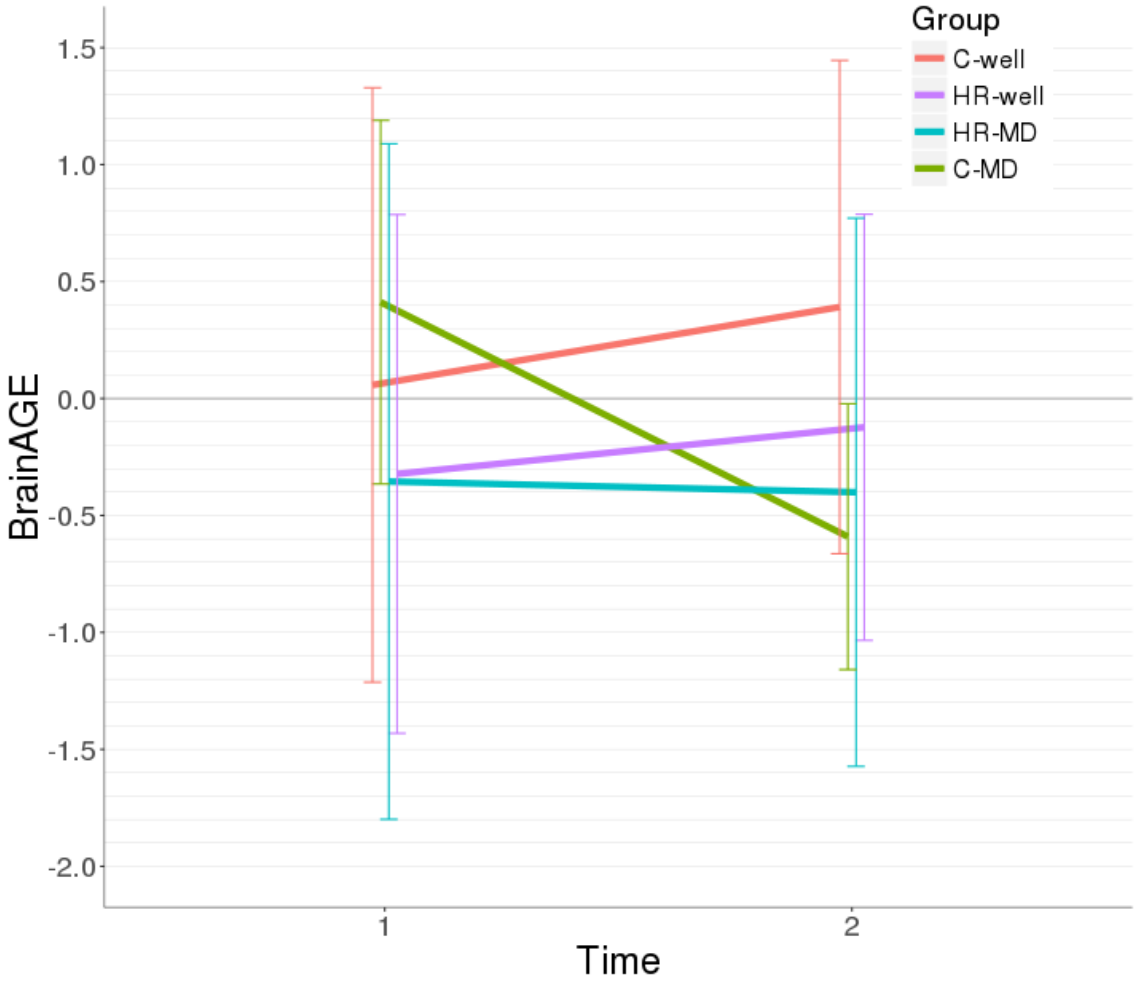


Figure 2. Group means of the brain age gap estimate (BrainAGE). The error bars represent standard deviations.

Table 2

Fixed effects of the mixed effects model applied to predict BrainAGE.

Fixed effect	Value of β-coefficient	Standard Error	*Df*	t-value	*p*-value
(Intercept)	0.06	0.11	218	0.49	.62
Time2	0.36	0.13	134	2.77	.007**
HR-well	-0.38	0.18	218	-2.13	.03*
HR-MD	-0.41	0.22	218	-1.87	.06
C-MD	0.35	0.34	218	1.04	.30
Time2*HR-well	-0.18	0.19	134	-0.97	.33
Time2*HR-MD	-0.39	0.21	134	-1.84	.07
Time2*C-MD	-1.33	0.38	134	-3.53	< .001***

** p < .05, ** p < .01, *** p < .001*

maturation at baseline of 0.41 years for HR-MD was not significant, although we highlight the relatively smaller sample size and greater heterogeneity in BrainAGE within this group (i.e., higher standard deviation) as compared to HR-well. Group means also show a positive BrainAGE at baseline within C-MD that was not significant when compared to C-well.

Brain maturation trajectories. Average brain maturation trajectories per group, indicated by mean change in BrainAGE over time, are displayed in Figure 2. Additionally, Figure 3 shows the brain maturation trajectories per participant (see also Figure S3 in supplemental results). The mixed effect model results (see Table 2) showed a significant which indicated an unexpected increase in BrainAGE over time of 0.36 years on average within the control group, and the mean BrainAGE trajectory for the HR-well group did not significantly differ from this control group brain maturation trajectory. In contrast, as shown by the time2*C-MD interaction effect, the route to mood disorder onset within C-MD was characterised by a deceleration in brain maturation of 1.33 years on average between baseline

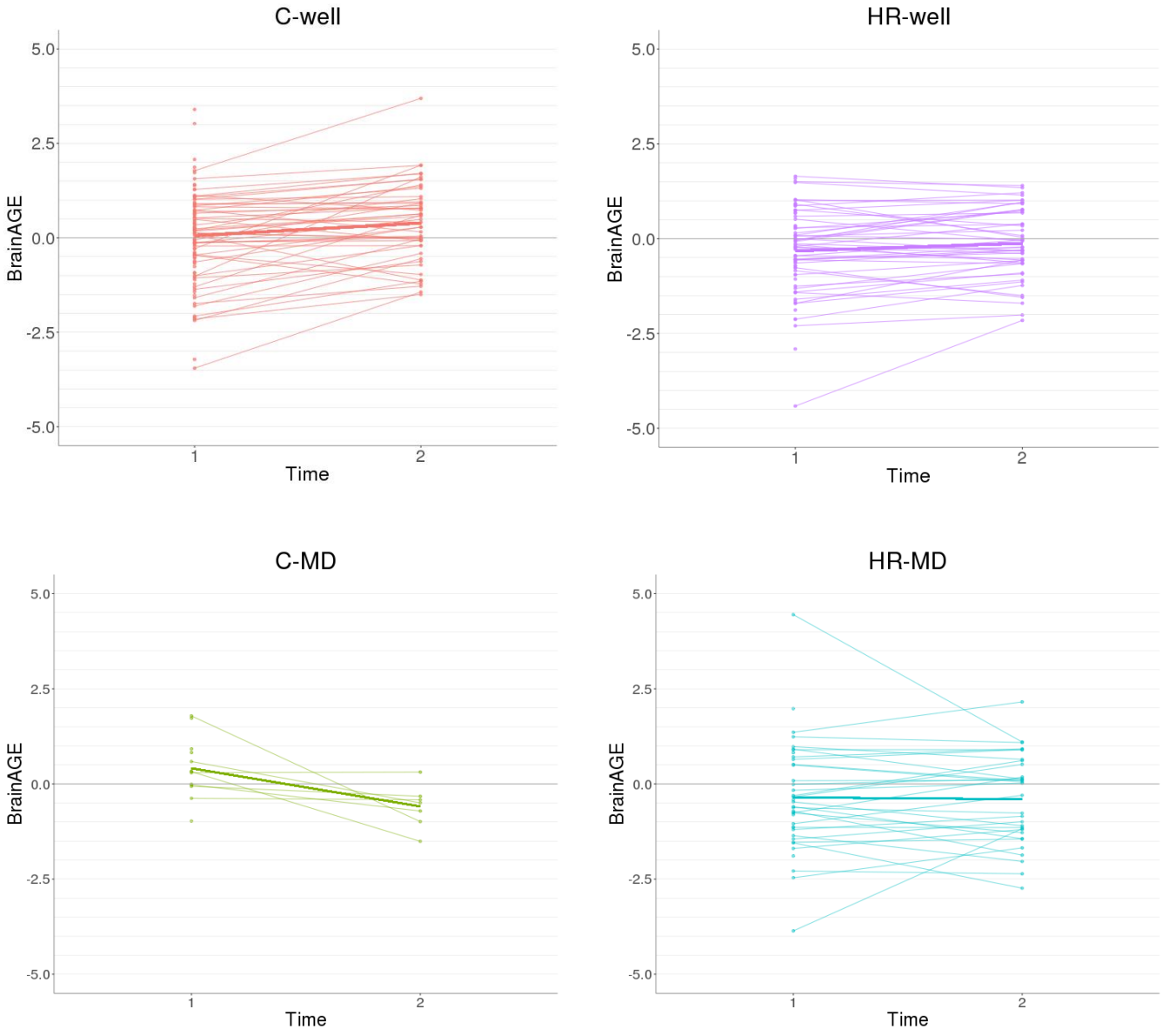


Figure 3. Brain maturation trajectories per participant displayed per group, as indicated by a changing brain age gap estimate (BrainAGE) between time1 and time2, two years apart. Thicker lines represent group mean maturation trajectories. The upper left and right panel show participants who remained well who were without familial risk for mood disorder (C-well) or at high familial risk for mood disorder (HR-well), respectively. The lower panels show participants who developed a mood disorder; the left lower panel displays brain maturation trajectories for participants without familial risk for mood disorder who became ill (C-MD), while the right lower panel displays brain maturation trajectories for those at familial risk who developed a mood disorder (HR-MD).

and follow-up, relative to C-well. HR-MD showed a non-significant deceleration in brain maturation of 0.39 years as compared to C-well.

Comparison at follow-up. At follow-up, two years later, the mean BrainAGEs were 0.39 (*SD* = 1.05) for C-well, -0.12 (*SD* = 0.91) for HR-well, -0.59 (*SD* = 0.57) for C-MD, and -0.40 (*SD* = 1.17) for HR-MD (Figure 2). As HR-well showed a brain maturation trajectory similar to C-well, their initial delay in brain maturation relative to C-well remained at follow-up with a size of 0.41 years ($t(96.3) = 2.6$, $p = .02$). For C-MD, the deceleration in brain maturation resulted in the development of a significant delay in brain maturation of 0.98 years on average at follow-up, relative to C-well ($t(12.7) = 3.8$, $p = .008$). Similarly, although deceleration of brain maturation was not significant within HR-MD, this group nevertheless showed a significant delay in maturation of 0.79 years on average at follow-up as compared to C-well ($t(60.8) = 3.1$, $p = .008$). The delay in brain maturation at follow-up did not significantly differ between HR-well and HR-MD ($t(55.1) = 1.13$, $p = .32$), HR-well and C-MD ($t(11.1) = 1.9$, $p = .13$), or HR-MD and C-MD ($t(19.1) = 0.6$, $p = .53$).

Alternative analyses. Application of a brain age prediction model based on Simple Linear Regression yielded similar results, but with a significant time2*HR-MD inteaction effect ($\beta = -0.71$, $p = .004$) and no significant time2*C-MD interaction effect ($\beta = -0.45$, $p = .28$; see Table S5 and Figure S4 in supplemental results). The results of analyses that included HR-MD and C-MD participants within one heterogeneous group of participants who developed mood disorder (MD) also showed an overall deceleration in brain maturation as indicated by a significant time2*MD interaction effect ($\beta = -0.58$, $p = .005$; see Table S6 and Figure S5 in supplemental results).

Discussion

The results of the study showed an initial delay in brain maturation within young individuals at high familial risk for mood disorders who remained well, as compared to those

with no family history of mood disorders who remained well. As brain maturation accelerated comparably across both groups of individuals who remained well, the initial delay in adolescent brain maturation remained constant over time. In contrast, our findings indicate no deviating brain maturation status at the start of the study for those who were initially well and without family history of mood disorders, but who nevertheless developed a mood disorder over time. Instead, their brain maturation decelerated significantly over time demonstrating a slower pace of brain development, which had resulted in a cross-sectional delay in brain maturation two years later. Furthermore, the results show considerable within-group heterogeneity in brain maturation trajectories which was largest within the group of individuals who were at high familial risk and who also developed a mood disorder, so that within this group mean brain maturation trajectories were characterised by a non-significant initial delay in brain maturation in combination with a non-significant deceleration in brain maturation; at follow-up two years later, this had resulted in a more negative average BrainAGE indicating a significant delay in maturation. For young individuals at high familial risk who become ill, it may thus be that the development of mood disorders is only associated with a deceleration in brain maturation within a subset of individuals. Importantly, cross-sectional results at follow-up indicated no differences in the size of the brain maturation delay between individuals who were at high risk and those who developed a mood disorder, meaning that the presence of a brain age gap was not specifically associated with illness. In conclusion, our findings suggest that familial risk for mood disorder is reflected by an initial delay in brain maturation that remains constant over time for those that remain well, while the development of a mood disorder may be related to the emergence of a lag maturational trajectory on route to illness.

Hajek et al. (2017) had previously investigated the same association between adolescent brain maturation and mood disorders cross-sectionally. Although they did not find any

significant differences, their results did indicate a possible trend in the same direction: a negative mean BrainAGE both within participants at high familiar risk who remained well, and within those that developed a mood disorder. On the other hand, our findings contrast the accelerated ageing hypothesis for mood disorders (Sibille, 2013; Wolkowitz et al., 2011), which was corroborated by a study on BrainAGE that predominantly included adult participants (Koutsouleris et al., 2014).

The strength of the current study is that the SBFS data allowed us to investigate adolescent brain maturation longitudinally over an interval of two years, thus revealing the dynamics of adolescent brain maturation trajectories. That development of a mood disorder was found to be associated with decelerated brain maturation could not have been identified within a cross-sectional design. Furthermore, clinical information extended this two-year period, as it was available for up to six years depending on continual of study participation. Such an extended clinical follow-up is desired to for reliable group categorisation. However, one should take into account that it also leads to heterogeneity within the mood disorder groups, because some individuals had already experienced mood disorder onset at follow-up while others would only experience it years later (with registration of the mood disorder at later clinical assessments). Moreover, as participants were initially well and matched on age and sex, our results are more likely to reflect disease processes related to the onset of mood disorders – although the possibility that our results partly reflect early disease related change or drug treatment effects cannot be excluded. Taken together, these study characteristics show the relative uniqueness of the SBFS sample within adolescent mental health research.

However, due to practical limitations that come with such a research design the sample size of the SBFS was relatively modest for the application of BrainAGE methods, which in turn resulted in certain important limitations. Firstly, one needs to take into consideration that the model was trained on all individuals who remained well including those at high familial

risk for mood disorder. BrainAGE slightly increased over time within the control group, which indicates the bias in our model that typical brain maturation would be characterised by the emergence of a small brain age gap. Although a BrainAGE of zero did not indicate typical brain maturation, this bias is unlikely to affect the investigation of the relative difference in BrainAGE between groups. Another limitation is that the brain age model predictions showed a relatively low explained variance with regards to chronological age, but the adoption of the BrainAGE residuals approach accounted for this suboptimal relationship and prevented the bias of a correlation between BrainAGE and chronological age. Although we did not have the possibility to validate our BrainAGE model within an independent sample, dimensionality reduction and sparse modelling via RVR were applied in an attempt to prevent overfitting.

Future research should aim to replicate the current results within a larger sample, as study with increased sample size will not suffer from the above limitations, enhance the performance of the brain age prediction model, and yield more reliable results for group comparisons (Button et al., 2013). Furthermore, it would allow investigation of similarities and differences between MDD and BD by considering these separately. More research is needed to establish the precise relationship between mood disorders and brain maturation, as well as biological and psychological mechanisms underlying this relationship.

Ineffective emotion regulation is one mechanism plausibly related to the development of mood disorders within adolescent, because it has been associated to subsequent depressive symptoms within multiple longitudinal studies (for a meta-analytic review see: Aldao, Nolen-Hoeksema & Schweizer, 2010), and has therefore been implemented in many models of psychopathology (e.g., Hofmann, Sawyer, Fang & Asnaani, 2012; Phillips, Ladouceur & Drevets, 2008; Nolen-Hoeksema, Wisco & Lyubomirsku, 2008). As adolescents and young adults age they increasingly need to regulate their behaviour and affect in order to achieve long-term goals, so that insufficient regulatory capacities puts them at risk for the

development of a mood disorder (Steinberg, 2005). In light of this, it might be that deceleration in adolescent brain maturation generates an imbalance in the brain in which emotion regulation and cognitive control functions delay in their development. This delay may be reflected by anatomical immaturity within brain regions related to these functions – especially within large regions of the prefrontal cortex (Philips et al., 2008). To investigate this hypothesis further, one could implement the BrainAGE methodology within a large sample to produce BrainAGEs based on the prefrontal cortex, and see if these are associated with the development of a mood disorder and/or (sub)clinical depressive symptoms.

Alternatively, one could investigate what areas of grey matter are influential for the brain age prediction by use of orthonormal projective non-negative matrix factorisation (OPNMF) of grey matter (Soritas, Resnick & Davatzikos, 2015), a method which produces components that represent biologically meaningful clusters within the brain (Soritas et al., 2017). Subsequent exploration of OPNMF components contributing to the brain age prediction can give insight into structural changes related to brain ageing, thus increasing model interpretability. Previously, Varikuti and colleagues (2018) have successfully implemented OPNMF within a brain age prediction model for a cohort of adults and elderly. Similarly, one could implement OPNMF to identify crucial regions for the prediction of brain age within young individuals and explore local differences in brain maturation associated with mood disorder development.

Furthermore, as longitudinal clinical research often includes a relatively limited sample size due to practical constraints, it would be useful to explore machine learning techniques that expand resources and thus enhance the reliability of results. Data augmentation methods synthetically increase sample size by the addition of artificial training images that are modifications of original images. Although this does not create authentic information, it does often decrease overfitting. Cole et al. (2017) applied data augmentation for their brain age

predictions by adding rotated and translated MRI images to their training sample, and this was indeed found to increase model performance. However, as Cole et al. (2017) developed a brain age prediction model based on raw T1-weighted images, the validity and utility of data augmentation methods when using normalised and pre-processed images within a BrainAGE framework is yet to be established.

References

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217-237. doi:10.1016/j.cpr.2009.11.004

Arnone, D., McIntosh, A. M., Ebmeier, K. P., Munafò, M. R., & Anderson, I. M. (2011). Magnetic resonance imaging studies in unipolar depression: Systematic review and meta-regression analysis. *European Neuropsychopharmacology, 22*, 1-16. doi:10.1016/j.euroneuro.2011.05.003

Bora, E., Fornito, A., Pantelis, C., & Yücel, M. (2012). Gray matter abnormalities in major depressive disorder: A meta-analysis of voxel based morphometry studies. *Journal of Affective Disorders, 138*, 9-18. doi:10.1016/j.jad.2011.03.049

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J. et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365-376.

Chen, B. H., Marioni, R. E., Colicino, E., Peters, M. J., Ward-Caviness, C. K., Tsai, P.-C. et al. (2016). DNA methylation-based measures of biological age: meta-analysis predicting time to death. *Aging*, *8*(9), 1844-1859.

Cole, J. H., Marioni, R. E., Harris, S. E., & Deary, I. J. (2018). Brain age and other bodily 'ages': implicaions for neuropsychiatry. *Molecular Psychiatry*. Online publication. doi:10.1038/s41380-018-0098-1.

Cole, J. H., Poudel, R. P. K., Tsagkrasoulis, D., Caan, M. W. A., Stewes, C., Spector, T. D. et al. (2017). Predicting brain age with deep learning from raw imaging data results in a

reliable and heritable biomarker. *NeuroImage, 163,* 115-124. doi:10.1016/j.neuroimage.2017.07.059

de Girolamo, G., Dagani, J., Purcell, R., Cocchi, A., & McGorry, P. D. (2012). Age of onset of mental disorders and use of mental health services: needs opportunities and obstacles. *Epidemiology and Psychiatric Sciences, 21,* 47-57. doi:10.1017/S2045796011000746

First, M. B., Spitzer, .R.L, Gibbon, M., Williams, J. B. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition with Psychotic Screen.* Biometrics Research, New York State Psychiatric Institute: New York.

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage, 50,* 883-892.

Gaser, C. & Dahnke, R. (2018). CAT – A Computational Anatomy Toolbox for SPM [computer software]. Retrieved from http://www.neuro.uni-jena.de/cat/index.html#DOWNLOAD

Giorgio, A., Watkins, K. E., Chadwick, M., James, S., Winmill, L., Douad, G. et al. (2010). Longitudinal changes in grey and white matter during adolescence. *NeuroImage, 49,* 94-103.

Gueorguieva, R., & Krystal, J. H. (2004). Move over anova: progress in analyzing repeated-measures data and its reflection in papers published in the archives of general psychiatry. *Archives of General Psychiatry, 61*(3), 310-317.

Hajek, T., Franke, K., Kolenic, M., Capkova, J., Matejka, M., Propper, L. et al. (In press). Brain age in early stages of bipolar disorders or schizophrenia. *Schizophrenia Bulletin*. doi:10.1093/schbul/sbx172

Hibar, D. P., Westlye, L. ., Doan, N. T., Jahanshad, N., Cheung, J. W., Ching, R. K. et al. (2018). Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular Psychiatry*, *23*, 932-942.

Hofmann, S. G., Sawyer, A. T., Fang, A., & Asnaani, A. (2012). Emotion dysregulation of mood and anxiety disorders. *Depression and Anxiety*, *29*, 409-416.

Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biology*, *14*, R115. doi:10.1186/gb-2013-14-10-r115

Kempton, M. J., Salvador, Z., Munafo, M. R., Geddes, J. R., Simmons, A., Frangou, S. et al. (2011). Structural Neuroimaging studies in major depressive disorder: Meta-analysis and comparison with bipolar disorder. *Archives of General Psychiatry, 68*(7), 675-690.

Kessler, R.C., & Bromet, E.J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, *34*, 119-138.

Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T. et al. (2014). Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders. *Schizophrenia Bulletin*, *40*(5), 1140-1153. doi:10.1093/schbul/sbt142

Luppino, F. S., de Wit, L. M., Bouvy, P. F., Stijnen, T., Cuijpers, P., Penninx, B. W. J. H. et al. (2010). Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies.

Mezuk, B., Eaton, W. W., Albrecht, S., & Golden, S. H. (2008). Depression and type 2 diabetes over the lifespan: A meta-analysis. *Diabetes Care*, *31*(12), 2383-2390.

Nenadić, I., Dietzek, M., Langbein, K., Sauer., H., & Gaser, C. (2017). BrainAGE score indicates accelerated brain aging in schizophrenia, but not in bipolar disorder. *Psychiatry Research: NeuroImaging*, *266*, 86-89. doi:10.1016/j.pscychresns.2017.05.006

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking Rumination. *Perspectives on Psychological Science*, *3*(5), 400-424.

Ösby, U., Brandt, L., Correia, N., Ekbom, A., & Sparén, P. (2001). Excess mortality in bipolar and unipolar disorder in Sweden. *Archives of General Psychiatry*, *58*, 844-850.

Pan, A., Sun, Q., Okereke, O. I., Rexrode, K. M., & Hu, F. B. (2011). Depression and risk of stroke morbidity and mortality: A meta-analysis and systematic review. *Journal of the American Medical Association*, *306*(11), 1241-1249. doi:10.1001/jama.2011.1282

Papmeyer, M., Sussmann, J. E., Hall, J., McKirdy, J., Peel, A., Macdonald, A. et al. (2015). Neurocognition in individuals at high familial risk of mood disorders with or without subsequent onset of depression. *Psychological Medicine*, *45*, 3317-3327. doi:10.1017/S0033291715001324

Phillips, M. L., Ladouceur, C. D., & Drevets, W. C. (2008). A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Molecular Psychiatry*, *13*, 833-857.

Regenold, W. T., Thapar, R. T., Marano, C., Gavirneni, S., & Kondapavuluru, P. V. (2002). Increased prevalence of type 2 diabetes mellitus among psychiatric inpatients with bipolar I affective and schizoaffective disorders independent of psychotropic drug use. *Journal of Affective Disorders*, *70*(1), 19-26. doi:10.1016/S0165-0327(01)00456-6

Ridout, K. K., Ridout, S. J., Price, L. H., Sen, S., & Tyrka, A. R. (2016). Depression and telomere length: A meta-analysis. *Journal of Affective Disorders, 191*, 237-247. doi:10.1016/j.jad.2015.11.052

Rizzo, L. B., Costa, L. G., Rodrigo, B. M., Swardfager, W., lole Belangero, S., Grassi-Oliveira, R. et al. (2014). The theory of bipolar disorder as an illness of accelerated aging: Implications for clinical care and research. *Neuroscience & Biobehavioral Reviews, 42*, 157-169.

Sanders, J. L., & Newman, A. B. (2013). Telomere length in epidemiology: A biomarker of aging, age-related disease, both, or neither? *Epidemiologic Reviews, 35*(1), 112-131.

Scahill, R. I., Frost, C., Jenkins, R., Whitwell, J. L., Rosssor, M. N., & Fox, N. C. (2003). A longitudinal study of brain volume changes in normal aging using serial registered Magnetic Resonance Imaging. *Archives of Neurology, 60*, 989-994.

Schmaal, L., Hibar, D. P., Sämann, P. G., Hall, G. B., Baune, B. T., Jahanshad, N. et al. (2017). Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder working group. *Molecular Psychiatry, 22*, 900-909.

Schmaal, L., Veltman, D. J., van Erp, T. G. M., Sämann, P. G., Frodl, T., Jahanshad, N. et al. (2016). Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Molecular Psychiatry, 21*, 806-812.

Schulz, R., Beach, S. R., Ives, D. G., Martire, L. M., Ariyo, A. A., & Kop, W. J. (2000). Association between depression and mortality in older adults: The cardiovascular health study. *Archives of Internal Medicine, 160*, 1761-1768.

Schutte, N. S., & Malouf, J. M. (2015). The association between depression and leukocyte telomere length: A meta-analysis. *Depression and Anxiety, 32,* 229-238. doi:10.1002/da.22351

Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Sibille, E. (2013) Molecular aging of the brain, neuroplasticity, and vulnerability to depression and other brain-related disorders. *Dialogues in Clinical Neuroscience*, *15*(1), 53-65.

Simon, N. M., Smoller, J. W., McNamara, K. L., Maser R. S., Zalta, A., Pollack, M. H. et al. (2006). Telemore shortening and mood disorders: Preliminary support for a chronic stress model of accelerated ageing. *Biological Psychiatry*, *60*(5), 432-435.

Smoller, J. W., & Finn, C. T. (2003). Family, twin, and adoption studies of bipolar disorder. *American Journal of Medical Genetics Part C*: *Seminars in Medical Genetics*, *123C*, 48–58.

Soritas, A., Resnick, S. M., & Davatzikos, C. (2015). Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *NeuroImage, 108,* 1-16. doi:10.1016/j.neuroimage.2014.11.045

Soritas, A., Toledo, J. B., Gur, R. E., Gur, R. C., Satterthwaite, T. D., & Davatzikos, C. (2017). Patterns of coordinated cortical remodeling during adolescence and their associations with functional specialization and evolutionary expansion. *Proceedings of the National Academy of Sciences of the United States of America, 114*(13), 3527-3532. doi:10.1073/pnas.1620928114

Sprooten, E., Sussmann, J. E., Clugston, A., Peel, A., McKirdy, J., Moorhead, T. W. et al. (2011). White matter integrity in individuals at high genetic risk of bipolar disorder. *Biological Psychiatry*, *70*, 350–356.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*(2), 69-74.

Tamnes, C. K., Østby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2010). Brain maturation in adolescence and young adulthood: Regional age-related changes in cortical thickness and white matter volume and microstructure. *Cerebral Cortex*, *20*, 534-548. doi:10.1093/cercor/bhp118

Varikuti, D. P., Genon, S., Sotiras, A., Schwender, H., Hoffstaedter, F., Patil, K. R. et al. (2018). Evaluation of non-negative matrix factorization of grey matter in age prediction. *NeuroImage*, *173*, 394-410. doi:10.1016/j.neuroimage.2018.03.007

Whalley, H. C., Sussmann, J. E., Romaniuk, L., Stewart, T., Kielty, S., Lawrie, S. M. et al. (2011). Dysfunction of emotional brain systems in individuals at high risk of mood disorder with depression and predictive features prior to illness. *Psychological Medicine*, *45*, 1207-1218. doi:10.1017/S0033291714002256

Whittle, S., Lichter, R., Dennison, M., Vijayakumar, N., Schwartz, O., Byrne, M. L. et al. (2014). Structural brain development and depression onset during adolescence: A prospective longitudinal study. *American Journal of Psychiatry*, *171*, 564-571.

Whooley, M. A., de Jonge, P., Vittinghoff, E., Otte, C., Moos, R., Carney, R. M. et al. (2008). Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. *Journal of the American Medical Association*, *300*(20), 2379-2388.

Wise, T., Radua, J., Via, E., Cardoner, N., Abe, O., Adams, T. M. et al. (2017). Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: evidence from voxel-based meta-analysis. *Molecular Psychiatry, 22*, 1455-1463. doi:10.1038/mp.2016.72

Wolkowitz, O. M., Reus, V. I., & Mellon, S. H. (2011). Of sound mind and body: depression, disease, and accelerated aging. *Dialogues in Clinical Neuroscience*, *13*(1), 25-39.

World Health Organization (2017). *Depression and other common mental disorders: Global health estimates.* Geneva: WHO Press.

Supplemental Materials

CONTENTS

Supplemental method

Excluded participants

Table S1

Excluded participants.

Rationale for exclusion	Number of excluded participants		
	Control group	HR-group	**Total**
Missing data	17	23	**40**
Insufficient MRI image quality or segmentation quality	4	9	**13**
Other psychiatric disorder or ambiguous diagnosis	1[a]	3[b]	**4**
High risk participant without follow-up	N.A.	9	**9**
Multiple reasons	0	1[c]	**1**
Total	**22**	**45**	**67**

HR group = High Risk group; MRI = Magnetic Resonance Imaging; N.A. = Not Applicable
[a] Obsessive Compulsive Disorder, [b] Low mood confirmed by General Practitioner (2x), single episode of psychosis, [c] No baseline scan, time2 scan of insufficient quality, alcohol dependence.

Cat12 segmentation

The CAT12 toolbox (version CAT12.3 (r1318); Gaser & Dahnke, 2018; www.neuro.uni-jena.de/cat/) was used with default settings to segment the T1-weighted MRI scans into grey matter (GM), white matter (WM) and cerebrospinal fluid (CSF); subsequently, smoothed grey matter maps (GMM) were used as input for the brain age prediction model. With default settings, the CAT12 toolbox applies internal interpolation and Spatial-adaptive Non-Local Means (SANLM; Manjón, Coupe, Marti-Bonmati, Collins & Robles, 2010) as a pre-processing step in order to reduce noise normalisation, before the T1-weighted MRI scans

are normalised to Montreal Neurological Institute (MNI) space using affine and non-linear registration by integration of Diffeomorphic Anatomic Registration Through Exponentiated Lie algebra algorithm (DARTEL; Ashburner, 2007) and Geodesic Shooting Normalisation (Ashburner & Friston, 2011). Segmentation is achieved by implementation of Adaptive Maximum A Posterior (AMAP) segmentation and Partial Volume Segmentation (PVE; Tohka, Zijdenbos & Evans, 2004). As a part of AMAP, Classical Markov Random Field (MRF; Rajapakse, Giedd & Rapoport, 1997) includes spatial information of adjacent voxels in the segmentation estimation.

After segmentation, the software provides a segmentation quality assurance in the form of percentage rating points (range 0-100%) for resolution, noise and bias, and a weighted average of these three components. A score above 70% is considered at least satisfactory. Our quality standards excluded segmentations with weighted average scores below 70% and/or resolution, noise or bias scores below 65% from analysis. Image quality of baseline scans was also ensured by visual inspection and exclusion of raw scans with major artefacts; this resulted in the additional exclusion of one scan (Figure S1).

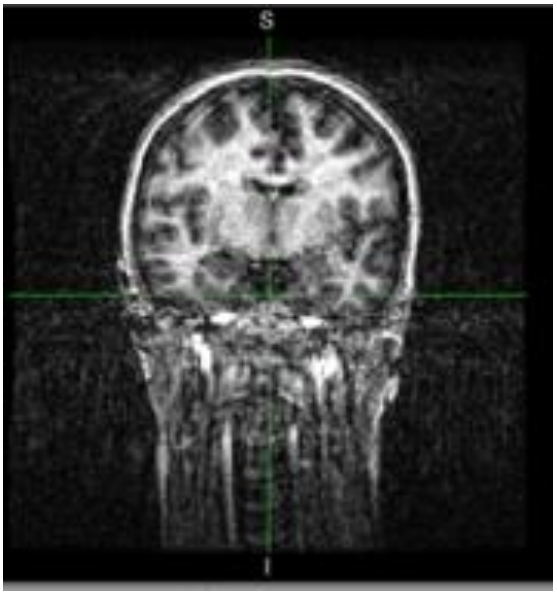


Figure S1. Middle coronal brain slice of the scan that was excluded because of insufficient image quality as determined by visual inspection

Principal Components

Principal Component Analysis with Singular Value Decomposition on the training sample data was further explored within R (version 3.2.3). Principal Components (PC) corresponding to an eigenvalue greater than one were retained (Kaiser criterion), and this resulted in 73 components with a total explained variance of 83.6%. The Pearson correlation between the rotated matrix components and age was significant for five components, none of which remained significant after False Discovery Rate (FDR) correction for multiple testing ($n = 73$, $\alpha = 0.05$). The results are displayed in Table S2.

Voxel loadings on retained components ranged from -0.46 to 0.21. The five components of which the correlation was nominally significant explained 4.3% of the total variance in the brain and 15.2% of the relationship with age prediction; the GMM voxel loadings of these components were further explored. The images in Figure S2 provide an indication of the voxel loadings for each component in one smoothed middle axial GMM slice.

Table S2

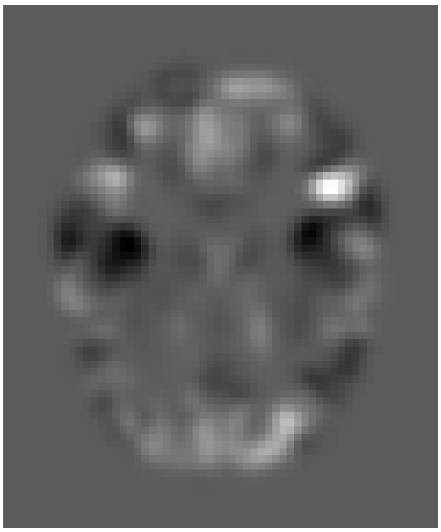
Principal Components within the brain that significantly correlated with brain age predictions.

Principal Component	Pearson correlation *r*	Uncorrected *p*-value
12	-0.16	.02*
13	-0.16	.04*
16	-0.23	.002*
17	0.16	.03*
30	-0.17	.03*

** p < .05, ** p < .01*

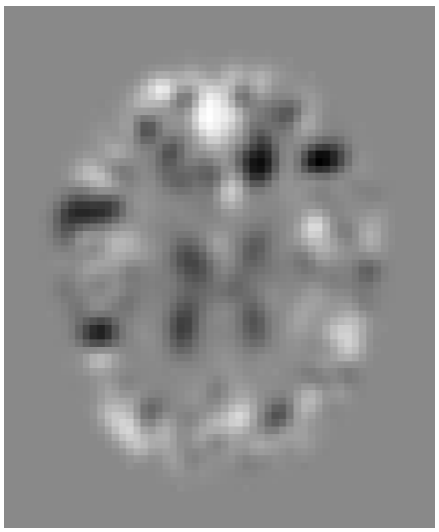
Component 12

Range (-0.017, 0.028)



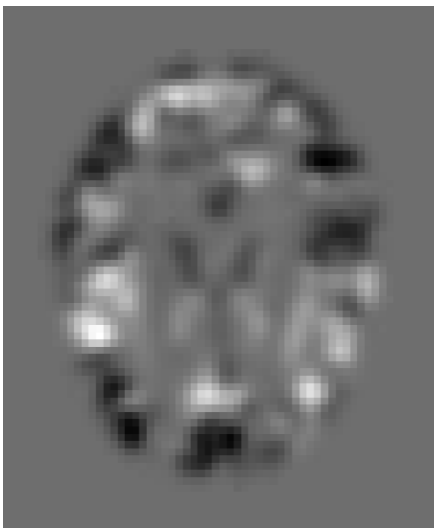
Component 13

Range (-0.017, 0.020)



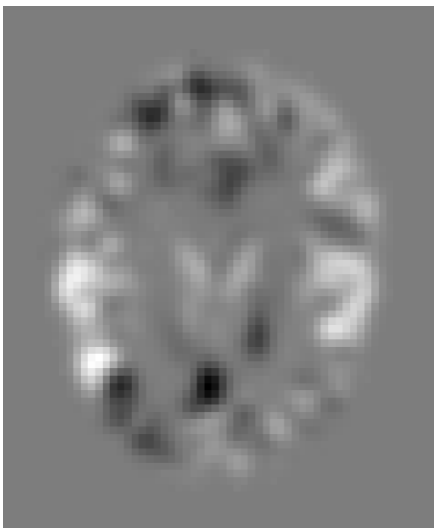
Component 16

Range (-0.017,0.025)



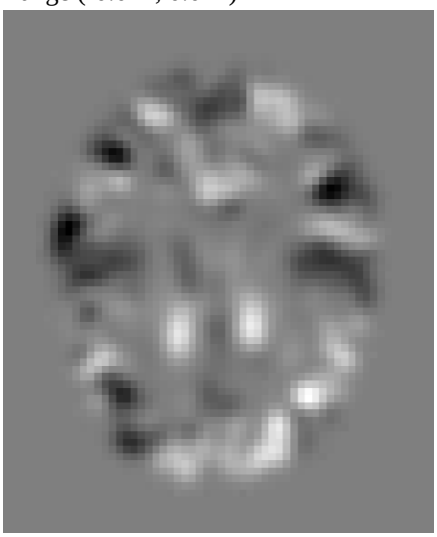
Component 17

Range (-0.024, 0.021)



Component 30

Range (-0.021, 0.017)



GMM of participant

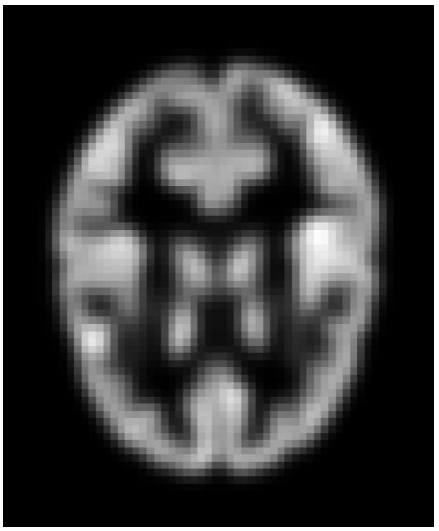


Figure S2. Exploration of voxel loadings for components that significantly correlated with brain age predictions. Within all images, the scale is relative: the background colour represents a value of zero, while the blackest pixel represents the minimum (i.e., most negative) voxel loading value, and a white colour represents the maximum (i.e., most positive) value. The frontal lobe is facing downwards. For reference, the range of voxel loading values is reported with each image, as well as the GMM of one of the participants (lower right picture).

Rationale for current method

The implemented brain age prediction model deviates from the proposed method (in 'Thesis Proposal') on the following aspects:

- The training sample consists on both control participants and HR-well participants, as opposed to solely control participants.
- The training sample measurements are balanced with regard to time1 and time2 MRI scans, as opposed to including solely time1 measurements.
- The RVR model implements a linear kernel, as opposed to a radial basis function (rbf) kernel.

The changes in methodology were necessary to ensure sufficient quality of the brain age prediction model, which is indicated by three important characteristics: (i) A positive relationship between predicted brain age and chronological age as indicated by a high positive correlation (and thus a high explained variance R^2), (ii) Accurate brain age predictions, indicated by relatively small Mean Absolute Error (*MAE*) for brain age predictions, and (iii) No systematic over- or underestimation of brain age predictions as indicated by a mean brain age prediction close to the mean of chronological age (M = 22.35) within the sample.

To illustrate the necessity of the changes in the brain age prediction model described above, we evaluated performance of models on the criteria *without* each of these changes, while keeping all other aspects of the implemented model the same. To ensure unbiased comparison of model performance, criteria of all models were evaluated based on resulting predictions for the eventually implemented training sample. Table S3 displays the model evaluation results.

Table S3

Model evaluation for models based on previously proposed methodology.

Change in methodology	[1] Pearson correlation *r*	[2] *MAE*	[3] Brain age prediction, *M*	Number of PC
Training solely on control sample	0.46***	**22.35**	22.13	45
Training with time1 measurements	0.31***	2.52	**21.44**	74
Rbf kernel	**0.14**	2.44	22.66	73
Implemented model	0.38***	2.27	22.38	73

MAE = Mean Absolute Error; PC = Principal Components; Rbf = radial basis function.

*** $p < .001$

Note. Problematic model characteristics displayed in bold for clarity; these show the necessity for change in methodology, leading up to the currently implemented model. For reference, the last row of the table presents the performance of the model implemented in this study. The number of principal components (PC) selected by the Kaiser criterion is also reported for each model as this differed across models.

Comparison of methods

The current study implements RVR with a linear kernel and PCA within the brain age prediction model. However, several alternative methods might also be appropriate. One alternative model would implement RVR without dimensionality reduction, but only with removal of features (i.e., GMM voxels) that show no variance across all training samples. Furthermore, a simple linear regression model is likely to yield similar results as the current model. For this method, dimensionality reduction must be implemented to avoid multicollinearity (and consequently rank-deficiency). A third alternative method is penalised regression with L1 regularisation, otherwise known as least absolute shrinkage and selection operator (Lasso). This method reaches a sparse solution by penalisation of model coefficients. The λ parameter determines the penalty, and the higher the penalty, the less coefficients are retained within the model. This parameter is normally optimised using cross-validation.

The three alternative models have been implemented in Python version 3.5.4 (for RVR models) or R version 3.2.3 (for Simple Linear Regression and Lasso) with default setting, unless otherwise specified. For Lasso, no cross-validation was applied because of our limited sample size; instead we considered λ values of 0.01, 0.001 and 0.0001. The alternative models were also evaluated according to the specified quality criteria (see 'rationale for current method'), and the results are shown in Table S4.

The results show that our RVR with PCA model performs considerably better than RVR without PCA and Lasso, and comparable to Simple Linear Regression with PCA. One can see that Simple Linear Regression might be favourable; however, these results are explorative as no cross-validation was applied, and therefore the currently used method were not changed accordingly. However, BrainAGE calculations and group comparisons based on a linear regression brain prediction model were investigated exploratively to find out whether the group comparison results are stable across different methodologies.

Table S4

Results of model evaluations for alternative models.

Prediction method	Implementation of PCA?	[1] Pearson correlation *r*	[2] *MAE*	[3] Brain age prediction, *M*
RVR	**Yes**	**0.38*****	**2.27**	**22.38**
RVR	No	0.22**	2.83	22.54
Simple Linear Regression	Yes	0.42***	2.20	22.38
Lasso (λ = 0.01)	No	0.32***	2.43	22.30
Lasso (λ = 0.001)	No	0.34***	2.47	22.31
Lasso (λ = 0.0001)	No	0.02	3.22	22.41

PCA = Principal Component Analysis, MAE = Mean Absolute Error, RVR = Relevance Vector Regression.

*** p < .01, ** p < .001.*

Note. The upper row shows the model performance for the implemented model; for clarity it is displayed in bold.

BrainAGE residuals approach

A residuals approach was applied by calculating BrainAGE based on the residuals from the regression model that regresses brain age prediction on chronological age and sex within the training sample. In other words, we fit a regression model according to the R formula lm(brain age prediction ~ chronological age + sex) on the training sample data which resulted in two regression lines corresponding to each sex (Figure 1, main body). Subsequently, for each participant the "observed brain age" was calculated as the point on the regression line corresponding to the chronological age and sex of the participant. BrainAGE was calculated by subtracting this "observed brain age" from the brain age as predicted by the brain age prediction model. This corresponds to the distance ('residual') of the participant's observation to the training sample regression line within a scatter plot of chronological age against brain age prediction. The BrainAGE residuals approach takes into account the inaccuracies of model predictions by using the observed relationship between brain age prediction and chronological age. Furthermore, BrainAGE calculations are inherently uncorrelated to chronological age, $r(170) = 0.00$, $p = 1$. Therefore, it is considered to be more reliable than a simple subtraction method.

Supplemental Results

Model evaluation: comparison with previous studies

The model implemented within the current study shows a Pearson correlation of 0.38 between brain age prediction and chronological age and an *MAE* of 2.27 years within the training sample. Previous studies with a much broader age range within their sample – including both adolescents and elderly participants – generally achieve a Pearson correlation of around 0.9 and an *MAE* of over 4 years when predicting brain age (Cole et al., 2017; Schnack et al., 2016). Franke et al. (2012) included children and adolescents aged 4 to 18 years and achieved Pearson correlations of >0.9 in combination with *MAEs* ranging between 1.1 and 1.3 years within their samples.

In comparison to brain age prediction models within previous studies, the currently implemented model showed a reasonable *MAE*, but a substantially lower correlation between brain age prediction and chronological age resulting in a low explained variance. However, one must take into account that the current study investigates brain age within a relatively narrow age range, and that it is therefore considerably more difficult to achieve a high correlation and explained variance. While this is also the case for the study of Franke et al. (2012), they investigated brain age during a period in which brain maturation leads to more prominent anatomical changes. Therefore, as the *MAE* is reasonable, we argue that the performance of our brain age prediction model is sufficiently reliable.

Individual brain maturation trajectories (by chronological age)

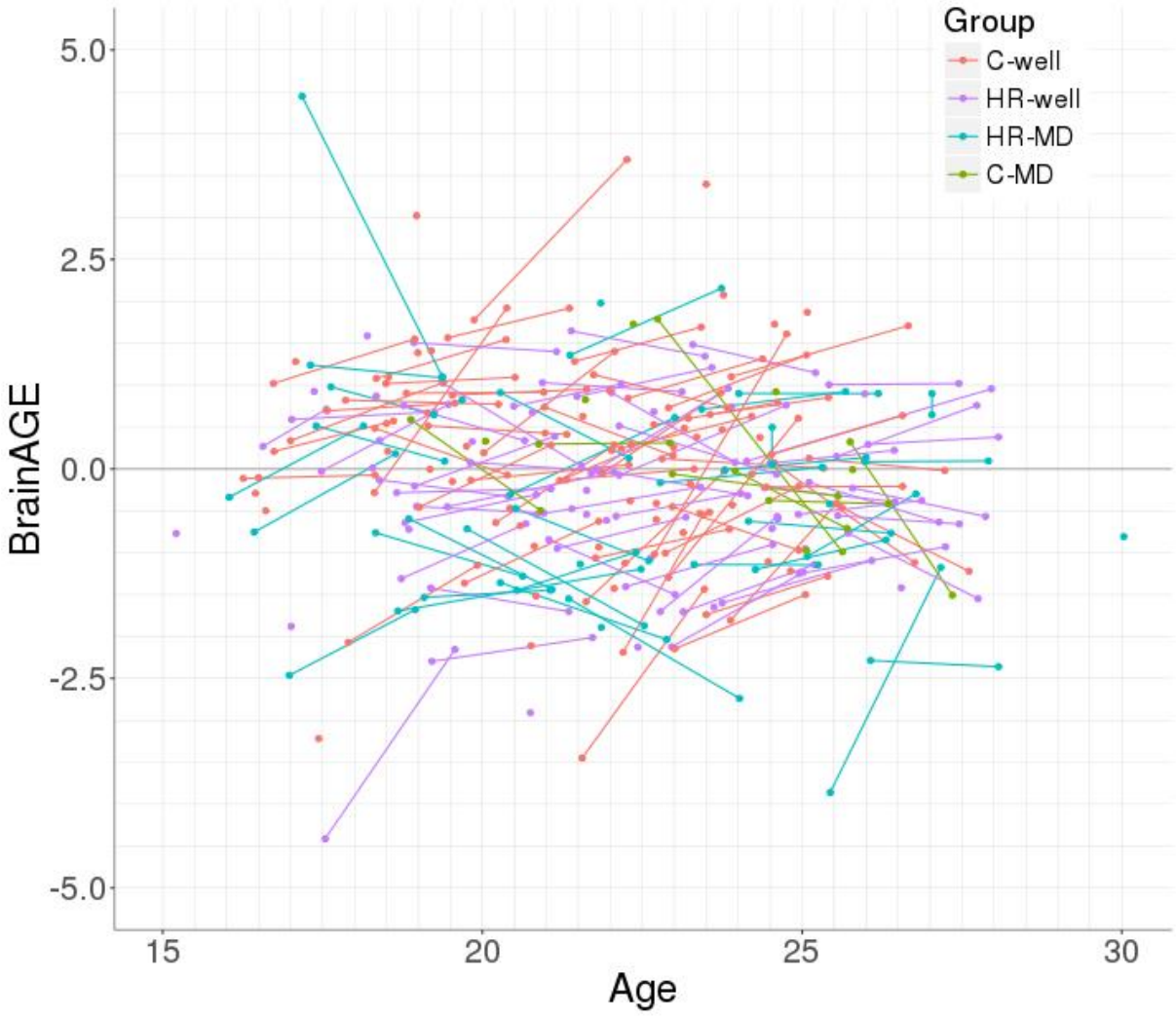


Figure S3. Individual brain maturation trajectories, displayed as a change in brain age gap estimate (BrainAGE) over a change in chronological age. Every point represents a measurement, and two connected points reflect BrainAGE changing over time for one participant.

Results with Simple Linear Regression model

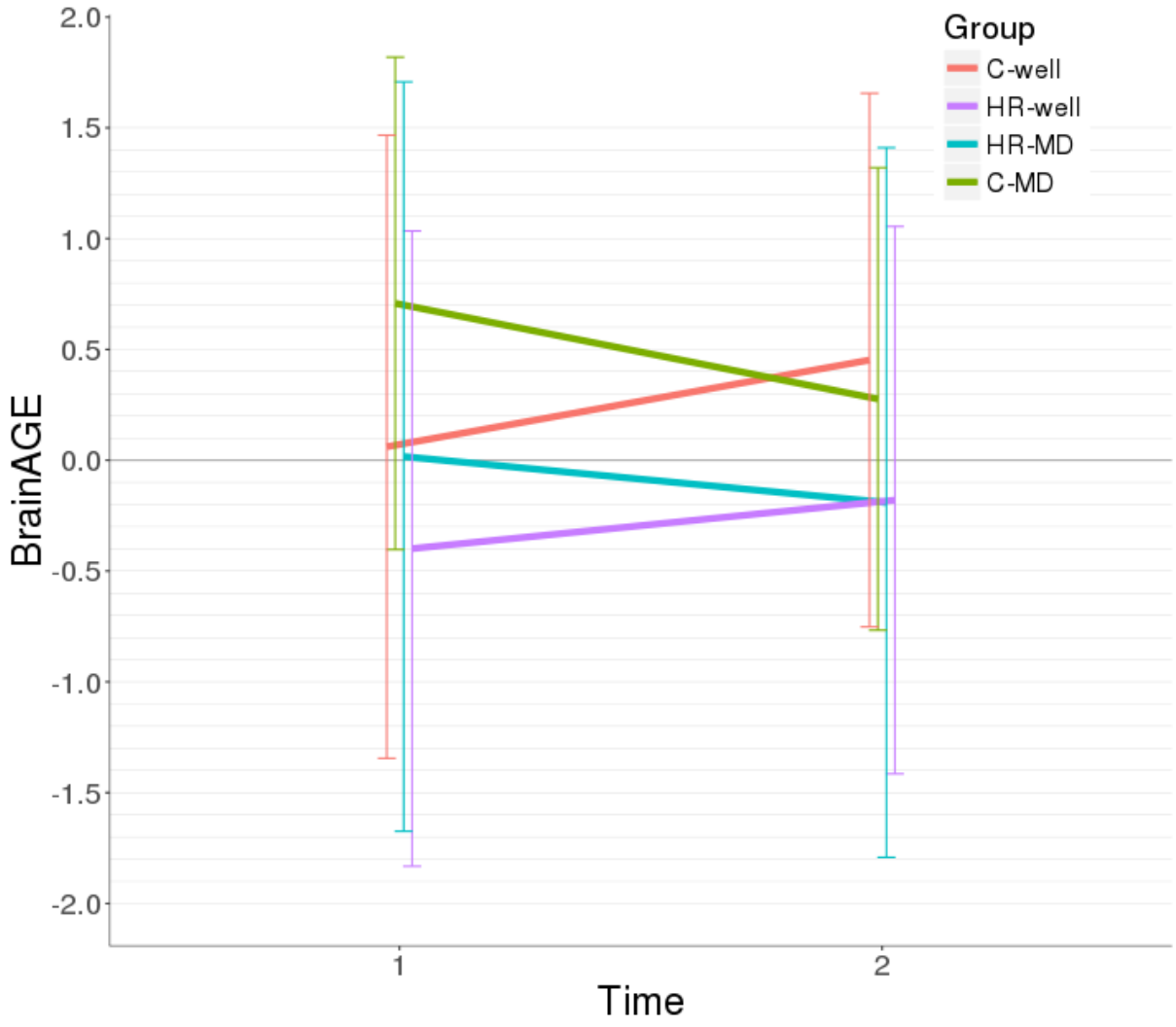


Figure S4. Group means of the brain age gap estimate (BrainAGE), when brain age predictions are obtained with a Simple Linear Regression model. The error bars represent standard deviations.

Table S5

Fixed effects within a mixed effects model on BrainAGEs obtained from a Simple Linear Regression brain age prediction model.

Fixed effect	Value of β-coefficient	Standard Error	*df*	t-value	*p*-value
(Intercept)	0.06	0.14	218	0.43	.67
Time2	0.40	0.15	134	2.74	.007**
HR-well	-0.46	0.22	218	-2.12	.04*
HR-MD	-0.04	0.27	218	-0.17	.87
C-MD	0.65	0.41	218	1.56	.12
Time2*HR-well	-0.23	0.21	134	-1.11	.27
Time2*HR-MD	-0.71	0.12	134	-2.94	.004**
Time2*C-MD	-0.45	0.42	134	-1.08	.28

** $p < .05$, ** $p < .01$*

Results for MD (combination of HR-MD and C-MD)

Table S6

Fixed effects within a mixed effects model on BrainAGE that included both HR-participants and control participants that developed a mood disorder as one heterogenous group (MD).

Fixed effect	Value of β-coefficient	Standard Error	*df*	t-value	*p*-value
(Intercept)	0.06	0.12	219	0.49	.62
Time2	0.36	0.13	135	2.72	.007**
HR-well	-0.38	0.18	219	-2.12	.03*
MD	-0.22	0.20	219	-1.09	.28
Time2*HR-well	-0.18	0.19	135	-0.95	.34
Time2*MD	-0.58	0.20	135	-2.85	.005**

** $p < .05$, ** $p < .01$*

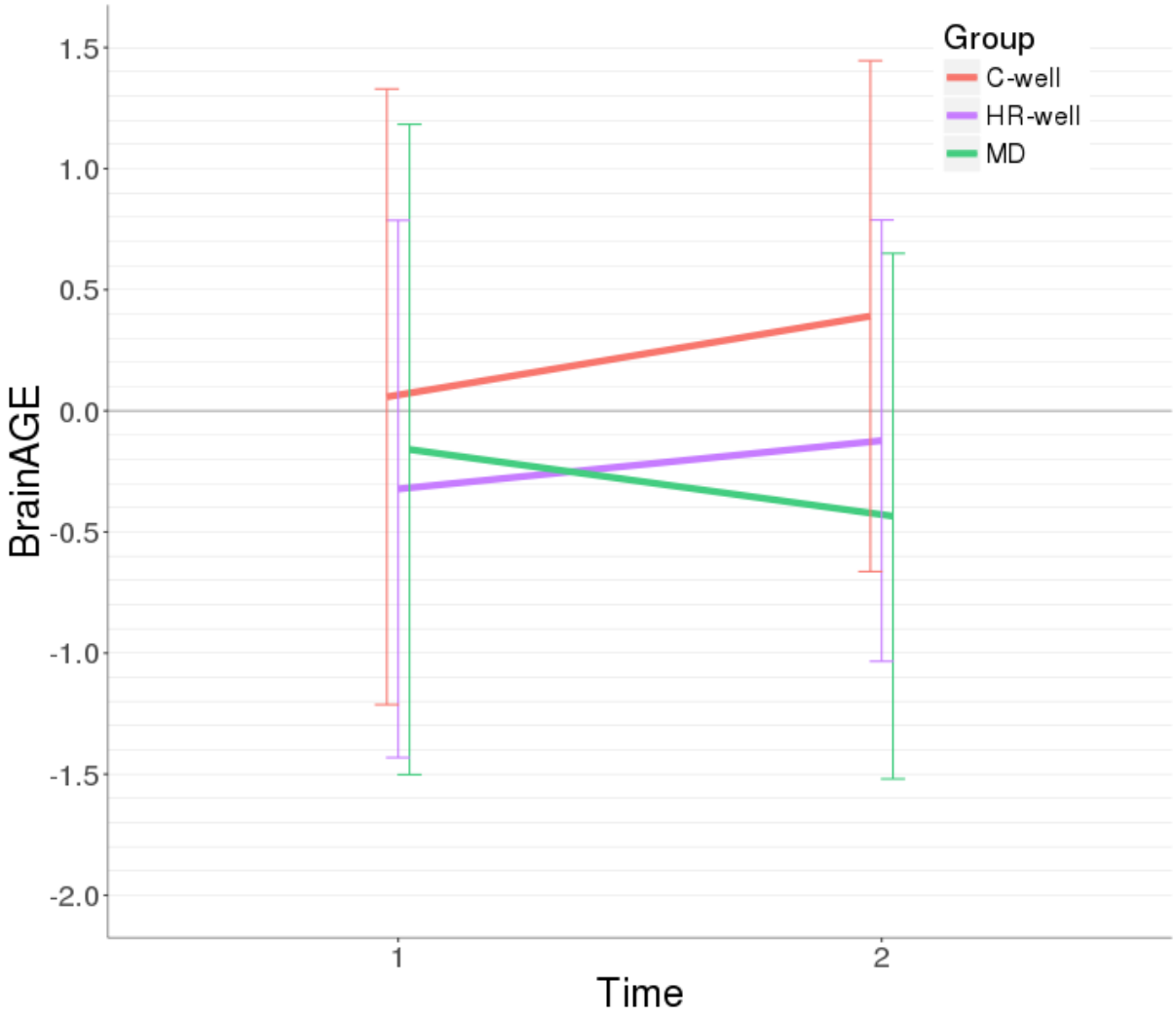


Figure S5. Group means of the brain age gap estimate (BrainAGE) with HR-participants who developed a mood disorder and control participants who developed a mood disorder considered as one heterogenous group (MD). The error bars represent standard deviations.

Supplemental references

Ashburner, J. (2007): A fast diffeomorphic image registration algorithm. *Neuroimage*, *38*(1), 95-113.

Ashburner, J., Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *Neuroimage*, *55*(3), 954-967.

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

Franke, K., Ziegler, G., Klöppel, S., & Gaser, C. (2010). Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters. *NeuroImage, 50,* 883-892.

Gaser, C. & Dahnke, R. (2018). CAT – A Computational Anatomy Toolbox for SPM [computer software]. Retrieved from http://www.neuro.uni-jena.de/cat/index.html#DOWNLOAD

Manjón, J., Coupe, P., Marti-Bonmati, L., Collins, D.L., & Robles, M. (2010). Adaptive Non-Local Means Denoising of MR images with spatially varying noise levels. *Journal of Magnetic Resonance Imaging*, *31*, 192-203.

Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, *16*(2), 176-186.

Schnack, H. G., van Haren, N. E. M., Nieuwenhuis, M., Hulshoff Pol, H. E., Cahn, W., & Kahn, R. S. (2016). Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern

Recognition Study. *American Journal of Psychiatry*, *173*(6), 607-616. doi:10.1176/appi.ajp.2015.15070922

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *Neuroimage, 23*(1), 84-97.