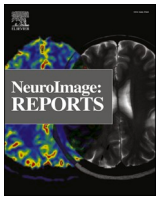
Neuroimage: Reports 2 (2022) 100149

Contents lists available at ScienceDirect

Neuroimage: Reports

journal homepage: www.sciencedirect.com/journal/neuroimage-reports



A large-scale ENIGMA multisite replication study of brain age in depression

Laura K.M. Han [a,b,*], Richard Dinga [c], Ramona Leenings [d,e], Tim Hahn [d], James H. Cole [f,g], Lyubomir I. Aftanas [h,i], Alyssa R. Amod [j], Bianca Besteher [k], Romain Colle [l,m], Emmanuelle Corruble [l,m], Baptiste Couvy-Duchesne [n,o], Konstantin V. Danilenko [h], Paola Fuentes-Claramonte [p,q], Ali Saffet Gonul [r], Ian H. Gotlib [s], Roberto Goya-Maldonado [t], Nynke A. Groenewold [j], Paul Hamilton [u,v], Naho Ichikawa [w,x], Jonathan C. Ipser [j], Eri Itai [w], Sheri-Michelle Koopowitz [j], Meng Li [k], Go Okada [w], Yasumasa Okamoto [w], Olga S. Churikova [h], Evgeny A. Osipov [i], Brenda W.J.H. Penninx [y], Edith Pomarol-Clotet [p,q], Elena Rodríguez-Cano [p,q], Matthew D. Sacchet [z], Hotaka Shinzato [w], Kang Sim [aa,ab], Dan J. Stein [ac], Aslihan Uyar-Demir [r], Dick J. Veltman [y], Lianne Schmaal [a,b]

[a] *Centre for Youth Mental Health, The University of Melbourne, Parkville, VIC, Australia*
[b] *Orygen, Parkville, VIC, Australia*
[c] *Department of Cognitive Science & Artificial Intelligence, Tilburg University, Tilburg, Netherlands*
[d] *Institute for Translational Psychiatry, University of Münster, Münster, Germany*
[e] *Faculty of Mathematics and Computer Science, University of Münster, Münster, Germany*
[f] *Centre for Medical Image Computing, Department of Computer Science, University College London, United Kingdom*
[g] *Dementia Research Centre, Queen Square Institute of Neurology, University College London, United Kingdom*
[h] *Scientific Research Institute of Neuroscience and Medicine, Novosibirsk, Russia*
[i] *Novosibirsk State University, Novosibirsk, Russia*
[j] *Department of Psychiatry & Mental Health, Neuroscience Institute, University of Cape Town, South Africa*
[k] *Department of Psychiatry and Psychotherapy, Jena University Hospital, Jena, Germany*
[l] *Service Hospitalo-Universitaire de Psychiatrie de Bicêtre, Mood Center Paris Saclay, Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires Paris-Saclay, Hôpital de Bicêtre, Le Kremlin Bicêtre, F-94275, France*
[m] *MOODS Team, INSERM 1018, CESP (Centre de Recherche en Epidémiologie et Santé des Populations), Université Paris-Saclay, Faculté de Médecine Paris-Saclay, Le Kremlin Bicêtre, F-94275, France*
[n] *Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de La Pitié Salpêtrière, F-75013, Paris, France*
[o] *Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, 4072, Australia*
[p] *FIDMAG Hermanas Hospitalarias Research Foundation, Barcelona, Spain*
[q] *CIBERSAM, Barcelona, Spain*
[r] *SoCAT Lab, Department of Psychiatry, School of Medicine, Ege University, Izmir, Turkey*
[s] *Department of Psychology, Stanford University, Stanford, CA, USA*
[t] *Laboratory of Systems Neuroscience and Imaging in Psychiatry (SNIP-Lab), University Medical Center Göttingen, Göttingen, Germany*
[u] *Center for Social and Affective Neuroscience, Department of Biomedical and Clinical Sciences, Linköping University, Sweden*
[v] *Center for Medical Image Science and Visualization, Department of Biomedical and Clinical Sciences, Linköping University, Sweden*
[w] *Department of Psychiatry and Neurosciences, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan*
[x] *Deloitte Analytics R&D, Deloitte Touche Tohmatsu LLC, Tokyo, Japan*
[y] *Amsterdam UMC, Location Vrije Universiteit, Department of Psychiatry and Amsterdam Neuroscience, Amsterdam, the Netherlands*
[z] *Meditation Research Program, Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA*
[aa] *West Region, Institute of Mental Health, Singapore*
[ab] *Yong Loo Lin School of Medicine, National University of Singapore, Singapore*
[ac] *SAMRC Unit on Risk & Resilience in Mental Disorders, Dept of Psychiatry & Neuroscience Institute, University of Cape Town, Cape Town, South Africa*

ARTICLE INFO

Keywords:
Brain age
Replication study

ABSTRACT

Background: Several studies have evaluated whether depressed persons have older appearing brains than their nondepressed peers. However, the estimated neuroimaging-derived "brain age gap" has varied from study to study, likely driven by differences in training and testing sample (size), age range, and used modality/features.

* Corresponding author. University of Melbourne, 35 Poplar road, 3052 Parkville, VIC, Australia.
E-mail address: laura.han@unimelb.edu.au (L.K.M. Han).

https://doi.org/10.1016/j.ynirp.2022.100149
Received 4 July 2022; Received in revised form 7 November 2022; Accepted 23 November 2022
Available online 29 November 2022

Depression
ENIGMA consortium
Biological aging

To validate our previously developed ENIGMA brain age model and the identified brain age gap, we aim to replicate the presence and effect size estimate previously found in the largest study in depression to date (N = 2126 controls & N = 2675 cases; +1.08 years [SE 0.22], Cohen's d = 0.14, 95% CI: 0.08–0.20), in independent cohorts that were not part of the original study.

Methods: A previously trained brain age model (www.photon-ai.com/enigma_brainage) based on 77 FreeSurfer brain regions of interest was used to obtain unbiased brain age predictions in 751 controls and 766 persons with depression (18–75 years) from 13 new cohorts collected from 20 different scanners. Meta-regressions were used to examine potential moderating effects of basic cohort characteristics (e.g., clinical and scan technical) on the brain age gap.

Results: Our ENIGMA MDD brain age model generalized reasonably well to controls from the new cohorts (predicted age vs. age: $r = 0.73$, $R^2 = 0.47$, MAE = 7.50 years), although the performance varied from cohort to cohort. In these new cohorts, on average, depressed persons showed a significantly higher brain age gap of +1 year (SE 0.35) (Cohen's d = 0.15, 95% CI: 0.05–0.25) compared with controls, highly similar to our previous finding. Significant moderating effects of FreeSurfer version 6.0 (d = 0.41, p = 0.007) and Philips scanner vendor (d = 0.50, p < 0.0001) were found, leading to more positive effect size estimates.

Conclusions: This study further validates our previously developed ENIGMA brain age algorithm. Importantly, we replicated the brain age gap in depression with a comparable effect size. Thus, two large-scale independent mega-analyses across in total 32 cohorts and >3400 patients and >2800 controls worldwide show reliable but subtle effects of brain aging in adult depression. Future studies are needed to identify factors that may further explain the brain age gap variance between cohorts.

1. Introduction

Recently, considerable literature has emerged around the theme of human aging. Aging is accompanied by complex biological changes, such as linear and nonlinear brain structural changes (Anderton, 2002). Machine learning algorithms can leverage these age-related brain patterns to predict **chronological age**, to explain individual differences in aging. If (structural) magnetic resonance imaging (MRI) data are used as input for these algorithms, the output can be considered as an estimate of brain-based biological age, or, predicted **brain age** (Cole and Franke, 2017). Over the past decade there has been an exponential increase in studies investigating brain age (Baecker et al., 2021), with this metric being used to quantify one's brain health state, as well as risk for aging-related diseases and mortality (Cole et al., 2018). These are important indicators of neurodegenerative diseases such as Alzheimer's or multiple sclerosis; however, these risks are also commonly increased (albeit to a lesser extent) in major depressive disorder (MDD) (Penninx, 2017).

The estimated neuroimaging-derived "brain age gap" (predicted brain age minus chronological age, i.e., brain-predicted age difference, or, **brain-PAD**) in depression varies across studies in terms of both effect size and statistical significance. These differences are likely driven by differences in sample properties (e.g., age range), but also training and testing sample (size), and methods used (e.g., modality/features). A recent systematic review and meta-analysis summarized that the majority (4 out of 7) of the existing studies of brain-PAD in depression did not establish a significant case-control difference (Ballester et al., 2022). Yet, effects were compatible across studies; thus, all studies identified a higher average brain age gap in depression compared to controls, with a pooled effect of approximately +1 year of added aging, although estimated gaps ranged from 0.13 to 4.92 years. Our previous ENIGMA MDD consortium study, the largest study of brain age in depression to date (N = 2126 controls and N = 2675 patients), showed a +1.08 year higher brain-PAD in depression (Cohen's d = 0.14), but with no evidence that this effect was driven by specific (clinical) characteristics such as age, age of onset, recurrence status, remission status, or antidepressant use (Han et al., 2021a). The subsequent addition of new cohorts to the ENIGMA MDD consortium since our previous brain age study provides us with an unique opportunity to perform an independent replication study in new data to validate our developed algorithm, as well as determine whether depression is consistently associated with older appearing brains (Wrigglesworth et al., 2021). Studying the impact of depressive psychopathology on age-related structural brain patterns may help to explain why persons with depression have an increased risk for poorer brain and physical health compared to their nondepressed peers.

Our ENIGMA brain age algorithm was trained on 952 male and 1236 female healthy controls from 19 cohorts. We used 77 FreeSurfer-derived ROI features (34 cortical thickness, 34 cortical surface area, 7 subcortical volumes, lateral ventricles, and intracranial volume) to predict chronological age using ridge regression. While other existing (deep neural network) algorithms may potentially provide more accurate predictions (Lombardi et al., 2020), most of them rely on using higher-dimensional imaging data as input (e.g., raw scans, individual-level voxels, or vertices). Within the ENIGMA consortium, many cohorts have shared data in the form of brain-derived summary measures (i.e., FreeSurfer brain regions of interest, or ROIs). The current FreeSurfer ROIs method is thus one of the more practical ways to perform a large multisite brain age mega-analysis in depression, facilitated by the collaborative nature of the ENIGMA MDD working group. Our first study showed good out-of-sample generalization to new and unseen controls and patients from the same cohorts as the model was trained on, as well as completely independent controls from cohorts not included in training (i.e., ENIGMA Bipolar Disorder controls) (Han et al., 2021a). Additionally, other ENIGMA studies have further demonstrated the validity of this model (Clausen et al., 2022), identifying a higher brain-PAD in schizophrenia (Constantinides et al., 2022).

This study aims to further validate the ENIGMA FreeSurfer ROI-based brain age prediction method, by evaluating the performance of our algorithm in 13 new and unseen cohorts of individuals with depression collected from 20 independent scanners. Importantly, we aim to contribute to the growing area of brain age research by attempting to replicate the magnitude of the brain age gap difference previously reported by the ENIGMA-MDD consortium between persons with depression (N = 766) and controls (N = 751) using this method.

2. Methods

2.1. Samples

Thirteen independent cohorts (N = 1517) from the ENIGMA MDD working group with data from people with major depression and controls (18–75 years old) participated in this replication study. Cohort-specific details on demographics, basic clinical characteristics, and exclusion criteria can be found in the **Supplement**. All sites obtained approval from their local institutional review boards and ethics committees. All study participants provided written informed consent.

2.2. ENIGMA brain age prediction model

Model development is described in more detail in (Han et al., 2021a), but in short, ridge regression was used to predict age from 77 FreeSurfer-derived features (7 subcortical volumes, 34 cortical thickness regions, 34 cortical surface area regions, lateral ventricles, and intracranial volume) in healthy controls (no history of mental or neurological illness). FreeSurfer features were averaged across hemispheres as this improved the model fit of the original algorithm. Separate models were trained for male (N = 952) and female (N = 1236) controls. The ENIGMA brain age model is publicly available (www.photon-ai.com/enigma_brainage) and was applied to the independent new ENIGMA MDD cohorts included in the current study. A schematic overview is displayed in Fig. 1.

2.3. Model generalization

The ENIGMA brain age prediction model has previously been validated in 646 unseen male and 757 unseen female control samples from 23 independent scanners that were not part of the training data (Han et al., 2021a), as well as in other disease working groups of ENIGMA (Clausen et al., 2022; Constantinides et al., 2022). In the current study, model generalization was evaluated in control samples collected from 20 additional independent scanners (N = 252 males and N = 499 females). To assess model performance in these data acquired from completely independent cohorts, we calculated *(1)* mean absolute error (MAE), *(2)* weighted MAE (i.e., $_{w}$MAE, an age range informed metric; *MAE ÷ age range of sample*), *(3)* Pearson correlation coefficients between predicted brain age and chronological age, and *(4)* the proportion of the variance explained by the model (R^2). R^2 was calculated using the caret package according to the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\left(y_i - \underline{y}\right)^2}$$

Where n is the number of subjects, y is the chronological age, $\hat{y}$ is the predicted age, and $\underline{y}$ is the average age of subjects in the test set. Please note that according to this formula, R^2 can be negative, even if a correlation between age and predicted age is positive. This happens when model predictions have larger errors than predicting the average age, such as when predictions are biased or have a relatively large variance.

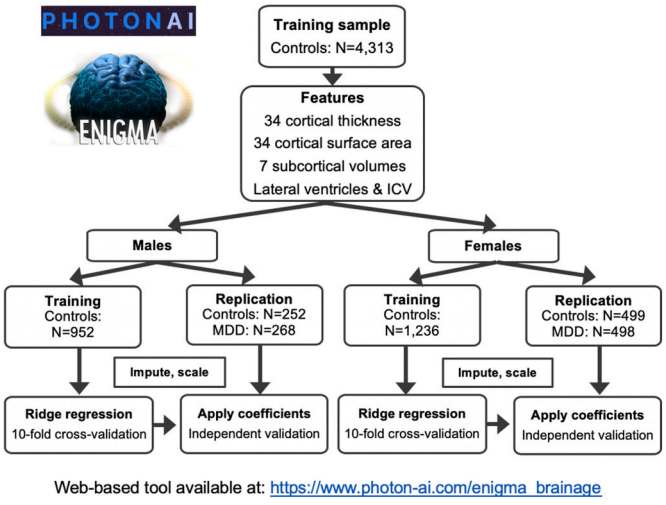


Fig. 1. Schematic overview of features used and data used to train the ENIGMA brain age model. Learned sex-specific ridge regression coefficients were applied to the current independent replication test data.

2.4. Statistical analyses

A mega-analytic approach was taken to replicate the presence and size of the brain age gap in depression by pooling data across all thirteen new cohorts. The brain age gap (predicted brain-based age minus chronological age, or, **brain-PAD**) was calculated for each individual and used as the outcome variable in analyses comparing the difference between brain-PAD in people with depression and controls and examining associations between brain-PAD and clinical characteristics. Each dependent measure of the ith individual at jth scanning site was modeled as follows:

1. $\text{Brain-PAD}_{ij} = \text{intercept} + \beta_1(\text{Dx}) + \beta_2(\text{sex}) + \beta_3(\text{age}) + \beta_4(\text{age}^2) + \beta_5(\text{Dx} \times \text{age}) + \beta_6(\text{Dx} \times \text{sex}) + \beta_7(\text{age} \times \text{sex}) + \beta_8(\text{Dx} \times \text{age} \times \text{sex}) + U_j + \varepsilon_{ij}$
2. $\text{Brain-PAD}_{ij} = \text{intercept} + \beta_1(\text{Dx}) + \beta_2(\text{sex}) + \beta_3(\text{age}) + \beta_4(\text{age}^2) + \beta_5(\text{Dx} \times \text{age}) + \beta_6(\text{Dx} \times \text{sex}) + U_j + \varepsilon_{ij}$
3. $\text{Brain-PAD}_{ij} = \text{intercept} + \beta_1(\text{Dx}) + \beta_2(\text{sex}) + \beta_3(\text{age}) + \beta_4(\text{age}^2) + U_j + \varepsilon_{ij}$

Intercept, Dx (MDD diagnosis), sex, and all age effects were fixed. The terms U_j and ε_{ij} are normally distributed and represent the random intercept attributed to the scanning site and the residual error, respectively. Standardized Cohen's *d* was calculated to indicate the size of the effect. Regression analyses were performed in a subset of participants to associate brain-PAD with depression severity as measured with the Beck's Depression Inventory (Beck et al., 1961) or Hamilton Depression Rating Scales (Hamilton, 2012), and partial-correlation Pearson's r-statistics appropriate for mixed-effects models were calculated to indicate the size of the effect (Nakagawa and Cuthill, 2007). Within the patient group, we also used linear mixed models to examine brain-PAD associations with clinical characteristics (i.e., recurrence status [first versus recurrent episode], antidepressant (AD) status [taking AD yes/no at time of scanning], remitted status [acute versus remitted], age of onset of depression [categorized as: early, <26 years; middle adulthood, >25 and < 56 years; and late adulthood onset, >55 years]). All models included age and age^2 covariates to statistically deal with the age bias of brain-PAD (i.e., correlation between brain-PAD and age) (Le et al., 2018). In addition to the **mega**-analytic approach, a **meta**-analytic approach was also performed to provide further insights into the generalization of the ENIGMA brain age model and case-control difference in individual cohorts. Exploratory effects of cohort specific characteristics (i.e., sample size, mean age, proportion of females) but also potential (scan) technical moderators (i.e., FreeSurfer version, field strength, scanner vendor, performance accuracy metrics [MAE, R^2]) on the brain-PAD outcome were examined by random effects meta-regressions analyses using the *metafor* package in *R* (a more detailed description on methods can be found in the **Supplementary Methods**). All statistical tests were tested two-sided and considered significant at $p < 0.05$.

3. Results

3.1. Participants

Participant characteristics are presented in Table 1. Thirty individuals from one cohort were excluded from the study based on having >10% missing structural brain ROIs data. Two individuals >75 years old from another cohort were also excluded from this study, given that the model was trained on data restricted within 18–75 years. Eight persons showed a brain-PAD with a calculated Z-score >3 (i.e., >3SD away from the global mean) and were excluded from analysis. In total, we included data from N = 1517 participants, including N = 751 controls (66% females) and N = 766 persons with (current) MDD (65% females). The **Supplement** includes cohort-specific information on participants (Supplementary Table S1), image acquisition and processing

Table 1
Participant characteristics per diagnostic group.

Characteristic		N	Controls, N = 751[a]	MDD, N = 766[a]
Cohort	AFFDIS	49	20 (2.7%)	29 (3.8%)
	CSAN	107	49 (6.5%)	58 (7.6%)
	DCHS	70	54 (7.2%)	16 (2.1%)
	FIDMAG	69	34 (4.5%)	35 (4.6%)
	Hiroshima	315	167 (22%)	148 (19%)
	TiPS	105	75 (10.0%)	30 (3.9%)
	MOODS	96	32 (4.3%)	64 (8.4%)
	MOTAR	108	68 (9.1%)	40 (5.2%)
	NESDA	219	65 (8.7%)	154 (20%)
	Novo	128	52 (6.9%)	76 (9.9%)
	Singapore	40	17 (2.3%)	23 (3.0%)
	SoCAT	179	100 (13%)	79 (10%)
	StanfFAA	32	18 (2.4%)	14 (1.8%)
Chronological age (years)		1517	38.80 ± 12.89 (17.00–73.00)	39.82 ± 12.70 (18.00–73.00)
Predicted brain age (years)		1517	41.75 ± 10.79 (14.78–74.60)	43.94 ± 10.65 (15.22–75.66)
Brain-PAD (years)		1517	2.95 ± 8.89 (−25.31-31.39)	4.12 ± 9.86 (−25.19-32.40)
Sex	Female	1517	499 (66%)	498 (65%)
Beck's Depression Inventory (severity)		756	5.40 ± 5.95 (0–42)	28.04 ± 10.87 (0–56)
Hamilton Depression Rating Scale (severity)		642	1.45 ± 2.29 (0–16)	19.07 ± 8.04 (0–41)
Recurrent status	First	1130		225 (38%)
	Recurrent			375 (62%)
Antidepressant use	AD-free	1131		357 (59%)
	AD-using			244 (41%)
Remitted status	Remitted	1004		41 (7.5%)
	Acute			508 (93%)
FreeSurfer version	5.0	1517	65 (8.7%)	154 (20%)
	5.3		403 (54%)	336 (44%)
	6.0		134 (18%)	139 (18%)
	7.1		100 (13%)	79 (10%)
	7.2		49 (6.5%)	58 (7.6%)
Field strength	1.5T	1517	34 (4.5%)	35 (4.6%)
	3T		717 (95%)	731 (95%)
Scanner vendor	GE	1517	104 (14%)	125 (16%)
	Philips		182 (24%)	281 (37%)
	Siemens		465 (62%)	360 (47%)

[a] n (%); Mean ± SD (Minimum-Maximum).

(Supplementary Table S2) and instruments used for depression ascertainment (Supplementary Table S3).

3.2. Brain age prediction performance

Model generalizability was evaluated in the healthy control samples, split by sex. Both pooled and cohort-specific model performances are presented in Supplementary Table S4. While the generalization power varied from cohort to cohort, the pooled performance accuracy was comparable to the out-of-sample generalizability previously reported in (Han et al., 2021a), with current metrics between predicted brain age and chronological age of Pearson's $r = 0.73$, $R^2 = 0.47$, MAE = 7.50 years, and $_{w}$MAE = 0.13 in males and $r = 0.73$, $R^2 = 0.47$, MAE = 7.50 years, and $_{w}$MAE = 0.13 in females. Fig. 2 shows the predicted brain age against chronological age in the pooled sample (Fig. 2A) and per cohort (Fig. 2B), with separate regression lines for controls and patients. Cohorts showing a negative R^2 showed negative mean cortical thickness deviations compared to the grand mean of combined cohorts (Supplementary Fig. S1). Performance metrics in patient samples (separately for males and females) can also be found in the Supplementary Table S5.

3.3. Replication of higher brain age in depression

On average, depressed persons showed a significantly higher brain-PAD of +0.99 (SE 0.35) years (Cohen's $d = 0.15$, 95% CI: 0.04–0.25) compared with controls ($p < 0.01$), Fig. 3. No significant three-way interaction between diagnosis by age and by sex, nor significant two-way interactions (diagnosis by age or diagnosis by sex) were found. The supplementary meta-analytic approach resulted in a slightly higher but similar pooled brain-PAD of +1.20 years and associated effect size of Cohen's $d = 0.19$ between cases and controls ($I^2 = 52.6\%$, indicating moderate heterogeneity). Forest plots are depicted in Supplementary Fig. S2. A significant positive association between brain-PAD and BDI was found across diagnostic groups (N = 756, b = 0.04 years per symptom [SE = 0.02]; r = 0.08 [SE = 0.04], 95% CI: 0.00–0.15, p = 0.02), but this did not reach statistical significance within the patient group (N = 350, b = 0.07 years per symptom [SE = 0.04]; r = 0.09 [SE = 0.05], 95% CI: −0.01-0.20, p = 0.055). Using the clinician-administered HDRS scores, no significant associations were found across diagnostic groups (N = 642, b = 0.05 years per symptom [SE = 0.03]; r = 0.06 [SE = 0.04], 95% CI: −0.01-0.14, p = 0.09) nor within the depressed group (N = 458, b = 0.02 years per symptom [SE = 0.05]; r = 0.02 [SE = 0.05], 95% CI: −0.07-0.11, p = 0.68).

3.4. Patient group analyses and meta-regressions with moderators

No significant differences in brain-PAD were found between patient subgroups (recurrent versus first episode depression [b = 0.00 years, p = 0.99], AD-free versus AD-using patients [b = 0.83 years, p = 0.20], acute versus remitted depression [b = −1.07 years, p = 0.43], or age of onset of depression in middle [b = 0.55, p = 0.49] or late adulthood [b = 0.84, p = 0.66] compared to early onset). The meta-regressions with sample size, mean age, proportion of females, proportion of first/recurrent episode patients, proportion of AD-free/AD-using patients, proportion of remitted/acute patients, field strength and performance accuracy metrics (MAE, R^2) did not significantly moderate the Cohen's d effect size estimates of brain-PAD (all QMp's > 0.05, Supplementary Table S6). However, significant moderating effects of FreeSurfer version 6.0 (d = 0.41, p = 0.007) (Fig. 4A) and Philips scanner vendor (d = 0.50, p < 0.0001) were found (Fig. 4B), leading to more positive effect size estimates.

4. Discussion

The current study replicated the finding that persons with depression reliably show older appearing brains, with a similar estimated gap and associated effect size (+1 year, Cohen's d = 0.14) as previously found in our largest mega-analysis of brain age in depression to date (+1.08 years, Cohen's d = 0.14) (Han et al., 2021a). While the generalization of our algorithm varied from cohort to cohort, pooled metrics were comparable to the performance accuracy found in the out-of-test samples in the original study. Importantly, post-hoc sensitivity analyses revealed that the exclusion of cohorts showing poor generalization did not change our replication findings (Supplementary Tables S7 and S8). In addition, a meta-analytic approach resulted in a highly similar pooled effect size (+1.20 years, Cohen's d = 0.19), providing robust evidence for significant but subtle age-related structural brain patterns in depression compared to controls.

The current multi-site replication study provides further evidence that the brain age gap in depression is an estimated +1 year (Cohen's d = 0.14), consistent with our previous mega-analysis in 19 other cohorts (Han et al., 2021a) and another meta-analysis including an additional 6 studies (Ballester et al., 2022). Taken together, the impact of depression on age-related structural brain differences thus seems to be rather subtle. However, it is important to note that the small, pooled effect size did not result from consistent small effects in each individual cohort, as can be seen from the forest plots of the meta-analyses in Supplementary Fig. S2. Instead, the subtlety of the effect seemed to be driven by the fact that four of the cohorts showed larger positive effects (Cohen's d's ranging from 0.40 to 0.67, mean = 0.50), whereas remaining cohorts showed no significant effects. However, effect sizes were not moderated by (Supplementary Table S6) or related to model generalization (i.e., small, or negative effect sizes were not only observed in cohorts showing

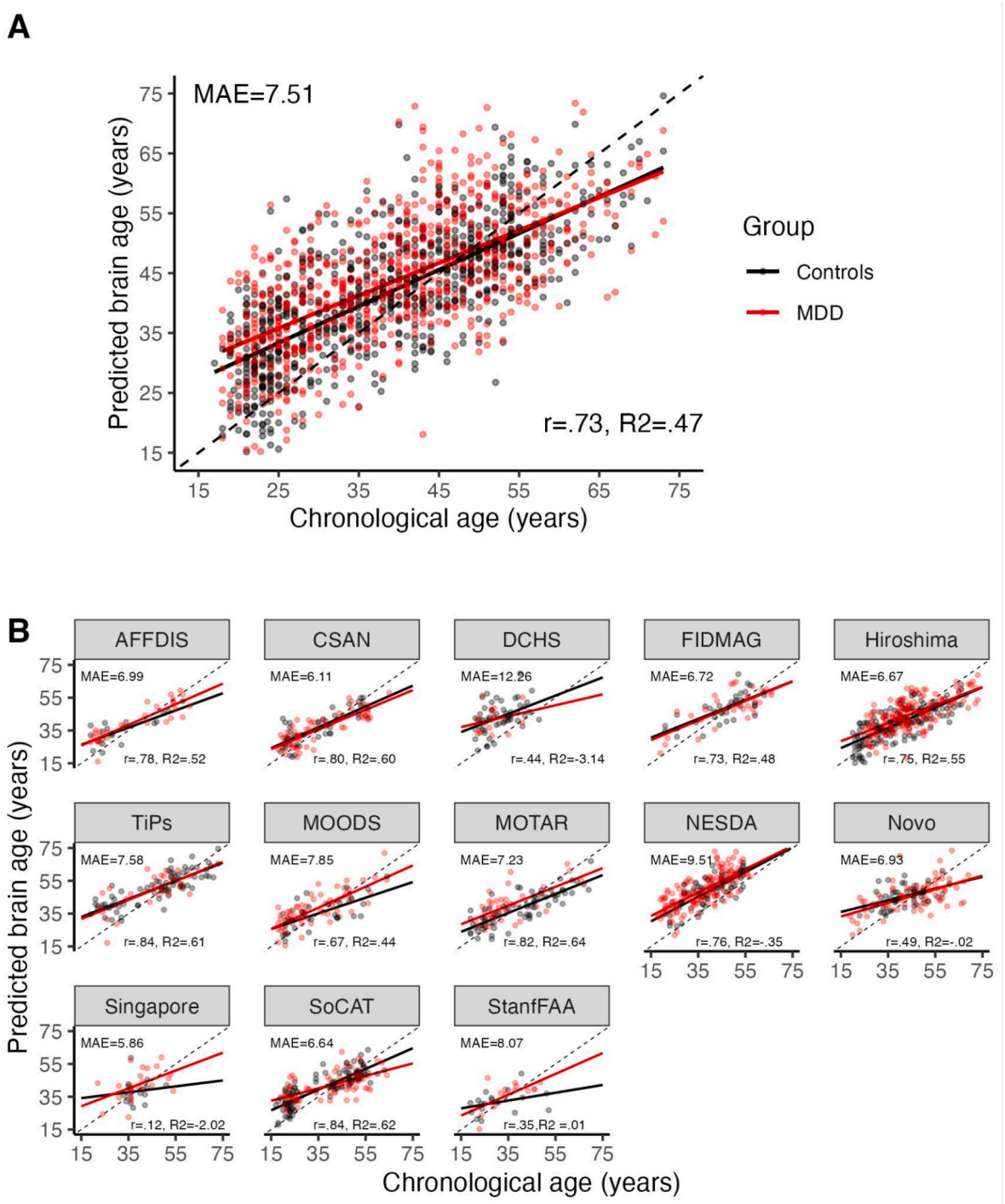


Fig. 2. Brain age prediction using the ENIGMA algorithm in 13 new and unseen cohorts from 20 different scanners. (A) Chronological age against predicted brain age in the pooled sample and **(B)** per cohort. Separate regression lines are plotted for controls (black) and persons with depression (red). Diagonal dashed line reflects the line of identity (x = y). Predictions were pooled across males and females to calculate the performance metrics in the control group. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

poor performance accuracy) (Supplementary Table S8). In addition, post-hoc correlation analyses (Pearson's r) and scatter plots (Supplementary Fig. S3) showed no direct relationship between model performance (R^2: r = 0.02, p = 0.95; MAE: r = −0.01, p = 0.96) and brain-PAD. While negative R^2 observed in some cohorts can likely be explained by lower values of the cortical thickness features in those particular cohorts (Supplementary Fig. S1), the inconsistency in effect sizes between cohorts may rather be due to other sources of variation unrelated to basic cohort or clinical characteristics such as first episode vs. recurrent, antidepressant free vs. antidepressant using or acute vs. remitted patients, as we did not observe any differences between these subgroups.

Depressive state as measured by the BDI but not clinician-rated HDRS seemed to be weakly related to the brain age gap, indicated by the lack of significant associations with depression severity scores in the patient group, although the latter may also be due to the reduced sample size. Interestingly, symptom severity scores were however highly correlated in overlapping samples that included both BDI and HDRS

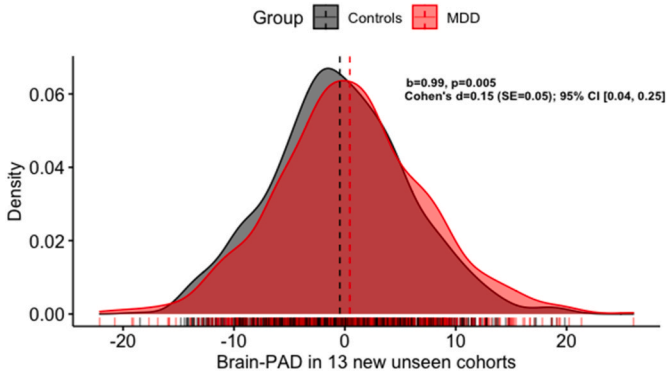


Fig. 3. Replication of the brain age gap difference between controls and persons with depression. Brain-PAD (predicted brain age minus chronological age) in persons with major depressive disorder (MDD) and controls. Group-level analyses showed significantly higher brain-PAD in persons with MDD than controls in pooled samples of thirteen cohorts ($b = 0.99$ years, $p = 0.005$). The brain-PAD estimates are adjusted for chronological age, age^2, sex and scanning site.

measures (N = 462, Pearson's r = 0.81, p < 0.0001), although the coefficient was reduced in patient samples only (N = 318, r = 0.48, p < 0.0001). The discrepancy could perhaps be explained by different ways the HDRS may be scored, inter-rater reliability or the differential proportion of items emphasizing cognitive and affective (BDI-II) or somatic and behavioral dimensions (HDRS-17) (Brown et al., 1995). Alternatively, the brain age gap may potentially be more sensitive to subjective (BDI) than to objectively (HDRS-17) rated experiences, consistent with the study finding that subjective experience of aging was closely related to the brain age gap (Kwak et al., 2018). Future studies may further investigate other clinical and symptom characteristics to explain such differences.

The current study did examine several potential technical sources of bias such as field strength, scanner vendor, and FreeSurfer version. In terms of scan technical moderators, we found that image acquisition with a Philips scanner (in contrast to Siemens or General Electric vendors) and FreeSurfer version for processing images (v6.0, in contrast to v5.0, v5.3, v7.1, v7.2) showed significant moderating effects on the effect size of the case-control difference in the brain age gap. The distribution and variety of scanner vendors used to train our brain age algorithm can be found in Supplementary Table S9. A qualitative comparison shows that the Philips scanner vendor was not necessarily overrepresented in the training sample (28%), and is unlikely to explain the significant moderating effect of the Philips scanner vendor in the current study. Yet to gain more insight, we repeated the meta-analytic approach used here on the test data from the original Han et al. (2021) study (Supplementary Fig. S4). When retrospectively performing meta regressions for scanner vendor and FreeSurfer version on the test data used in the original study, we found significant moderating effects of Philips (d = 0.28, p < 0.0001) but also Siemens scanner vendor (d = 0.11, p = 0.01), as well as for FreeSurfer v5.3 (d = 0.16, <0.0001), leading to more positive effect size estimates compared to General Electric or Bruker scanner vendors, or FreeSurfer v5.0, v5.1, and v6.0, respectively (Supplementary Figs. S5 and S6). Scanner manufacturer differences may potentially lead to non-negligible differences in cortical thickness, surface area, and volume (Potvin et al., 2016), and a recent study also found that different FreeSurfer pipelines may generate different statistical outcomes in case-control comparison studies (Filip et al., 2022). Further work is required to thoroughly examine the influence of scan technical variables on brain age prediction performance and statistical outcomes. Future studies may, for example, consider using retrospective techniques to accommodate site-effects in multi-site neuroimaging studies (Bayer et al., 2022). However, importantly, current site effects were corrected for in both the current mega-analysis and in that of the original study, and it also seems plausible that other heterogeneous demographic, psychosocial, clinical, or biological cohort-specific characteristics, which we did not measure, coincided with the scanner vendor variable (i.e., biological sampling bias).

A recent systematic review, for example, suggests a role for epigenetic factors, and work investigating whether (genetic risk) for epigenetic aging contributes to the brain-PAD metric is underway in the ENIGMA consortium. While other literature suggests differential brain aging effects in older adults compared to middle-aged adults (i.e., only significantly higher brain-PAD in geriatric sample) (Christman et al., 2020), females and males (i.e., brain-PAD only associated with depressive severity in males) (Dunlop et al., 2021), or stage-dependent relationships with depression (i.e., only occurring at illness onset) (Han et al., 2021), we did not confirm this in the current study. To interpret these different findings, we must acknowledge that this may be due to the older age of the geriatric sample (mean = 66.41 years [SD = 5.45] in Christman et al. vs. mean = 39.31 years [SD = 12.80] here) and different imaging-derived modality (functional MRI in Dunlop et al. vs. structural MRI used here). Furthermore, detailed information on ethnicity, socioeconomic and psychosocial variance were not available and its impact on (the performance of the) brain age (prediction model) could not be evaluated in more detail here. However, an independent study including the NESDA cohort showed selectively older appearing brains in those with high *somatic* symptom severity (Han et al., 2021b). Future studies with more detailed (clinical) characterization (e.g., individual or clusters of depressive symptoms) are needed to gain more insight into which factors consistently contribute to the brain-PAD metric.

A major strength of this replication study is the harmonized approach of data preprocessing, quality checking, and brain age prediction algorithm across cohorts, potentially limiting the sources of bias that may stem from these decisions. This study is therefore a good example of the advantage of consortium efforts and collaborative team science. A note of caution is however due, since within individual cohorts, the case-control difference may not be consistent, present, or significant, also explaining the inconsistent findings across individual studies (Ballester et al., 2022). Unfortunately, due to a lack of harmonized clinical, demographic, and psychosocial information in consortia like ENIGMA MDD, we are limited in our ability to identify factors that could explain the variance in brain-PAD between cohorts. Finally, while the brain age predictions may be more accurate with higher-dimensional data from multimodal sources, it remains an open question whether models with improved performance accuracy show increased sensitivity in detecting subsequent associations with clinical psychopathology.

5. Conclusion

This replication study using data from 13 cohorts around the world confirmed our previous findings that persons with major depressive disorder show advanced brain aging compared to controls by approximately +1 year. Thus, two large-scale independent but harmonized mega-analyses across 32 cohorts and >3400 patients and >2800 controls show a reliable but subtle pattern of brain aging in adult depression. It is important to note that the small, pooled effect is not due to consistent small effects across cohorts but may be driven in part by the heterogeneity across scanning sites. Although we did not find a relation between basic patient properties and the effect size difference in the brain age gap, future work is needed to examine which scan technical, clinical or biological characteristics may underlie the individual variation in the brain age gap.

Funding

BCD is supported by a NHMRC CJ Martin fellowship (APP1161356).

DJS received financial support from the South African Medical Research Council (SAMRC).

JPH is funded by the ALF Grants, Region Östergötland, Sweden.

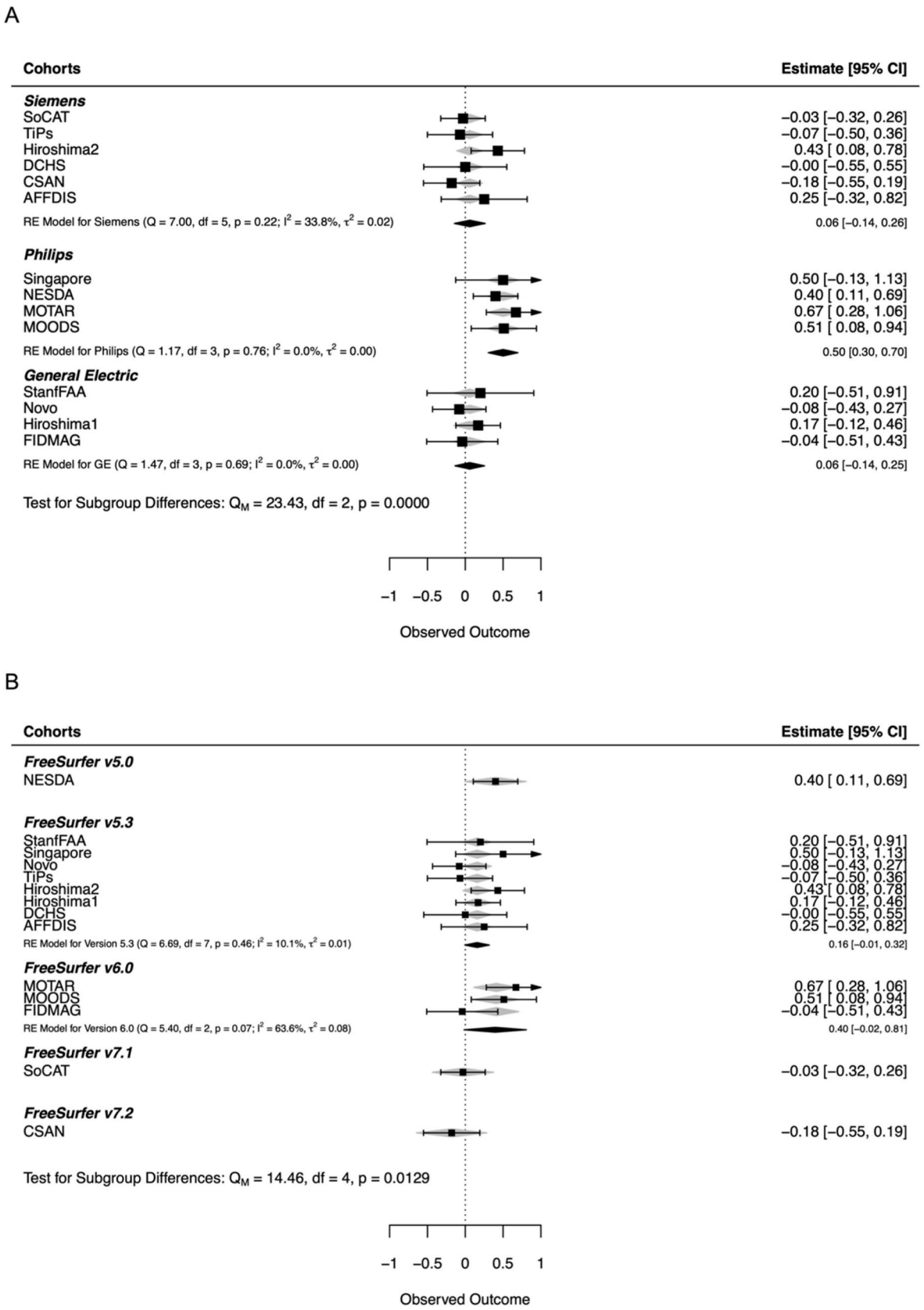


Fig. 4. Forest plots with subgroups from meta regressions with (A) scanner vendors used for image acquisition and **(B)** FreeSurfer version used for image processing, as moderators. Observed outcome and estimate indicate Cohen's d effect size.

KS is funded by National Healthcare Group, Singapore (SIG/15012) for the project.

LA is supported by budgetary financing from the Russian Science Foundation grant (#16-15-00128) by the Institute of Neurosciences and Medicine in 2014–2021.

LH was funded by the Rubicon award (grant number 452020227) from the Dutch NWO.

LS is supported by an MRFF CDF fellowship (APP1140764).

MDS is supported by the National Institute of Mental Health (Project Number R01MH125850), Dimension Giving Fund, Ad Astra Chandaria Foundation, Brain and Behavior Research Foundation (Grant Number 28972), BIAL Foundation (Grant Number 099/2020), Emergence Benefactors, The Ride for Mental Health, Gatto Foundation, and individual donors.

RGM was supported by the University Medical Center Göttingen (UMG) and the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF: 01 ZX 1507, ‘‘PreNeSt - e:Med‘‘)

The DCHS cohort is funded by the Bill & Melinda Gates Foundation [OPP 1017641].The Hiroshima cohort was supported by AMED (Grant Number: JP18dm0307002). This study was further supported by NIH RO1 grants with award numbers MH129832 (LS), MH117601 (LS).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

We would like to thank all research participants, ENIGMA MDD working group contributors and collaborators for sharing their data and promoting team science.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ynirp.2022.100149.

References

Anderton, B.H., 2002. Ageing of the brain. Mech. Ageing Dev. 123, 811–817.

Baecker, L., Garcia-Dias, R., Vieira, S., Scarpazza, C., Mechelli, A., 2021. Machine learning for brain age prediction: introduction to methods and clinical applications. EBioMedicine 72, 103600.

Ballester, P.L., Romano, M.T., de Azevedo Cardoso, T., Hassel, S., Strother, S.C., Kennedy, S.H., Frey, B.N., 2022. Brain age in mood and psychotic disorders: a systematic review and meta-analysis. Acta Psychiatr. Scand. 145, 42–55.

Bayer, J.M.M., Thompson, P.M., Ching, C.R.K., Liu, M., Chen, A., Panzenhagen, A.C., Jahanshad, N., Marquand, M.F., Schmaal, L., Saemann, P.G., 2022. Site effects how-to & when: an overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. Front. Neurol. 13, 923988.

Beck, A.T., Ward, C., Mendelson, M., Mock, J., Erbaugh, J., 1961. Beck depression inventory (BDI). Arch. Gen. Psychiatr. 4, 561–571.

Brown, C., Schulberg, H.C., Madonia, M.J., 1995. Assessment depression in primary care practice with the Beck depression inventory and the Hamilton rating scale for depression. Psychol. Assess. http://psycnet.apa.org/fulltext/1995-27650-001.html.

Christman, S., Bermudez, C., Hao, L., Landman, B.A., Boyd, B., Albert, K., Woodward, N., Shokouhi, S., Vega, J., Andrews, P., Taylor, W.D., 2020. Accelerated brain aging predicts impaired cognitive performance and greater disability in geriatric but not midlife adult depression. Transl. Psychiatry 10, 317.

Clausen, A.N., Fercho, K.A., Monsour, M., Disner, S., Salminen, L., Haswell, C.C., Rubright, E.C., Watts, A.A., Buckley, M.N., Maron-Katz, A., Sierk, A., Manthey, A., Suarez-Jimenez, B., Olatunji, B.O., Averill, C.L., Hofmann, D., Veltman, D.J., Olson, E.A., Li, G., Forster, G.L., Walter, H., Fitzgerald, J., Théberge, J., Simons, J.S., Bomyea, J.A., Frijling, J.L., Krystal, J.H., Baker, J.T., Phan, K.L., Ressler, K., Han, L. K.M., Nawijn, L., Lebois, L.A.M., Schmaal, L., Densmore, M., Shenton, M.E., van Zuiden, M., Stein, M., Fani, N., Simons, R.M., Neufeld, R.W.J., Lanius, R., van Rooij, S., Koch, S.B.J., Bonomo, S., Jovanovic, T., deRoon-Cassini, T., Ely, T.D., Magnotta, V.A., He, X., Abdallah, C.G., Etkin, A., Schmahl, C., Larson, C., Rosso, I.M., Blackford, J.U., Stevens, J.S., Daniels, J.K., Herzog, J., Kaufman, M.L., Olff, M., Davidson, R.J., Sponheim, S.R., Mueller, S.C., Straube, T., Zhu, X., Neria, Y., Baugh, L.A., Cole, J.H., Thompson, P.M., Morey, R.A., 2022. Assessment of brain age in posttraumatic stress disorder: findings from the ENIGMA PTSD and brain age working groups. Brain Behav 12, e2413.

Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends Neurosci. 40, 681–690.

Cole, J.H., Ritchie, S.J., Bastin, M.E., Valdés Hernández, M.C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., Wray, N.R., Redmond, P., Marioni, R.E., Starr, J.M., Cox, S.R., Wardlaw, J.M., Sharp, D.J., Deary, I.J., 2018. Brain age predicts mortality. Mol. Psychiatr. 23, 1385–1392.

Constantinides, C., Han, L.K.M., Alloza, C., Antonucci, L., Arango, C., Ayesa-Arriola, R., Banaj, N., Bertolino, A., Borgwardt, S., Bruggemann, J., Bustillo, J., Bykhovski, O., Carr, V., Catts, S., Chung, Y.-C., Crespo-Facorro, B., Díaz-Caneja, C.M., Donohoe, G., Plessis, S.D., Edmond, J., Ehrlich, S., Emsley, R., Eyler, L.T., Fuentes-Claramonte, P., Georgiadis, F., Green, M., Guerrero-Pedraza, A., Ha, M., Hahn, T., Henskens, F.A., Holleran, L., Homan, S., Homan, P., Jahanshad, N., Janssen, J., Ji, E., Kaiser, S., Kaleda, V., Kim, M., Kim, W.-S., Kirschner, M., Kochunov, P., Kwak, Y.B., Kwon, J.S., Lebedeva, I., Liu, J., Mitchie, P., Michielse, S., Mothersill, D., Mowry, B., de la Foz, V. O.-G., Pantelis, C., Pergola, G., Piras, F., Pomarol-Clotet, E., Preda, A., Quidé, Y., Rasser, P.E., Rootes-Murdy, K., Salvador, R., Sangiuliano, M., Sarró, S., Schall, U., Schmidt, A., Scott, R.J., Selvaggi, P., Sim, K., Skoch, A., Spalletta, G., Spaniel, F., Thomopoulos, S.I., Tomecek, D., Tomyshev, A.S., Tordesillas-Gutiérrez, D., van Amelsvoort, T., Vázquez-Bourgon, J., Vecchio, D., Voineskos, A., Weickert, C.S., Weickert, T., Thompson, P.M., Schmaal, L., van Erp, T.G.M., Turner, J., Cole, J.H., Dima, D., Walton, E., 2022. Brain ageing in schizophrenia: evidence from 26 international cohorts via the ENIGMA Schizophrenia consortium. bioRxiv. https://doi.org/10.1101/2022.01.10.21267840.

Dunlop, K., Victoria, L.W., Downar, J., Gunning, F.M., Liston, C., 2021. Accelerated brain aging predicts impulsivity and symptom severity in depression. Neuropsychopharmacology 46, 911–919.

Filip, P., Bednarik, P., Eberly, L.E., Moheet, A., Svatkova, A., Grohn, H., Kumar, A.F., Seaquist, E.R., Mangia, S., 2022. Different FreeSurfer versions might generate different statistical outcomes in case-control comparison studies. Neuroradiology 64, 765–773.

Hamilton, M., 2012. Hamilton Rating Scale for Depression. PsycTESTS Dataset. https://doi.org/10.1037/t04100-000.

Han, L.K.M., Dinga, R., Hahn, T., Ching, C.R.K., Eyler, L.T., Aftanas, L., Aghajani, M., Aleman, A., Baune, B.T., Berger, K., Brak, I., Filho, G.B., Carballedo, A., Connolly, C. G., Couvy-Duchesne, B., Cullen, K.R., Dannlowski, U., Davey, C.G., Dima, D., Duran, F.L.S., Enneking, V., Filimonova, E., Frenzel, S., Frodl, T., Fu, C.H.Y., Godlewska, B.R., Gotlib, I.H., Grabe, H.J., Groenewold, N.A., Grotegerd, D., Gruber, O., Hall, G.B., Harrison, B.J., Hatton, S.N., Hermesdorf, M., Hickie, I.B., Ho, T.C., Hosten, N., Jansen, A., Kähler, C., Kircher, T., Klimes-Dougan, B., Krämer, B., Krug, A., Lagopoulos, J., Leenings, R., MacMaster, F.P., MacQueen, G., McIntosh, A., McLellan, Q., McMahon, K.L., Medland, S.E., Mueller, B.A., Mwangi, B., Osipov, E., Portella, M.J., Pozzi, E., Reneman, L., Repple, J., Rosa, P.G. P., Sacchet, M.D., Sämann, P.G., Schnell, K., Schrantee, A., Simulionyte, E., Soares, J. C., Sommer, J., Stein, D.J., Steinsträter, O., Strike, L.T., Thomopoulos, S.I., van Tol, M.-J., Veer, I.M., Vermeiren, R.R.J.M., Walter, H., van der Wee, N.J.A., van der Werff, S.J.A., Whalley, H., Winter, N.R., Wittfeld, K., Wright, M.J., Wu, M.-J., Völzke, H., Yang, T.T., Zannias, V., de Zubicaray, G.I., Zunta-Soares, G.B., Abé, C., Alda, M., Andreassen, O.A., Bøen, E., Bonnin, C.M., Canales-Rodriguez, E.J., Cannon, D., Caseras, X., Chaim-Avancini, T.M., Elvsåshagen, T., Favre, P., Foley, S. F., Fullerton, J.M., Goikolea, J.M., Haarman, B.C.M., Hajek, T., Henry, C., Houenou, J., Howells, F.M., Ingvar, M., Kuplicki, R., Lafer, B., Landén, M., Machado-Vieira, R., Malt, U.F., McDonald, C., Mitchell, P.B., Nabulsi, L., Otaduy, M.C.G., Overs, B.J., Polosan, M., Pomarol-Clotet, E., Radua, J., Rive, M.M., Roberts, G., Ruhe, H.G., Salvador, R., Sarró, S., Satterthwaite, T.D., Savitz, J., Schene, A.H., Schofield, P.R., Serpa, M.H., Sim, K., Soeiro-de-Souza, M.G., Sutherland, A.N., Temmingh, H.S., Timmons, G.M., Uhlmann, A., Vieta, E., Wolf, D.H., Zanetti, M.V., Jahanshad, N., Thompson, P.M., Veltman, D.J., Penninx, B.W.J.H., Marquand, A.F., Cole, J.H., Schmaal, L., 2021a. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. Mol. Psychiatr. 26, 5124–5139.

Han, L.K.M., Schnack, H.G., Brouwer, R.M., Veltman, D.J., van der Wee, N.J.A., van Tol, M.-J., Aghajani, M., Penninx, B.W.J.H., 2021b. Contributing factors to advanced brain aging in depression and anxiety disorders. Transl. Psychiatry 11, 402.

Han, S., Chen, Y., Zheng, R., Li, S., Jiang, Y., Wang, C., Fang, K., Yang, Z., Liu, L., Zhou, B., Wei, Y., Pang, J., Li, H., Zhang, Y., Cheng, J., 2021. The stage-specifically accelerated brain aging in never-treated first-episode patients with depression. Hum. Brain Mapp. 42, 3656–3666.

Kwak, S., Kim, H., Chey, J., Youm, Y., 2018. Feeling how old I Am: subjective age is associated with estimated brain age. Front. Aging Neurosci. 10, 168.

Le, Trang T., Rayus, T., Kuplicki, Brett A., McKinney, Hung-Wen, Yeh, Wesley, K., Thompson, Martin P., Paulus, 2018. A nonlinear simulation framework supports adjusting for age when analyzing BrainAGE. Tulsa 1000 Investigators Front. Aging Neurosci. 10 (October), 317.

Lombardi, A., Monaco, A., Donvito, G., Amoroso, N., Bellotti, R., Tangaro, S., 2020. Brain age prediction with morphological features using deep neural networks: results from predictive analytic competition 2019. Front. Psychiatr. 11, 619629.

Nakagawa, Shinichi, Cuthill, Innes C., 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol. Rev. Camb. Phil. Soc. 82 (4), 591–605.
Penninx, B., 2017. Depression and its somatic consequences: allostatic load as the connecting link. Eur. Psychiatr. 41. S62–S62.
Potvin, O., Mouiha, A., Dieumegarde, L., Duchesne, S., 2016. P4-163: effects of scanner manufacturer and strength on cortical surfaces, thicknesses and volumes in the aging brain. Alzheimers. Dement. 12, P1077–P1079.
Wrigglesworth, J., Ward, P., Harding, I.H., Nilaweera, D., Wu, Z., Woods, R.L., Ryan, J., 2021. Factors associated with brain ageing - a systematic review. BMC Neurol. 21, 312.