

## Original Article

\*Shared last authors.

**Cite this article:** van der Markt A *et al* (2023). Accelerated brain aging as a biomarker for staging in bipolar disorder: an exploratory study. *Psychological Medicine* 1–10. <https://doi.org/10.1017/S0033291723002829>

Received: 30 August 2022

Revised: 15 August 2023

Accepted: 4 September 2023

**Keywords:**










aging; biomarkers; bipolar disorder; brainAGE; brain-PAD; staging models; staging

**Corresponding author:**

Afra van der Markt;

Email: [a.vandermarkt@ggzingeest.nl](mailto:a.vandermarkt@ggzingeest.nl)

# Accelerated brain aging as a biomarker for staging in bipolar disorder: an exploratory study

Afra van der Markt<sup>1,2,3</sup> , Ursula Klumpers<sup>1,4</sup> , Annemiek Dols<sup>1,4,5,6</sup> ,  
Nicole Korten<sup>1,2</sup>, Marco P. Boks<sup>6</sup> , Roel A. Ophoff<sup>7,8,9</sup> ,  
Aartjan Beekman<sup>1,2,3</sup> , Ralph Kupka<sup>1,2,3</sup> , Neeltje E. M. van Haren<sup>9,10,\*</sup> ,  
and Hugo Schnack<sup>6,11,\*</sup> 

<sup>1</sup>Psychiatry, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; <sup>2</sup>GGZ inGeest Specialized Mental Health Care, Amsterdam, The Netherlands; <sup>3</sup>Mental Health, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands; <sup>4</sup>Amsterdam Neuroscience, Mood, Anxiety, Psychosis, Sleep & Stress, Amsterdam, The Netherlands; <sup>5</sup>Amsterdam Neuroscience, Neurodegeneration, Amsterdam, The Netherlands; <sup>6</sup>Department of Psychiatry, UMC Utrecht Brain Center, University Utrecht, Utrecht, The Netherlands; <sup>7</sup>Department of Psychiatry and Biobehavioral Science, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; <sup>8</sup>Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; <sup>9</sup>Department of Psychiatry, Erasmus University Medical Center, Rotterdam, The Netherlands; <sup>10</sup>Erasmus Medical Center – Sophia, Child and Adolescent Psychiatry and Psychology, Rotterdam, The Netherlands and <sup>11</sup>Department of Languages, Literature and Communication, Faculty of Humanities, Utrecht University, Utrecht, The Netherlands

**Abstract**

**Background.** Two established staging models outline the longitudinal progression in bipolar disorder (BD) based on episode recurrence or inter-episodic functioning. However, underlying neurobiological mechanisms and corresponding biomarkers remain unexplored. This study aimed to investigate if global and (sub)cortical brain structures, along with brain-predicted age difference (brain-PAD) reflect illness progression as conceptualized in these staging models, potentially identifying brain-PAD as a biomarker for BD staging.

**Methods.** In total, 199 subjects with bipolar-I-disorder and 226 control subjects from the Dutch Bipolar Cohort with a high-quality T1-weighted magnetic resonance imaging scan were analyzed. Global and (sub)cortical brain measures and brain-PAD (the difference between biological and chronological age) were estimated. Associations between individual brain measures and the stages of both staging models were explored.

**Results.** A higher brain-PAD (higher biological age than chronological age) correlated with an increased likelihood of being in a higher stage of the inter-episodic functioning model, but not in the model based on number of mood episodes. However, after correcting for the confounding factors lithium-use and comorbid anxiety, the association lost significance. Global and (sub)cortical brain measures showed no significant association with the stages.

**Conclusions.** These results suggest that brain-PAD may be associated with illness progression as defined by impaired inter-episodic functioning. Nevertheless, the significance of this association changed after considering lithium-use and comorbid anxiety disorders. Further research is required to disentangle the intricate relationship between brain-PAD, illness stages, and lithium intake or anxiety disorders. This study provides a foundation for potentially using brain-PAD as a biomarker for illness progression.

**Introduction**

Bipolar disorder (BD) is a major mental illness characterized by alternating manic and depressive mood episodes, separated by relatively symptom-free intervals. Originally, Kraepelin (1921) suggested that patients with BD reach complete recovery after each mood episode. However, a growing body of evidence has established that patients suffer from illness progression in the form of decreased functioning and disability (Sanchez-Moreno *et al.*, 2009), cognitive decline (El-Badri, Ashton, Moore, Marsh, & Ferrier, 2001), in combination with cellular aging (Fries *et al.*, 2017), and less efficient cellular resilience mechanisms (Fries *et al.*, 2012). The influence of illness progression on research findings is often overlooked. Although the included patients may appear similar from a cross-sectional perspective, they commonly range from having a short to long illness duration with either a favorable or unfavorable longitudinal course, resulting in a heterogeneous sample.

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

**Table 1.** Two staging models for bipolar disorders

| Staging model based on number of episodes (Berk et al., 2007), with adaptations by van der Markt et al. (2020) in italic font |  | Subjects per stage |
|---|--|--------------------|
| Stage 0   | Increased risk of severe mood disorder (e.g. family history, abuse, substance use). No specific symptoms       | 0                  |
| Stage 1   | Mild or non-specific symptoms of mood disorder and prodromal features: ultra-high risk                         | 0                  |
| Stage 2   | First threshold mood episode (depressive or manic)   | 0                  |
| Stage 3   | Recurrence of any depressive, hypomanic, or manic/mixed episode  |                    |
| A   | Recurrence of sub-threshold mood symptoms  | 6                  |
| B   | First threshold relapse  | 12                 |
| C   | <i>Multiple relapses ≤5 episodes*</i>  | 33                 |
|   | <i>Multiple relapses 6–10 episodes*</i>  | 79                 |
|   | <i>Multiple relapses &gt;10 episodes*</i>  | 60                 |
| Stage 4   | Persistent unremitting illness; chronic (>1 year) depressive, manic or mixed episodes, including rapid cycling | 9                  |
| Staging model based on inter-episodic functioning (Kapczinski et al., 2009)   |  |                    |
| Latent  | At risk for developing BD, positive family history, mood or anxiety symptoms without criteria for threshold BD | 0                  |
| Stage I   | Well-defined periods of euthymia without overt psychiatric symptoms  | 21                 |
| Stage II  | Symptoms in inter-episodic periods related to comorbidities  | 65                 |
| Stage III   | Marked impairment in cognition and functioning   | 98                 |
| Stage IV  | Unable to live autonomously due to cognitive and functional impairment   | 15                 |

Staging models in medicine have been developed to define illness progression. For BD, there are two prevailing clinical staging models. Berk et al. (2007) introduced a model based on the occurrence and recurrence of mood episodes while Kapczinski et al. (2009) based their model on impaired inter-episodic functioning (Table 1). In an earlier study, we found clinical markers for illness progression that were related to both models, i.e. familial loading, childhood abuse, psychiatric comorbidity, and treatment resistance (van der Markt et al., 2021). However, we also found evidence suggesting that each model reflects different aspects of illness progression (van der Markt et al., 2020). Earlier age at onset of BD was associated with higher stages in the model based on the recurrence of episodes (Berk et al., 2007), whereas a longer illness duration was associated with higher stages in the model based on inter-episodic functioning (Kapczinski et al., 2009). Based on these findings, we recommended for future studies aimed at understanding illness progression to apply both models.

Malhi, Rosenberg, and Gershon (2014) argued that staging models are only meaningful if there is an understanding of the underlying pathophysiology. While our (van der Markt et al., 2019, 2020, 2021) and other (Goi et al., 2015a, 2015b; Rosa et al., 2014) studies thus far established the clinical utility of the two staging models for BD, underlying mechanisms and corresponding biomarkers have not yet been identified. Structural brain measures, such as global and (sub)cortical volumes, cortical thickness, and cortical surface area may be interesting markers in relation to illness progression. Indeed, when comparing groups of patients with BD with healthy individuals, differences in brain structure are consistently reported. The ENIGMA Bipolar Disorder Working Group published some of the most comprehensive studies on structural brain deviations in BD. They found a widespread pattern of a thinner cortex, but not a smaller surface area in individuals with BD (Hibar et al., 2018). Also, volumes of subcortical regions, such as the hippocampus,

amygdala, and thalamus, have been found to be significantly smaller, alongside larger ventricular volumes (Hibar et al., 2016).

Accelerated brain aging, which has been related to several psychiatric disorders (Cole et al., 2018), may be an additional brain measure of interest. Brain age is a composite brain measure related to aging reflecting the biological age, derived from several regional brain metrics, including volume, cortical thickness, and surface area measurements from structural magnetic resonance imaging (MRI) scans. The relative weighting of the individual brain metrics has been learned by training a machine learning model to predict age from these brain metrics in a large dataset. By design, it thus estimates the biological age of the brain, which can be compared with chronological age. The difference between the biological age and chronological age is called the brain-predicted age difference (brain-PAD), also referred to as the brainage gap or brainAGE (brain Age Gap Estimate). Several larger studies found an increase in brain-PAD for patients with BD compared to controls (de Nooij et al., 2020; Kaufmann et al., 2019; Tønnesen et al., 2020; Van Gestel et al., 2019), implying that the brains of subjects with BD experience accelerated aging. However, these findings were not confirmed in smaller studies (Nenadić, Dietzek, Langbein, Sauer, & Gaser, 2017; Shahab et al., 2019), nor in young participants at risk or in the early stages of BD (Hajek et al., 2019). Although brain-PAD in BD has been studied extensively, none of these studies have controlled for illness progression, thereby neglecting a possible association between the clinical course of the illness in BD and brain-PAD.

The current study aims to explore and gain initial insights into whether global and (sub)cortical brain measures, expressed as volume, cortical thickness, cortical surface area, or brain-PAD may function as biomarkers for illness progression as conceptualized in both staging models in patients with BD-I. We hypothesized (1) a decrease in global and (sub)cortical volumes and a thinner cortex, and (2) an increase of brain-PAD across subsequent stages.

We further hypothesized that the likelihood of being in higher stages was more strongly related to brain-PAD than to individual global and regional (sub)cortical brain measures as the calculated brain age is a composite quantity, derived from tens of individual brain regions related to brain aging, thereby reducing the influence of individual variations in brain morphology and noise.

## Methods

### Study sample

Data were acquired from the Dutch Bipolar Cohort (DBC) (Vreeker et al., 2016). The DBC is a large case-control study, and a collaboration between the University of California Los Angeles (UCLA) and Dutch healthcare institutes of the University Medical Center Utrecht (UMC Utrecht), GGZ Altrecht, GGZ inGeest, University Medical Center Groningen, Delta Center for Mental Health Care, Dimence, Parnassia (PsyQ), and Reinier van Arkel Group, investigating genetic and phenotypic information of subjects with bipolar disorder type I (BD-I), their first-degree relatives, and controls.

Data of 1396 subjects with BD-I, 589 first-degree relatives, and 266 control subjects were collected. Subjects were recruited between June 2011 and April 2015 through various means, i.e. via clinicians, the Dutch BD patient association, pharmacies, advertisements, self-referral, and participation in previous UMCU studies.

Inclusion criteria were: a minimum age of 18, at least three biological grandparents of Dutch ancestry to acquire a homogeneous genetic sample, and a thorough understanding of the Dutch language. For the subjects with BD, an additional requirement was the diagnosis of BD-I as verified by the Structured Clinical Interview for DSM-IV (SCID-I [First, Spitzer, Gibbon, & Williams, 2002]). Inclusion criteria for the controls were no diagnosis of BD, schizophrenia, or any other psychotic disorder, nor in a first-degree relative. Controls were screened for psychiatric diagnoses using the M.I.N.I. (Mini International Neuropsychiatric Interview [Sheehan et al., 1998]). The presence of co-occurring psychiatric conditions, such as substance use disorders and anxiety disorders, did not lead to exclusion from the study.

The assessments were done by a group of carefully trained staff members and consisted of PhD-students, research assistants, trainee psychiatrists, psychologists, and medical students, under the supervision of experienced psychiatrists. Subjects with BD, relatives, and controls received a small allowance (respectively 40, 20, or 20 euros) to cover expenses. The study was approved by the medical ethical committee of the UMC Utrecht and all participants gave written informed consent.

In the current study, we used a subsample of the DBC of 256 subjects with BD and 166 control subjects, of whom high-quality T1-weighted MRI scans were acquired. All subjects with a chronological age above 60 (38 subjects with BD and 19 controls) were excluded from the main analysis as the study by Han et al. (2020) showed higher errors in determining the brain age in individuals over 60 years. The sample thus consisted of 218 subjects with BD and 147 controls.

### Baseline characteristics

For all participants, demographic information was obtained, including age at inclusion, sex, left or right-handedness, and IQ.

The questionnaire for bipolar disorders (QBP [Akkerhuis et al., 2013]), an adaption of the Enrolment Questionnaire as previously used in the Stanley Foundation Bipolar Network by Leverich et al. (2001) and Suppes et al. (2001), was administered to the subjects with BD only, asking questions including age at onset, number of depressive episodes, number of manic and/or hypomanic episodes, and lithium use at the time of MRI-scanning (yes/no).

### Staging

Each subject with BD was assigned to a stage from the models based on the number of episodes (Berk et al., 2007) and based on inter-episodic functioning (Kapczinski et al., 2009). We implemented substages for stage 3c with thresholds at 5 and 10 mood episodes in line with suggestion from a prior study (van der Markt et al., 2020). Subjects were classified into stages using a decision flowchart that was previously applied in two other studies (van der Markt et al., 2020, 2021) (online Supplementary Fig. S1).

To determine the stage in the model based on the number of episodes (Berk et al., 2007), items from the QBP (Akkerhuis et al., 2005) were used. With the QBP, we assessed current or previous episodes and recovery in the last year. Out of the 218 subjects with BD, a total of 199 subjects could be assigned to a stage, resulting in  $n = 6$  for stage 3a,  $n = 12$  for stage 3b,  $n = 33$  for stage 3c (fewer than 5 episodes),  $n = 79$  for stage 3c (6–10 episodes),  $n = 60$  for stage 3c (over 10 episodes),  $n = 9$  for stage 4.

For the model based on inter-episodic functioning (Kapczinski et al., 2009), the GAF-score from the QBP questionnaire, information on current mood episodes, and items on work, work limitations, and independent functioning were used to determine the stage. A total of 199 subjects could be staged, resulting in  $n = 21$  for stage I,  $n = 65$  for stage II,  $n = 98$  for stage III, and  $n = 15$  for stage IV.

Missings ( $n = 19$  for each model) were due to missing items on the questionnaires. See Table 1.

### Brain imaging

Three-dimensional T1-weighted images were acquired on a 3 T Philips Achieva scanner (Philips Healthcare, Best, The Netherlands), equipped with an eight-channel SENSE head coil. Fast field echo scans with 200 contiguous sagittal slices (TE = 4.6 ms, TR = 10 ms, flip angle = 8°, FOV = 240 mm,  $0.75 \times 0.75 \times 0.80$  mm<sup>3</sup> voxels) were made. An independent radiologist evaluated the brain scans of all participants.

Image processing was done on the neuroimaging computer network of the University Medical Center Utrecht, Utrecht, the Netherlands. All images were processed using the FreeSurfer 5.1.0 image analysis suite (<http://surfer.nmr.mgh.harvard.edu>) for cortical reconstruction and segmentation (Fischl & Dale, 2000). According to the findings of Janssen et al. (2022), utilizing this version of FreeSurfer yielded outcomes comparable to the more recent version 6.0, indicating that the older version remains a valid and reliable option for image analysis. T1-weighted images were registered to the Talairach atlas (Laitinen, 1989) and intensity variations were corrected. The image was then skull-stripped and the remaining voxels were classified as white matter or non-white matter based on intensity and neighbor constraints (Ségonne et al., 2004). Cutting planes were computed to separate the hemispheres and remove the cerebellum and brain stem. Any

interior holes in the components representing white matter were filled. The initial triangular tessellation was formed on the surface of this white matter mass to create a surface mesh representation and then smoothed using a deformable surface algorithm to form the grey/white surface. The algorithm was further used to expand the surface to obtain the grey matter–cerebrospinal fluid interface, i.e. the cortical surface (Dale, Fischl, & Sereno, 1999; Fischl & Dale, 2000). Images were registered to a spherical atlas and cortical thickness measures were obtained by calculating the distance between the grey/white matter boundary and the cortical surface at approximately 320 000 points across the cortex (Dale *et al.*, 1999). All images were coded to ensure investigator blindness to subject identification and group.

### Global measures and (sub)cortical volumes

Segmentation of gray and white matter was visually checked and control points were added where necessary. This ensured reliable cortical volume, thickness, and surface area measures. The global brain measures cortical gray and white matter volumes, cerebellum, and lateral ventricles were estimated per hemisphere. Intracranial volume, third ventricle, mean cortical thickness, and total cortical surface area were also estimated. For the volumetric measures, left and right volumes were added up. For the cerebellum, left and right white and gray matter were combined. For the lateral ventricles, left and right inferior and superior lateral ventricles were combined.

In addition, subcortical volumes of the thalamus, caudate nucleus, putamen, pallidum, hippocampus, amygdala, and nucleus accumbens were extracted for each hemisphere separately (Buckner *et al.*, 2004). These were quality checked according to the guidelines provided by the ENIGMA consortium (<http://enigma.ini.usc.edu/>), using Surfscan Visualiser (<http://ibowman.com/surfscan/>). Poorly segmented volumes were removed from the dataset. Volumes of 34 cortical regions were also obtained for each hemisphere based on the Desikan–Killiany cortical atlas (Desikan *et al.*, 2006). For regional cortical and subcortical volumes, the mean was calculated by combining left and right hemisphere volumes, then dividing by two. These means were used for further analysis.

### Brain age and brain-PAD

To calculate the brain age, pre-trained ENIGMA-based models for males and females by Schirrmeyer *et al.*, (2020) were used, freely available from the photon-ai website ([www.photon-ai.com](http://www.photon-ai.com)). Following the example of the ENIGMA paper (Han *et al.*, 2020), we estimated brain age separately for males and females, because of differential brain developmental trajectories.

These ENIGMA-based models calculate the brain age by combining 77 FreeSurfer-derived brain measures, including intracranial volume, the means ( $[(\text{left} + \text{right})/2]$ ) of eight left and right subcortical volumes (ventricles, thalamus, caudate nucleus, putamen, pallidum, hippocampus, amygdala, and nucleus accumbens) and the 34 left and right cortical thickness and surface area measures. The brain age was then estimated using ridge regression from the Python-based *sklearn* package.

Brain-PAD was then calculated by subtracting the chronological age at the time of the scan from the biological brain-age predictions.

### Statistical analysis

All analyses were primarily exploratory in nature and were performed in SPSS version 27 (IBM Corp., 2020). Descriptive characteristics were compared between groups using  $\chi^2$  analyses for dichotomous variables and independent samples *t* tests for normally distributed variables. We used ANOVA and  $\chi^2$  analyses to explore whether the patients in different stages of the staging models differed on baseline characteristics.

### Global and (sub)cortical brain regions

Volume measures were corrected for intracranial volume, sex, and age. The mean cortical thickness measure and cortical surface area were corrected for sex and age. Standardized residuals were saved and used for subsequent analyses. Separate ordinal regression analyses were performed for each brain structure with staging as a dependent ordinal variable, both for the model based on the number of episodes (Berk *et al.*, 2007) and inter-episodic functioning (Kapczynski *et al.*, 2009). Analyses of the global and (sub)cortical areas were corrected for false discovery rate (Benjamini & Hochberg, 1995) due to multiple testing for 15 tests per staging model for the global and subcortical areas and 34 tests per staging model for the cortical areas.

### Brain-PAD

Separate ordinal logistic regression analyses were performed for each model to examine whether the brain-PAD significantly increased over subsequent stages in individuals with BD. Following the example by Han *et al.* (2020), to correct for the influence of chronological age on brain-PAD, age and age<sup>2</sup> were added as covariates to the regression model.

### Sensitivity analyses

To investigate the effect of lithium use (associated with normalizing decreased brain volumes in BD [Hafeman, Chang, Garrett, Sanders, & Phillips, 2012] and lower discrepancy between brain age and chronological age [Van Gestel *et al.*, 2019]), all analyses were repeated by adding the binary variable lithium use (yes [Li+]/no [Li−]) at the time of the MRI scan as a covariate to the models. For exploratory purpose, analyses were repeated in both Li+ and Li− groups separately. To further explore the significant finding of the brain-PAD for the staging model based on inter-episodic functioning, the average brain-PAD per stage was plotted and relevant *t* tests were carried out to investigate group differences.

To examine the impact of current IQ on our results, analyses were repeated adding current IQ as a covariate to the models. Analyses were also repeated including subjects aged over 60 to investigate if excluding subjects over 60 indeed led to a better age prediction, as suggested by a previous paper by Han *et al.* (2020), since there is too much variety in aging not related to BD above this age. Lastly, analyses were repeated correcting for comorbid anxiety disorders and comorbid substance use disorders respectively, to examine their impact on the findings.

## Results

### Characteristics of the study sample

The sociodemographic and psychiatric characteristics of the study sample are listed in Table 2. The mean age differed



**Table 2.** Characteristics of the study sample

|  | Bipolar patients ( <i>N</i> = 218) | Controls ( <i>N</i> = 147) | <i>t</i> test/ $\chi^2$ , <i>p</i> value          |
|--|------------------------------------|----------------------------|---|
| <b>Sociodemographic characteristics</b>                    |                                    |                            |   |
| Age at MRI scan, mean years (s.d.)                         | 44.7 (10.5)                        | 41.8 (12.8)                | <b><i>T</i>(271) = 2.30, <i>p</i> = 0.022</b>     |
| Sex, m/f (%m)  | 105/113 (48.2%)                    | 74/73 (50.3%)              | $\chi^2(1) = 0.17, p = 0.684$                     |
| Handedness, R/L/B (%right)                                 | 187/21/7 (87.0%)                   | 119/19/5 (83.2%)           | $\chi^2(2) = 1.11, p = 0.574$                     |
| Premorbid IQ, mean (s.d.)                                  | 106.5 (9.48)                       | 108.6 (8.63)               | <b><i>T</i>(352) = 2.07, <i>p</i> = 0.039</b>     |
| Current IQ, mean (s.d.)                                    | 100.2 (13.35)                      | 109.3 (15.91)              | <b><i>T</i>(272) = 5.49, <i>p</i> &lt; 0.001</b>  |
| <b>Psychiatric characteristics of the bipolar patients</b> |                                    |                            |   |
| <b>Age at onset (years)</b>                                |                                    |                            |   |
| Manic symptoms   | 27.2 (10.06)                       | n/a                        |   |
| Depressive symptoms  | 23.0 (9.72)                        | n/a                        |   |
| Number of depressive episodes                              | 17.6 (31.9)                        | n/a                        |   |
| Number of hypomanic / manic episodes                       | 10.9 (24.9)                        | n/a                        |   |
| Lithium use <i>N</i> (%) <sup>a</sup>                      | 145 (39.7)                         | n/a                        |   |
| Comorbid anxiety disorders y/n (%yes)                      | 46/205 (18%)                       | n/a                        |   |
| <b>Comorbid substance use disorders</b>                    |                                    |                            |   |
| Alcohol, units/week (s.d.)                                 | 26.4 (49.6)                        | 26.1 (41.8)                | <i>T</i> (222) = -0.39, <i>p</i> = 0.969          |
| Substance abuse, y/n (%yes)                                | 15/200 (7.0%)                      | 7/74 (8.6)                 | $\chi^2(1) = 0.24, p = 0.626$                     |
| Substance dependence, y/n (%yes)                           | 18/196 (8.4%)                      | 5/66 (7.0%)                | $\chi^2(1) = 0.14, p = 0.714$                     |
| <b>Brain age</b>   |                                    |                            |   |
| Brain age, mean years (s.d.)                               | 47.7 (13.3)                        | 41.8 (14.8)                | <b><i>T</i>(363) = 3.98, <i>p</i> = &lt;0.001</b> |
| Brain-PAD, mean years (s.d.)                               | 3.0 (8.54)                         | 0.0 (7.47)                 | <b><i>T</i>(363) = 3.45, <i>p</i> &lt; 0.001</b>  |
| Brain-PAD Li+, mean years (s.d.) <sup>b</sup>              | 1.4 (8.04)                         |                            |   |
| Brain-PAD Li-, mean years (s.d.) <sup>b</sup>              | 6.1 (8.68)                         |                            |   |

<sup>a</sup>At the time of the scan.

<sup>b</sup>Lithium as regression coefficient (OR 1.962, range 1.085–3.550, *p* 0.026).

Significant results in bold and italic.

significantly between subjects with BD and controls (*p* = 0.022), i.e. subjects with BD were on average 2.9 years older. Compared to controls, subjects with BD had a significantly lower current IQ (*p* ≤ 0.001) and premorbid IQ (*p* = 0.039).

For both models, baseline characteristics were compared between subjects in the different stages, see online Supplementary Table S1. For the model based on the number of episodes, the age at the MRI scan differed across the stages with no clear pattern (*p* < 0.001). For the model based on inter-episodic functioning, current IQ decreased significantly across stages (*p* = 0.009).

The sociodemographic and psychiatric characteristics of the sample including subjects over 60 years of age are shown in online Supplementary Table S2. Subjects with BD were significantly older (*p* = 0.026) compared to controls. Subjects with BD had a significantly lower current IQ (*p* ≤ 0.001) but did not differ in premorbid IQ (*p* = 0.234).

### Global measures and subcortical volumes

Table 3 shows the results of the ordinal logistic regression analyses between global and subcortical regions and staging models, and online Supplementary Table S3 shows the results for the cortical volumes. No significant associations were found after applying correction for the false discovery rate.

Online Supplementary Table S7 demonstrates the numbers of participants with and without lithium use, anxiety disorders or substance use disorders per stage. Adding lithium use at the time of MRI-scanning, current IQ, or anxiety disorders as a covariate (online Supplementary Tables S4, S5, and S8 respectively) did not change the pattern of findings for both models, nor did including the subjects with a chronological age over 60 years (online Supplementary Table S6). However, adding comorbid substance use disorders as a covariate did change our findings; now, although odds ratios (ORs) were mostly similar or lower, the association between stages and volumes of lateral ventricles, putamen, pallidum, hippocampus, amygdala, and nucleus accumbens and mean cortical thickness reached significance after correcting for multiple testing.

### Brain aging

On average, subjects with BD showed a higher brain-PAD of 3 years compared with controls, implying increased aging of the brain in this group.

From online Supplementary Fig. S2, it can be observed that there is a rise in brain-PAD throughout the stages for both models. One could make a valid argument that obtaining a bigger sample size might result in statistically significant outcomes.

**Table 3.** Associations between global and subcortical brain measures and brain-PAD and illness progression for both models, with global and subcortical brain measures corrected for intracranial volume (except cortical thickness and cortical surface measures), age, and sex and brain-PAD corrected for age and age<sup>2</sup>

|                              | Model based on the number of episodes |             |       |                | Model based on inter-episodic functioning |             |       |                |
|------------------------------|---------------------------------------|-------------|-------|----------------|---|-------------|-------|----------------|
|                              | Odds ratio                            | Range       | Wald  | <i>p</i> value | Odds ratio                                | Range       | Wald  | <i>p</i> value |
| Global measures              |                                       |             |       |                |   |             |       |                |
| Intracranial volume          | 1.002                                 | 1.001–1.004 | 6.759 | <i>0.009</i>   | 1.000                                     | 0.999–1.002 | 0.279 | 0.597          |
| Cortical gray matter volume  | 0.997                                 | 0.990–1.005 | 0.573 | 0.449          | 0.993                                     | 0.985–1.000 | 3.394 | 0.065          |
| Cortical white matter volume | 0.995                                 | 0.989–1.001 | 2.376 | 0.123          | 0.997                                     | 0.991–1.003 | 0.823 | 0.364          |
| Cerebellum volume            | 0.978                                 | 0.957–0.999 | 4.336 | <i>0.037</i>   | 0.989                                     | 0.968–1.011 | 0.918 | 0.338          |
| Lateral ventricle volume     | 1.042                                 | 1.010–1.075 | 6.574 | <i>0.010</i>   | 1.045                                     | 1.011–1.080 | 6.721 | <i>0.010</i>   |
| Third ventricle volume       | 2.123                                 | 1.110–4.062 | 5.177 | <i>0.023</i>   | 1.360                                     | 0.699–2.645 | 0.819 | 0.365          |
| Mean cortical thickness      | 0.534                                 | 0.120–2.376 | 0.678 | 0.410          | 0.242                                     | 0.048–1.225 | 2.939 | 0.086          |
| Mean surface area            | 1.001                                 | 0.999–1.003 | 0.457 | 0.499          | 0.999                                     | 0.997–1.001 | 0.563 | 0.453          |
| Subcortical volumes          |                                       |             |       |                |   |             |       |                |
| Thalamus                     | 0.973                                 | 0.657–1.440 | 0.019 | 0.890          | 0.919                                     | 0.617–1.368 | 0.173 | 0.677          |
| Caudate nucleus              | 0.825                                 | 0.432–1.573 | 0.343 | 0.558          | 0.931                                     | 0.476–1.819 | 0.044 | 0.834          |
| Putamen                      | 0.736                                 | 0.471–1.150 | 1.811 | 0.178          | 0.586                                     | 0.371–0.926 | 5.230 | <i>0.022</i>   |
| Pallidum                     | 0.529                                 | 0.121–2.303 | 0.720 | 0.396          | 0.301                                     | 0.068–1.334 | 2.499 | 0.114          |
| Hippocampus                  | 0.506                                 | 0.273–0.939 | 4.661 | <i>0.031</i>   | 0.437                                     | 0.229–0.831 | 6.368 | <i>0.012</i>   |
| Amygdala                     | 0.375                                 | 0.120–1.171 | 2.849 | 0.091          | 0.199                                     | 0.060–0.667 | 6.857 | <i>0.009</i>   |
| Nucleus accumbens            | 0.510                                 | 0.136–1.913 | 0.996 | 0.318          | 0.300                                     | 0.075–1.209 | 2.867 | 0.090          |
| Brain-PAD <sup>a</sup>       |                                       |             |       |                |   |             |       |                |
| Brain-PAD                    | 1.015                                 | 0.985–0.895 | 1.044 | 0.344          | 1.033                                     | 1.001–1.065 | 4.187 | <b>0.041</b>   |

<sup>a</sup>Corrected for age and age<sup>2</sup>.*p* < 0.05 in italic.

Significant results in bold.

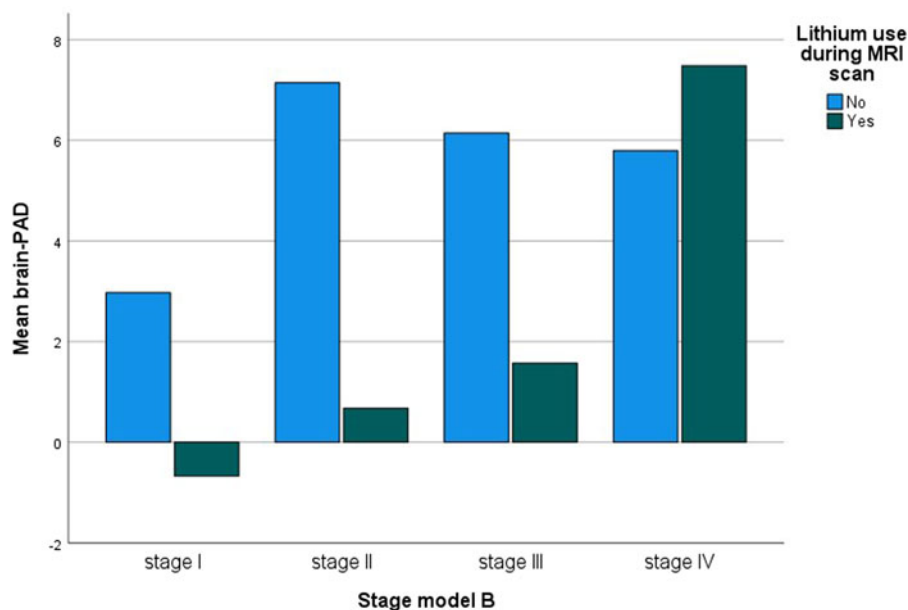
An increase in brain-PAD was not significantly associated with the odds of being in a higher stage for the model based on the number of episodes (OR 1.015, *p* = 0.344; Table 3) (Berk et al., 2007), but was significantly associated for the model based on inter-episodic functioning (Kapczinski et al., 2009) (OR 1.033, *p* = 0.041; Table 3). After correcting for lithium use (online Supplementary Table S4, Fig. S1), an increase in brain-PAD was no longer significantly associated with a higher stage for either model (model based on recurrence *p* = 0.366 and functioning *p* = 0.167). The brain-PAD for Li+ was 1.4 years and for Li– 6.1 years, indicating that the brain age of patients who used lithium is 1.4 years older than their chronological age while in patients not on lithium this is 6.1 years. In other words, the effect of lithium use at the time of MRI on the brain-PAD was 4.7 years (*p* = 0.026) (see Table 2). Figure 1 shows the brain-PAD per stage for Li+ and Li– subjects separately. This figure shows a brain-PAD which increases across the stages for the Li+ subjects and a higher brain-PAD for the Li– subjects. Comparison of stage I, II, and III combined (*n* = 125) to stage IV (*n* = 7) for Li+: mean brain-PAD stage I–III combined was 0.92 years (s.d. 8.02) and for stage IV 7.48 years (s.d. 8.85). Independent sample *t* test showed that this is a significant difference between these groups, *t*(130) = –2094, *p* = 0.019 with a moderate to high effect size, Cohen's *d* = 0.81.

The association between an increase in brain-PAD and a higher stage still reached trend-level significance after correction for IQ

(online Supplementary Table S5), after inclusion of subjects over 60 years of age (online Supplementary Table S6) and after correcting for substance use disorders (online Supplementary Table S9) for the model based on functioning (*p* = 0.056, *p* = 0.057, and *p* = 0.008, respectively) but remained non-significant for the model based on recurrence (*p* = 0.482, *p* = 0.252, and *p* = 0.384, respectively). Correction for comorbid anxiety disorders led to a slight reduction in effect size for the brain-PAD but did not remain significant (from OR 1.033, range 1.001–1.065, *p* = 0.041 to OR 1.025, range 0.993–1.058, *p* = 0.124).

## Discussion

The present study aimed to explore whether global and (sub)cortical brain measures and brain-PAD can be considered as a marker for illness progression in subjects with BD when conceptualizing illness progression in two different staging models. No global or (sub)cortical brain regions were associated with odds of reaching a higher stage. We found evidence, albeit, with a small effect size (OR 1.033), that brain-PAD may function as a marker for illness progression, using a staging model based on impaired inter-episode functioning (Kapczinski et al., 2009) but not in a staging model based largely on the recurrence of mood episodes (Berk et al., 2007). These results are of an exploratory nature and are therefore meant to take a first step in investigating the influence of illness progression in relation to brain aging.



**Figure 1.** Bar charts of the mean brain-PAD for subjects with BD with and without the current use of lithium.N

Previous studies have focused on identifying deviations in global and (sub)cortical brain measures (i.e. volumes, cortical thickness, and cortical surface areas) in subjects with BD compared with healthy controls (Abramovic et al., 2016; Hibar et al., 2016, 2018), but did not study the relation between illness progression and these brain deviations. Indeed, we previously reported a significantly smaller volume of the total brain and the globus pallidum, a thinner cortex in small clusters in frontal, parietal, and cingulate regions (Abramovic et al., 2018), and increased ventricle volumes in subjects with BD as compared with controls in the same cohort (Abramovic et al., 2016, 2018). As we have not identified any areas that significantly changed across the stages, we can now conclude that global or (sub)cortical brain measures are not associated with illness progression, regardless of whether illness progression was based on the recurrence of episodes (Berk et al., 2007) or on impaired inter-episodic functioning (Kapczinski et al., 2009). This may imply that BD-related deviations in individual brain regions occur prior to or right after the onset of BD and do not further alter as illness progresses. However, longitudinal studies are necessary to confirm this hypothesis.

Irrespective of staging, we found a larger brain-PAD of 3 years in subjects with BD compared to controls. In other words, the brains of subjects with BD are estimated to be on average 3 years older than the chronological age suggests. Although several earlier smaller studies did not find a significant brain-PAD difference in subjects with BD compared to controls (Nenadić et al., 2017; Shahab et al., 2019), larger studies found a brain-PAD difference ranging from 0.29 to 4.28 years (de Nooij et al., 2020; Kaufmann et al., 2019; Tønnesen et al., 2020; Van Gestel et al., 2019). This variability in brain-PAD could potentially be attributed to the differences between lithium and non-lithium users, as Van Gestel et al. (2019) reported a brain-PAD of non-lithium users of 4.28 years and for lithium users of 0.48 year.

The brain-PAD significantly increased with illness progression in the model based on impaired inter-episodic functioning (Kapczinski et al., 2009), but not in the model based on the recurrence of mood episodes (Berk et al., 2007). This suggests that brain-PAD may be specifically related to the level of functioning in-between episodes. Our previous papers on these staging models

already suggested that both models reflect different aspects of illness progression in BD since they had a low correlation (van der Markt et al., 2020) and differences in associated clinical markers (van der Markt et al., 2021). The markers that we found to be associated with illness progression as reflected by the staging models had found to be related to volume or thickness alterations in certain brain areas in other studies, e.g. familial risk (de Zwart, Brouwer, Agartz, & van Haren, 2019), childhood abuse (Cassiers et al., 2018), illness duration (Vita, De Peri, & Sacchetti, 2009), psychiatric comorbidity (Biederman et al., 2008), and cognitive functioning expressed as IQ (Pietschnig, Penke, Wicherts, Zeiler, & Voracek, 2015). It is important to note that we do not know the causality in the relation between brain-PAD and illness progression expressed as functioning. To identify whether an increase in brain-PAD causes a poorer level of functioning or whether poor functioning precedes a more pronounced brain-PAD, longitudinal studies are needed that cover the full adult age range.

A significant difference across the stages for the model based on functioning was found in brain-PAD, but not for individual global and regional (sub)cortical brain measures. This may be explained by brain age being a composite quantity, derived from many individual brain regions related to brain aging, thus reducing the influence of individual variations in brain morphology and noise. Individual brain areas for this model showed an effect in the same direction, including a larger volume of the lateral ventricles, and a smaller volume of the putamen, hippocampus, and amygdala, however did not reach significance after correction for multiple testing.

We investigated whether our findings could be explained by a difference in baseline characteristics between the stages. Not surprisingly, a higher stage in the staging model based on episode recurrence was associated with increased age. However, since age was introduced as a covariate for the global and (sub)cortical measures and age and age-squared were included as covariates in our brain age analyses, we expect that these associations did not confound our results. After including individuals with BD aged over 60 years, the difference in brain-PAD across the stages for the model based on functioning went to trend-level significance (online Supplementary Table S6), confirming the model's lower

accuracy in higher age groups, as was already suggested by Han et al. (2020).

Correcting for IQ did not change our results for the global and subcortical brain measures (online Supplementary Table S5). It did marginally change the results of the brain-PAD for the model based on inter-episodic functioning from significant ( $p=0.041$ ) to (just) not significant ( $p=0.056$ ). Stages in this model use criteria such as the ability to work and function independently, which are closely related to IQ, as subjects with a higher IQ are described to have more cognitive reserve (Martino, Valerio, Szmulewicz, & Strejilevich, 2017) and thus have more possibilities to continue working less demanding jobs and have more cognitive capacity to remain functioning autonomously. This is in line with the findings by Reinares et al. (2013), who found a relation between verbal IQ and better functioning, defined as the ability to work and function socially in subjects with BD.

We found that correcting for current lithium use did not change our results for the global and subcortical brain measures. However, correcting for lithium did change the results of the brain-PAD for the model based on impaired functioning from significant (OR 1.033, range 1.001–1.065,  $p=0.041$ ) to non-significant (OR 1.023, range 0.991–1.056,  $p=0.167$ ). The share of lithium users decreased across the stages of the functioning-based model, although not significantly ( $\chi^2[3,199] 7.40$ ,  $p=0.060$ ). This indicates that lithium use unlikely explains the effect of staging on brain-PAD. The change in significance when correcting for lithium is mostly caused by the influence of lithium on the brain-PAD itself. When assessing the lithium users and non-lithium-users separately, the average brain-PADs of these groups differ 4.7 years (1.4 years larger than controls for Li+ and 6.1 years larger than controls for Li-). A possible explanation would be the effect of lithium on the brain itself, which is in line with the study by Van Gestel et al. (2019) who found a normalizing effect on the brain age in lithium users compared to non-lithium users. It is also consistent with the effect of lithium on brain volumes, where lithium users do not (or to a lesser extent) show the smaller gray matter volumes that are found in non-lithium users as compared to controls, as shown by Abramovic et al. (2016) in the current sample and by many others (for review, see Hafeman et al., 2012). It has been suggested that gray matter volume reduction in BD is caused by actual brain deterioration through different pathways, such as modulation of nerve growth factors, inflammation, mitochondrial function, oxidative stress, and programmed cell death mechanisms, such as autophagy and apoptosis, and lithium possibly serve as a neuroprotective agent by influencing these pathways (Ochoa, 2022). Figure 1 shows the brain-PAD per stage in lithium users and non-lithium users for the model based on inter-episodic functioning. For lithium users, the protective effect of lithium is present in stages I, II, and III but decreases significantly in stage IV. This might mean that the protective effect of lithium is no longer present when functioning is severely affected. Due to the small sample size, further studies are needed to confirm this hypothesis. Alternatively, patients not on lithium could be more treatment resistant, as lithium is the first choice in the Dutch clinical guidelines, and thus experience more severe mood episodes leading to more rapid aging. In conclusion, the use of lithium may be a protective factor as it possibly prevents brain volume loss in subjects with BD and it may explain the less divergent brain-PAD compared to controls after correcting for lithium.

We did not assess the effect of other psychotropic drugs. A study by Abramovic et al. (2016) in the current cohort found

that the use of antipsychotics in BD was subtly associated with a larger third ventricle and smaller hippocampus and supramarginal cortex volume, albeit with smaller effect sizes than the more extensive effects of lithium on the brain. Indeed, the review by Hafeman et al. (2012) found that the use of antipsychotics was generally not associated with volumetric differences. However, the ENIGMA study found a reduced cortical surface area for atypical antipsychotics and an increased cortical surface area related to typical antipsychotics, suggesting that different types of medication are differently associated with brain measures (Hibar et al., 2018), although the authors acknowledge that the effect may be related to the small number of subjects using typical antipsychotics in their study. Due to these varying findings, it may be interesting for future studies to assess the influence of neuroleptic drugs on the brain from the perspective of illness progression.

In the main analyses, we intentionally did not exclude individuals with co-occurring anxiety and substance use disorders to maintain the naturalistic nature of our outpatient sample. While excluding such subjects would improve sample homogeneity, it would compromise the representativeness of our outcomes for real-world outpatients. In order to ensure comprehensive analysis, we corrected our models for the presence or absence of anxiety and substance abuse to explore their potential impact on the observed associations between brain-PAD and staging models. The effect sizes only showed minimal changes as compared to the analyses in the full sample, but some significance levels changed, e.g. after correcting for the presence of anxiety disorders the association between the staging model based on functioning and the brain-PAD did not retain significance. Whether the presence of anxiety disorders had a significant influence on this association needs further study in a larger sample. It needs emphasis that comorbid anxiety disorders are common in a naturalistic sample and that this might just be an indicator of the severity of BD.

Also, correction for substance use disorders did alter the pattern of findings for the brain regions. Now, although odds ratios were again mostly similar, the association between stages in the model based on functioning and lateral ventricles, mean cortical thickness, putamen, pallidum, hippocampus, amygdala, and nucleus accumbens reached significance after correcting for multiple testing, i.e. thickness and volumes decreased with an increase in stages. Notably, almost all subjects with substance use disorders clustered in stages II and III. As substance use disorders are relatively common comorbid to BD, it is difficult to explain why findings now reached the significance threshold. Again, further studies with larger cohorts are needed to determine the consequences of these findings.

To our knowledge, this is the first study to address the potential underlying neurobiology of staging models. Analyses were done in a relatively large sample. However, some limitations must be considered. We found a large variation in the predicted brain age within the stages of both staging models. This means an even larger sample may be needed to better estimate the mean brain age and its standard deviation per stage, since differences in the brain measures across stages may be too small to detect in a sample of our size. The model based on the recurrence of episodes (Berk et al., 2007) consists of nine categories of which the six highest categories of the model were assigned to the subjects with BD, as the cohort did not include individuals in the prodromal and very early stages of BD. In contrast, the model based on inter-episodic functioning (Kapczinski et al., 2009) consists of four categories. The implication of a difference in the number of stages per model may be that it becomes more likely



to find significant differences in variables across the stages for the model based on functioning as sample sizes per stage are larger. On the other hand, a larger number of stages could lead to a more refined grouping, which decreased within-stage variation, thus leading to more statistically significant results when testing the differences in variables across the stages. The reported number of previous mood episodes may be subject to recall bias, which may have led to some inaccuracy in the demographics table. Also, our outpatient sample may be less suitable for testing the staging model based on the number of episodes, as an outpatient sample does not typically contain patients in the very early stages of bipolar illness as they are not enrolled in treatment settings yet. Therefore, our cohort contained a large number of subjects clustering in similar stages. Lastly, our results are meant to be exploratory in nature. Our results need to be interpreted with caution as correction for multiple testing was only performed for each model separately and not for the two models combined.

In conclusion, our study provides suggestive evidence that brain-PAD may be associated with illness progression in the model based on impaired inter-episodic functioning (Kapczinski et al., 2009). We found a significant association between illness progression reflected by this staging model and the brain-PAD, however, the role of lithium appears relevant as correcting for lithium led to this association no longer being significant. This relationship between brain-PAD and stage of inter-episodic functioning should therefore be studied in larger and longitudinal cohorts. This study emphasized the importance of accounting for lithium-use in imaging studies in BD, as lithium was associated with less brain aging. Additionally, it emphasized the role of comorbid anxiety disorders as the association between brain aging and the staging model no longer reached significance after correction for subjects with anxiety disorders. New studies with larger samples are needed to further disentangle the complex relationship between brain-PAD, stage, and lithium intake or anxiety disorders.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291723002829>

**Acknowledgments.** This project was partly funded by the Dutch Research Agenda (NWA – Nationale Wetenschapsagenda), part of the Dutch Research Council (NWO; Nederlandse Organisatie voor Wetenschappelijk Onderzoek), file number 1228.192.002, received by A. v. d. M., N. K., and H. S. and the National Institute of Mental Health grants R01 MH090553 to R. A. O. Funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

**Competing interests.** None.

## References

- Abramovic, L., Boks, M. P. M., Vreeker, A., Bouter, D. C., Kruiper, C., Verkooijen, S., ... van Haren, N. E. M. (2016). The association of antipsychotic medication and lithium with brain measures in patients with bipolar disorder. *European Neuropsychopharmacology*, 26(11), 1741–1751. <https://doi.org/10.1016/j.euroneuro.2016.09.371>
- Abramovic, L., Boks, M. P. M., Vreeker, A., Verkooijen, S., van Bergen, A. H., Ophoff, R. A., ... van Haren, N. E. M. (2018). White matter disruptions in patients with bipolar disorder. *European Neuropsychopharmacology*, 28(6), 743–751. <https://doi.org/10.1016/j.euroneuro.2018.01.001>
- Akkerhuis, G. W., Nolen, W. A., Kupka, R. W., Drexhage, H. A., Fekkes, D., & Wiersma, J. E. (2013). Development of a questionnaire for bipolar phenotypes (QBP-NL): a pilot study. *Journal of Affective Disorders*, 150(3), 1049–1053.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Berk, M., Conus, P., Lucas, N., Hallam, K., Malhi, G. S., Dodd, S., ... McGorry, P. (2007). Point of view setting the stage: From prodrome to treatment resistance in bipolar disorder. *Bipolar Disorders*, 9, 671–678. <https://doi.org/10.1111/j.1399-5618.2007.00484.x>
- Biederman, J., Makris, N., Valera, E. M., Monuteaux, M. C., Goldstein, J. M., Buka, S., ... Seidman, L. J. (2008). Towards further understanding of the co-morbidity between attention deficit hyperactivity disorder and bipolar disorder: A MRI study of brain volumes. *Psychological Medicine*, 38(7), 1045–1056. <https://doi.org/10.1017/S0033291707001791>
- Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., & Snyder, A. Z. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: Reliability and validation against manual measurement of total intracranial volume. *NeuroImage*, 23(2), 724–738. <https://doi.org/10.1016/j.neuroimage.2004.06.018>
- Cassiers, L. L. M., Sabbe, B. G. C., Schmaal, L., Veltman, D. J., Penninx, B. W. J. H., & Van Den Eede, F. (2018). Structural and functional brain abnormalities associated with exposure to different childhood trauma subtypes: A systematic review of neuroimaging findings. *Frontiers in Psychiatry*, 9(8), 1–17. <https://doi.org/10.3389/fpsyg.2018.00329>
- Cole, J. H., Ritchie, S. J., Bastin, M. E., Valdés Hernández, M. C., Muñoz Maniega, S., Royle, N., ... Deary, I. J. (2018). Brain age predicts mortality. *Molecular Psychiatry*, 23(5), 1385–1392. <https://doi.org/10.1038/mp.2017.62>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- de Nooij, L., Harris, M. A., Hawkins, E. L., Clarke, T. K., Shen, X., Chan, S. W. Y., ... Whalley, H. C. (2020). Longitudinal trajectories of brain age in young individuals at familial risk of mood disorder from the Scottish Bipolar Family Study. *Wellcome Open Research*, 4, 206. <https://doi.org/10.12688/wellcomeopenres.15617.2>
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., ... Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- de Zwarte, S. M. C., Brouwer, R. M., Agartz, I., & van Haren, N. E. M. (2019). The association between familial risk and brain abnormalities is disease specific: An ENIGMA-relatives study of schizophrenia and bipolar disorder. *Biological Psychiatry*, 86(7), 545–556. <https://doi.org/10.1016/j.biopsych.2019.03.985>
- El-Badri, S. M., Ashton, C. H., Moore, P. B., Marsh, V. R., & Ferrier, I. N. (2001). Electrophysiological and cognitive function in young euthymic patients with bipolar affective disorder. *Bipolar Disorders*, 3, 79–87. <https://doi.org/10.1034/j.1399-5618.2001.030206.x>
- First, M., Spitzer, R., Gibbon, M., & Williams, J. (2002). *Structured clinical interview for DSM-IV axis I disorders-patient edition (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute. <https://doi.org/10.1002/9781118625392.wbecp351>
- Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the USA*, 97(20), 11050–11055. <https://doi.org/10.1073/pnas.200033797>
- Fries, G. R., Bauer, I. E., Scaini, G., Wu, M. J., Kazimi, I. F., Valvassori, S. S., ... Quevedo, J. (2017). Accelerated epigenetic aging and mitochondrial DNA copy number in bipolar disorder. *Translational Psychiatry*, 7(12), 1283. <https://doi.org/10.1038/s41398-017-0048-8>
- Fries, G. R., Pfaffenseller, B., Stertz, L., Paz, A. V. C., Dargél, A. A., Kunz, M., & Kapczinski, F. (2012). Staging and neuroprogression in bipolar disorder. *Current Psychiatry Reports*, 14(6), 667–675. <https://doi.org/10.1007/s11920-012-0319-2>
- Goi, P. D., Bücker, J., Vianna-Sulzbach, M., Rosa, A. R., Grande, I., Chendo, I., ... Massuda, R. (2015a). Pharmacological treatment and staging in bipolar disorder: Evidence from clinical practice. *Revista Brasileira de Psiquiatria*, 37(2), 121–125. <https://doi.org/10.1590/1516-4446-2014-1554>
- Goi, P. D., Vianna-Sulzbach, M., Silveira, L., Grande, I., Chendo, I., Sodré, L. A., ... Gama, C. S. (2015b). Treatment delay is associated with more episodes

- and more severe illness staging progression in patients with bipolar disorder. *Psychiatry Research*, 227(2–3), 372–373. <https://doi.org/10.1016/j.psychres.2015.04.004>
- Hafeman, D. M., Chang, K. D., Garrett, A. S., Sanders, E. M., & Phillips, M. L. (2012). Effects of medication on neuroimaging findings in bipolar disorder: An updated review. *Bipolar Disorders*, 14(4), 375–410. <https://doi.org/10.1111/j.1399-5618.2012.01023.x>
- Hajek, T., Franke, K., Kolenic, M., Capkova, J., Matejka, M., Propper, L., ... Alda, M. (2019). Brain age in early stages of bipolar disorders or schizophrenia. *Schizophrenia Bulletin*, 45(1), 191–198. <https://doi.org/10.1093/schbul/sbx172>
- Han, L. K. M. K., Dinga, R., Hahn, T., Ching, C. R. K., Eyler, L. T., Aftanas, L., ... Schmaal, L. (2020). Brain aging in major depressive disorder: Results from the ENIGMA major depressive disorder working group. *Molecular Psychiatry*, 26, 5124–5139. <https://doi.org/10.1101/560623>
- Hibar, D. P., Westlye, L. T., Doan, N. T., Jahanshad, N., Cheung, J. W., Ching, C. R. K., ... Andreassen, O. A. (2018). Cortical abnormalities in bipolar disorder: An MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Molecular Psychiatry*, 23(4), 932–942. <https://doi.org/10.1038/mp.2017.73>
- Hibar, D. P., Westlye, L. T., Van Erp, T. G. M., Rasmussen, J., Leonardo, C. D., Faskowitz, J., ... Andreassen, O. A. (2016). Subcortical volumetric abnormalities in bipolar disorder. *Molecular Psychiatry*, 21(12), 1710–1716. <https://doi.org/10.1038/mp.2015.227>
- IBM Corp. (2020). *IBM SPSS advanced statistics for windows, version 27.0*. Armonk, NY: IBM Corp. <https://doi.org/10.1080/02331889108802322>
- Janssen, J., Alloza, C., Díaz-Caneja, C. M., Santonja, J., Pina-Camacho, L., Gordaliza, P. M., ... Schnack, H. G. (2022). Longitudinal allometry of sulcal morphology in health and schizophrenia. *The Journal of Neuroscience*, 42(18), 3704–3715. <https://doi.org/10.1523/JNEUROSCI.0606-21.2022>
- Kapczynski, F., Dias, V. V., Kauer-Sant'Anna, M., Frey, B. N., Grassi-Oliveira, R., Colom, F., & Berk, M. (2009). Clinical implications of a staging model for bipolar disorders. *Expert Review of Neurotherapeutics*, 9(7), 957–966. <https://doi.org/10.1586/ern.09.31>
- Kaufmann, T., van der Meer, D., Doan, N. T., Schwarz, E., Lund, M. J., Agartz, I., ... Westlye, L. T. (2019). Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nature Neuroscience*, 22(10), 1617–1623. <https://doi.org/10.1038/s41593-019-0471-7>
- Kraepelin, E. (1921). *Manic-depressive insanity and paranoia*. (R. M. Barclay & G. M. Robertson, Trans.). Edinburgh, Scotland: E. & S. Livingstone.
- Laitinen, L. (1989). Co-planar stereotaxic atlas of the human brain. *Clinical Neurology and Neurosurgery*, 91(3), 277–278. [https://doi.org/10.1016/0303-8467\(89\)90128-5](https://doi.org/10.1016/0303-8467(89)90128-5)
- Leverich, G. S., Nolen, W. A., Rush, A. J., McElroy, S. L., Keck, P. E., Denicoff, K. D., Suppes, T., ... Post, R. M. (2001). The Stanley Foundation Bipolar Treatment Outcome Network: I. Longitudinal methodology. *Journal of Affective Disorders*, 67(1–3), 33–44. [https://doi.org/10.1016/S0165-0327\(01\)00430-X](https://doi.org/10.1016/S0165-0327(01)00430-X)
- Malhi, G. S., Rosenberg, D. R., & Gershon, S. (2014). Staging a protest! *Bipolar Disorders*, 16(7), 776–779. <https://doi.org/10.1111/bdi.12254>
- Martino, D. J., Valerio, M. P., Szmulewicz, A. G., & Strejilevich, S. A. (2017). The effect of premorbid intelligence on neurocognitive and psychosocial functioning in bipolar disorder. *Journal of Affective Disorders*, 210, 226–229. <https://doi.org/10.1016/j.jad.2016.12.053>
- Nenadić, I., Dietzek, M., Langbein, K., Sauer, H., & Gaser, C. (2017). BrainAGE score indicates accelerated brain aging in schizophrenia, but not bipolar disorder. *Psychiatry Research - Neuroimaging*, 266, 86–89. <https://doi.org/10.1016/j.psychres.2017.05.006>
- Ochoa, E. L. M. (2022). Lithium as a neuroprotective agent for bipolar disorder: An overview. *Cellular and Molecular Neurobiology*, 42(1), 85–97. <https://doi.org/10.1007/s10571-021-01129-9>
- Pietschnig, J., Penke, L., Wicherts, J. M., Zeiler, M., & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience and Biobehavioral Reviews*, 57, 411–432. <https://doi.org/10.1016/j.neubiorev.2015.09.017>
- Reinares, M., Papachristou, E., Harvey, P., Mar Bonnín, C., Sánchez-Moreno, J., Torrent, C., ... Frangou, S. (2013). Towards a clinical staging for bipolar disorder: Defining patient subtypes based on functional outcome. *Journal of Affective Disorders*, 144(1–2), 65–71. <https://doi.org/10.1016/j.jad.2012.06.005>
- Rosa, A. R., Magalhães, P. V. S. S., Czepielewski, L., Sulzbach, M. V., Goi, P. D., Vieta, E., ... Kapczynski, F. (2014). Clinical staging in bipolar disorder: Focus on cognition and functioning. *Journal of Clinical Psychiatry*, 75(5), e450–e456. <https://doi.org/10.4088/JCP.13m08625>
- Sanchez-Moreno, J., Martinez-Aran, A., Tabarés-Seisdedos, R., Torrent, C., Vieta, E., & Ayuso-Mateos, J. L. (2009). Functioning and disability in bipolar disorder: An extensive review. *Psychotherapy and Psychosomatics*, 78(5), 285–297. <https://doi.org/10.1159/000228249>
- Schirmer, J., Schulte, P., Kessel, M., & Strobel, G. (2020). PHOTONAI: A high level Python API for machine learning pipelines. *Journal of Open Source Software*, 5(54), 2713.
- Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22(3), 1060–1075. <https://doi.org/10.1016/j.neuroimage.2004.03.032>
- Shahab, S., Mulsant, B. H., Levesque, M. L., Calarco, N., Nazeri, A., Wheeler, A. L., ... Voineskos, A. N. (2019). Brain structure, cognition, and brain age in schizophrenia, bipolar disorder, and healthy controls. *Neuropsychopharmacology*, 44(5), 898–906. <https://doi.org/10.1038/s41386-018-0298-z>
- Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., ... Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry*, 59(SUPPL. 20), 22–33. Retrieved from [https://www.psychiatrist.com/wp-content/uploads/2021/02/15175\\_mini-international-neuropsychiatric-interview-mini.pdf](https://www.psychiatrist.com/wp-content/uploads/2021/02/15175_mini-international-neuropsychiatric-interview-mini.pdf)
- Suppes, T., Leverich, G. S., Keck, P. E., Nolen, W. A., Denicoff, K. D., Altschuler, L. L., ... Post, R. M. (2001). The Stanley Foundation Bipolar Treatment Outcome Network II. Demographics and illness characteristics of the first 261 patients. *Journal of Affective Disorders*, 67(1–3), 45–59. [https://doi.org/10.1016/S0165-0327\(01\)00432-3](https://doi.org/10.1016/S0165-0327(01)00432-3)
- Tønnesen, S., Kaufmann, T., de Lange, A. M. G., Richard, G., Doan, N. T., Alnæs, D., ... Westlye, L. T. (2020). Brain age prediction reveals aberrant brain white matter in schizophrenia and bipolar disorder: A multisample diffusion tensor imaging study. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 5(12), 1095–1103. <https://doi.org/10.1016/j.bpsc.2020.06.014>
- van der Markt, A., Klumpers, U. M. H., Dols, A., Boks, M. P., Draisma, A., van Bergen, S., ... Kupka, R. W. (2020). Exploring the clinical utility of two staging models for bipolar disorder. *Bipolar Disorders*, 22(1), 38–45. <https://doi.org/10.1111/bdi.12825>
- van der Markt, A., Klumpers, U. M. H., Dols, A., Boks, M. P., Vreeker, A., Beekman, A. T., & Kupka, R. W. (2021). Clinical profiles of subsequent stages in bipolar disorder: Results from the Dutch Bipolar Cohort. *Bipolar Disorders*, 24, 424–433. <https://doi.org/10.1111/bdi.13159>
- van der Markt, A., Klumpers, U. M. H., Draisma, S., Dols, A., Nolen, W. A., Post, R. M., ... Kupka, R. W. (2019). Testing a clinical staging model for bipolar disorder using longitudinal life chart data. *Bipolar Disorders*, 21(3), 228–234. <https://doi.org/10.1111/bdi.12727>
- Van Gestel, H., Franke, K., Petite, J., Slaney, C., Garnham, J., Helmick, C., ... Hajek, T. (2019). Brain age in bipolar disorders: Effects of lithium treatment. *Australian and New Zealand Journal of Psychiatry*, 53(12), 1179–1188. <https://doi.org/10.1177/0004867419857814>
- Vita, A., De Peri, L., & Sacchetti, E. (2009). Gray matter, white matter, brain, and intracranial volumes in first-episode bipolar disorder: A meta-analysis of magnetic resonance imaging studies. *Bipolar Disorders*, 11(8), 807–814. <https://doi.org/10.1111/j.1399-5618.2009.00759.x>
- Vreeker, A., Boks, M. P. M., Abramovic, L., Verkooijen, S., Van Bergen, A. H., Hillegers, M. H. J., ... Ophoff, R. A. (2016). High educational performance is a distinctive feature of bipolar disorder: A study on cognition in bipolar disorder, schizophrenia patients, relatives and controls. *Psychological Medicine*, 46(4), 807–818. <https://doi.org/10.1017/S0033291715002299>