



OPEN

# Diagnostic accuracy of brain age prediction in a memory clinic population and comparison with clinically available volumetric measures

Karin Persson<sup>1,2</sup>✉, Esten H. Leonardsen<sup>3,4</sup>, Trine Holt Edwin<sup>2</sup>, Anne-Brita Knapskog<sup>2</sup>, Gro Gujord Tangen<sup>1,2</sup>, Geir Selbæk<sup>1,2,5</sup>, Thomas Wolfers<sup>3,4,6</sup>, Lars T. Westlye<sup>3,4</sup> & Knut Engedal<sup>1,2</sup>

The aim of this study was to assess the diagnostic validity of a deep learning-based method estimating brain age based on magnetic resonance imaging (MRI) and to compare it with volumetrics obtained using NeuroQuant (NQ) in a clinical cohort. Brain age prediction was performed on minimally processed MRI data using deep convolutional neural networks and an independent training set. The brain age gap (difference between chronological and biological age) was calculated, and volumetrics were performed in 110 patients with dementia (Alzheimer's disease, frontotemporal dementia (FTD), and dementia with Lewy bodies), and 122 with non-dementia (subjective and mild cognitive impairment). Area-under-the-curve (AUC) based on receiver operating characteristics and logistic regression analyses were performed. The mean age was 67.1 (9.5) years and 48.7% (113) were females. The dementia versus non-dementia sensitivity and specificity of the volumetric measures exceeded 80% and yielded higher AUCs compared to BAG. The explained variance of the prediction of diagnostic stage increased when BAG was added to the volumetrics. Further, BAG separated patients with FTD from other dementia etiologies with > 80% sensitivity and specificity. NQ volumetrics outperformed BAG in terms of diagnostic discriminatory power but the two methods provided complementary information, and BAG discriminated FTD from other dementia etiologies.

The number of patients suffering from dementia is rapidly increasing worldwide as populations grow older. To date no curative treatment is available for any dementia disorder<sup>1</sup>. However, new treatment possibilities targeted at Alzheimer's disease (AD) are evolving, prompting the need for tools to aid early diagnosis and predict future cognitive decline<sup>2</sup>. This is of utmost importance as potentially disease-modifying treatment strategies will target early pathophysiological changes.

The use of artificial intelligence (AI) to improve and facilitate early diagnosis, planning, and follow-up of treatment has emerged rapidly in many medical fields, including the field of cognitive impairment and dementia<sup>3</sup>. A recent report published by the Norwegian directorate of health encouraged increased use of AI in radiology to improve early diagnosis and provide decision support<sup>4</sup>.

AI based methods for the assessment of brain age based on magnetic resonance imaging (MRI) scans have recently evolved<sup>5</sup>. The purpose of such data-driven methods is to train models to identify characteristics in the MRI data that are robustly associated with age (or any other key characteristics) in a training set, and then apply the resulting model on a different set of brain scans to estimate the age of individual participants or patients in a clinical context. The discrepancy between the predicted and the chronological age, sometimes referred to as

<sup>1</sup>The Norwegian National Centre for Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway. <sup>2</sup>Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway. <sup>3</sup>Department of Psychology, University of Oslo, Oslo, Norway. <sup>4</sup>Norwegian Centre for Mental Disorders Research (NORMENT), Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway. <sup>5</sup>Faculty of Medicine, University of Oslo, Oslo, Norway. <sup>6</sup>Department of Psychiatry and Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Tübingen, Germany. ✉email: Karin.persson@aldringoghelse.no

the brain age gap (BAG), can then be used as a proxy of the integrity and health of the individual brain<sup>5</sup>. Most previous brain age studies have pursued feature-based machine learning techniques, which typically require extensive and computationally demanding image processing and feature extraction<sup>5</sup>. However, recent increases in available training data and computational developments have enabled approaches based on deep learning techniques that provide predictions based on minimally processed MRI data. Deep learning techniques may thus enable both more accurate and faster predictions with minimal computational engineering, which facilitates the implementation in a clinical setting<sup>5</sup>. Previous studies have found evidence for higher brain age in AD compared to healthy controls and promising findings for the prediction of disease progression, typically using traditional machine learning<sup>6–8</sup>, but also recently using deep learning<sup>9,10</sup>. A general limitation is that previously performed deep learning studies were not based on patients recruited from everyday clinical practice.

In the present study, we used a recently established deep learning method<sup>11</sup> to estimate brain age and the corresponding BAG using minimally processed structural MRI brain scans from a heterogeneous memory clinic cohort comprising patients with dementia, mild cognitive impairment (MCI) or subjective cognitive decline (SCD), diagnosed using current clinical criteria. The main purpose was to compare the diagnostic sensitivity and specificity of BAG with conventional brain volumetrics obtained using the clinically available, and previously validated, software NeuroQuant 3. version (NQ, CorTechs labs/University of California, San Diego, CA, USA)<sup>12,13</sup>, using disease stage and etiological diagnosis based on clinical criteria as the gold standard.

## Results

**Sample characteristics.** Characteristics of the patients are presented in Table 1. Mean age was 67.1 (SD 9.5) years and 48.7% (113) were females. Of the 232 included patients, 110 had dementia (76 with AD, 6 FTD, 12 DLB, and 16 with other or non-specific dementia) and 122 had non-dementia (45 with SCD and 77 with MCI) with a mean MMSE of 23.5 (SD 5.0) and 28.4 (SD 1.8) respectively, ( $p < 0.001$ ). Patients with dementia were older and had fewer years of education as compared to patients with non-dementia ( $p < 0.001$ ).

**Brain age prediction.** Across groups, age prediction accuracy was high, with a correlation between predicted and chronological age of 0.879 and a mean absolute error (MAE) of 4.29. Within groups the correlation was 0.913 (MAE 3.58) in non-dementia and 0.750 (MAE 5.09) in dementia. Across groups, the correlation between BAG and age was  $-0.046$  ( $p$  0.483);  $0.092$  ( $p$  0.313) in non-dementia and  $-0.503$  ( $p < 0.001$ ) in dementia.

Diagnostic associations with brain MRI features.

Group-wise summary stats for the MRI features are presented in Table 1. Linear models adjusted for age and sex revealed patients with dementia had higher BAG ( $t = 5.23$ ,  $p < 0.001$ ), smaller forebrain parenchyma volume ( $t = -6.67$ ,  $p \leq 0.001$ ), and smaller hippocampi ( $t = -7.36$ ,  $p = < 0.001$ ) compared to non-dementia patients (unadjusted Cohen's  $d$  of  $-0.59$ ,  $1.27$ , and  $1.34$  respectively).

Table 2 shows the results from the ROC analysis. AUCs were overall higher for the dementia versus SCD classification compared to the dementia versus non-dementia classification. The two NQ measures yielded higher AUCs for both stage classifications compared to BAG, with non-overlapping confidence intervals in the dementia versus non-dementia classification. For the dementia versus SCD classification, the NQ AUCs were both 0.89 (sensitivity 80%, specificity 86–88%) and BAG AUC was 0.78 (sensitivity 80%, specificity 67%). For the dementia versus non-dementia classification, the NQ AUCs were 0.82–0.83 (sensitivity 80%, specificity 67–68%) and BAG AUC was 0.68 (sensitivity 80%, specificity 48%).

Table 3 presents the results from the logistic regression analyses predicting disease stages. The model including hippocampus volume adjusted for demographic covariates gave the highest Nagelkerke  $R^2$  (0.44 vs. 0.40 and 0.36) for dementia/non-dementia prediction, while the model with BAG adjusted for demographic covariates revealed the highest Nagelkerke  $R^2$  for dementia/SCD prediction (0.60 vs. 0.58 and 0.59). Adding white matter hypointensity volume (WMH) to model 3 of both diagnostic predictions did not change the Nagelkerke

	SCD, n 45	MCI, n 77	Dementia, n 110	$p^*$	Non-dementia, n 122	$p^{**}$
Age, years	61.8 (8.9)	64.6 (9.9)	71.0 (7.7)	<b>&lt;0.001</b>	63.6 (9.6)	<b>&lt;0.001</b>
Female, n (%)	28 (62.2)	31 (40.3)	54 (49.1)	0.064	59 (48.4)	0.912
Education, years, n 212	15.0 (4.1)	14.3 (3.7)	13.3 (3.7)	<b>0.041</b>	14.6 (3.9)	<b>0.019</b>
MMSE score, n 212	29.1 (1.2)	28.0 (1.9)	23.5 (5.0)	<b>&lt;0.001</b>	28.4 (1.8)	<b>&lt;0.001</b>
CDR-SB, n 222	0.2 (0.3)	1.1 (0.7)	4.8 (2.5)	<b>&lt;0.001</b>	0.8 (0.7)	<b>&lt;0.001</b>
10 word delayed recall, n 176	6.9 (1.6)	4.8 (2.2)	2.0 (2.0)	<b>&lt;0.001</b>	5.5 (2.2)	<b>&lt;0.001</b>
BAG, years	$-1.7$ (3.8)	1.4 (4.4)	3.1 (5.3)	<b>&lt;0.001</b>	0.3 (4.4)	<b>&lt;0.001</b>
Forebrain parenchyma/ICV (%)	63.6 (2.6)	62.0 (3.3)	58.6 (3.2)	<b>&lt;0.001</b>	62.6 (3.1)	<b>&lt;0.001</b>
Hippocampi/ICV (%)	0.49 (0.06)	0.46 (0.08)	0.38 (0.07)	<b>&lt;0.001</b>	0.47 (0.07)	<b>&lt;0.001</b>
WMH/ICV (%)	0.23 (0.37)	0.36 (0.42)	0.57 (0.62)	<b>&lt;0.001</b>	0.31 (0.41)	<b>&lt;0.001</b>

**Table 1.** Patient characteristics, and comparisons between disease stages. All continuous variables expressed as mean (SD). \*Three groups comparison using ANOVA/ $\chi^2$ . \*\*Dementia versus non-dementia (SCD-MCI) comparison using t-test/ $\chi^2$ . SCD subjective cognitive decline, MCI mild cognitive impairment, SD standard deviations, MMSE mini mental status examination, CDR-SB clinical dementia rating scale sum of boxes, BAG brain age gap, ICV intracranial volume, WMH white matter hypointensities. Significant values are in bold.

	(a) Dementia versus non-dementia			(b) Dementia versus SCD		
	AUC	95% CI	<i>p</i>	AUC	95% CI	<i>p</i>
BAG	0.678	0.608–0.748	< <b>0.001</b>	0.780	0.703–0.856	< <b>0.001</b>
Forebrain parenchyma/ICV	0.817	0.762–0.872	< <b>0.001</b>	0.893	0.835–0.951	< <b>0.001</b>
Hippocampi/ICV	0.833	0.780–0.886	< <b>0.001</b>	0.890	0.834–0.946	< <b>0.001</b>

**Table 2.** Ability of MRI methods to distinguish dementia from non-dementia (a) and from SCD (b). *SCD* subjective cognitive decline, *BAG* brain age gap, *ICV* intracranial volume, *AUC* area under the receiver operating curve, *CI* confidence interval. Significant values are in bold.

$R^2$  substantially (0.36–0.36 and 0.60–0.61) and did not affect the odds ratio or *p* value of BAG. The correlation between forebrain parenchyma and BAG was  $-0.27$  ( $p < 0.001$ ) and between hippocampus volume and BAG it was  $-0.22$  ( $p < 0.001$ ). In model 4, adding BAG to forebrain parenchyma and covariates, the Nagelkerke  $R^2$  increased from 0.40 to 0.44, and in model 5, adding BAG to hippocampus and covariates, it increased from 0.44 to 0.48 for dementia/non-dementia prediction (Table 3a). In the dementia/SCD prediction (Table 3b), the Nagelkerke  $R^2$  increased from 0.59 to 0.69 when adding BAG to forebrain parenchyma and covariates and from 0.58 to 0.68 when adding BAG to hippocampus and covariates.

Table 4 summarizes the comparisons of AD, FTD and DLB. BAG and forebrain parenchyma volume, but not hippocampus volume, were significantly different between groups. Post hoc group comparisons of AD versus non-AD, FTD versus non-FTD, and DLB versus non-DLB showed highest BAG in patients with FTD, and largest forebrain parenchyma volume in patients with DLB, compared to the other etiologies ( $p = 0.005$  and  $p = 0.012$ , respectively). AUC of BAG separating FTD from non-FTD was 0.82 (95% CI 0.62–1.00,  $p = 0.009$ ) with sensitivity 83% and specificity 82% and of forebrain parenchyma volume separating DLB from non-DLB AUC was 0.73 (95% CI 0.60–0.87,  $p = 0.009$ ) with sensitivity 83% and specificity 57%.

FTD patients were younger than the other patients (64.7 vs. 72.0,  $p = 0.023$ ). When adjusting for age and sex, the association between FTD and BAG was no longer statistically significant ( $t = 1.74$ ,  $p = 0.086$ ). However, a sensitivity subanalysis including only patients 70 years and below was performed. In this analysis, including five patients with FTD and 34 with non-FTD (median ages 64 and 66 ( $p = 0.117$ ) in FTD and non-FTD respectively), the FTD patients had significantly higher BAG (median 11.3 vs. 5.1 ( $p = 0.004$ )), and significantly higher BAG when adjusting for age and sex ( $p = 0.007$ ). The AUC of this subgroup analysis was 0.88 ( $p = 0.006$ ).

## Discussion

This study of the diagnostic properties of MRI-based brain age prediction in a memory clinic setting revealed that BAG was associated with disease stage, but the discriminatory power was outperformed by the hippocampus and forebrain parenchyma volumes. BAG was however found to discriminate FTD from other dementia etiologies.

A higher BAG was associated with more impaired disease stage, i.e. dementia versus non-dementia stages. This was as expected as previous studies have found BAG to be associated with cognitive test results and to be higher in AD compared to healthy controls<sup>5</sup>. The association between BAG and disease stage was not confounded by degree of vascular comorbidity, as measured with FreeSurfer WMH. Despite the association with disease stage, BAG performed poorly at discriminating dementia from non-dementia, while volumetrics using NQ did better. We suggest including MCI patients to the non-dementia group might be the cause of the weak discriminating power of BAG as the distinction between MCI and dementia is indefinite and excluding MCI patients should help distinguishing the remaining groups. Thus, in analyses discriminating dementia from SCD, the discriminating power was higher for both MRI measures. Despite increasing the AUC of both methods, BAG did not achieve the sensitivity and specificity levels that are generally expected from a clinical biomarker, while both NQ measures did<sup>14</sup>.

In clinical practice the separation of dementia from non-dementia stages is based on clinical interviews and examinations and does not include biomarkers in the decision-making process. Therefore, the analyses on associations with disease stage were primarily performed to compare the results of the novel, and until now, research-intended brain age prediction with the clinically available hippocampus and forebrain parenchyma volumes. Hippocampal atrophy is a well-known marker of AD<sup>15</sup> and is often used as a supportive biomarker in the diagnostic workup. It is therefore not surprising that this measure reached clinically relevant discriminatory power as most of the dementia patients had probable AD. The brain age prediction method is trained to capture general brain age and an increased BAG has been associated with genetic, lifestyle, and psychiatric diseases, in addition to AD<sup>7</sup>. Increased BAG is likely to be less specific to neurodegenerative diseases than hippocampal volume, supported by the current findings. It is also possible that cognitive and brain reserve play a greater role when BAG is applied to cognitively impaired patients. We adjusted for educational level, but this only accounts for one part of the complex concept of cognitive reserve<sup>16</sup>. Further, brain age prediction integrating information across the whole brain is likely less sensitive to specific, small region, hippocampal atrophy than hippocampal volume itself. Therefore, it is conceivable that future work performing regional brain age prediction (e.g. Kaufmann et al.<sup>7</sup>) may increase clinical sensitivity and specificity.

NQ volumetrics and BAG were also compared in logistic regression analyses. The adjusted model including hippocampus resulted in a higher Nagelkerke  $R^2$  than the model with BAG in the dementia/non-dementia prediction, in line with the ROC results. In the dementia/SCD prediction, the adjusted model with BAG had the

	Unadjusted		Adjusted model 1		Adjusted model 2		Adjusted model 3		Adjusted model 4		Adjusted model 5	
	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
<b>(a)</b>												
Age	1.11 (1.07–1.16)	<0.001	1.03 (0.98–1.08)	0.237	1.01 (0.96–1.06)	0.707	1.13 (1.08–1.18)	<0.001	1.06 (1.00–1.11)	0.043	1.03 (0.98–1.09)	0.205
Female sex	1.03 (0.60–1.77)	0.909	1.04 (0.54–2.00)	0.917	1.59 (0.79–3.19)	0.193	1.12 (0.59–2.12)	0.736	0.93 (0.47–1.84)	0.840	1.38 (0.67–2.83)	0.378
Education	0.92 (0.85–0.99)	<b>0.021</b>	0.94 (0.86–1.03)	0.156	0.96 (0.88–1.06)	0.416	0.94 (0.86–1.02)	0.139	0.93 (0.85–1.02)	0.141	0.95 (0.87–1.05)	0.319
Forebrain pvc/ICV	0.68 (0.61–0.76)	<0.001	0.71 (0.63–0.81)	<0.001					0.76 (0.66–0.87)	<0.001		
Hippocampi/ICV	0.5 × 10 <sup>-8</sup> (0.3 × 10 <sup>-10</sup> –0.1 × 10 <sup>-5</sup> )	<0.001			0.8 × 10 <sup>-8</sup> (0.1 × 10 <sup>-10</sup> –0.5 × 10 <sup>-5</sup> )	<0.001					0.1 × 10 <sup>-8</sup> (1.7 × 10 <sup>-10</sup> –0.9 × 10 <sup>-4</sup> )	<0.001
BAG	1.12 (1.06–1.19)	<0.001					1.17 (1.09–1.26)	<0.001	1.13 (1.04–1.22)	<b>0.003</b>	1.12 (1.04–1.21)	<b>0.005</b>
R <sup>2</sup> *				0.40		0.44		0.36		0.44		0.48
<b>(b)</b>												
Age	1.15 (1.09–1.22)	<0.001	1.06 (0.99–1.13)	0.096	1.02 (0.94–1.09)	0.696	1.20 (1.12–1.29)	<0.001	1.11 (1.02–1.21)	0.013	1.08 (0.99–1.18)	0.068
Female sex	1.84 (0.87–3.88)	0.110	0.30 (0.10–0.92)	0.034	1.49 (0.51–4.35)	0.470	3.16 (1.04–9.59)	<b>0.042</b>	0.22 (0.06–0.81)	0.023	2.16 (0.61–7.65)	0.232
Education	0.89 (0.91–0.98)	<b>0.021</b>	0.96 (0.83–1.10)	0.521	0.98 (0.86–1.12)	0.773	0.94 (0.82–1.08)	0.382	0.95 (0.81–1.12)	0.575	0.95 (0.81–1.11)	0.520
Forebrain pvc/ICV	0.55 (0.45–0.67)	<0.001	0.59 (0.47–0.73)	<0.001					0.66 (0.52–0.84)	<0.001		
Hippocampi/ICV	0.2 × 10 <sup>-11</sup> (0.2 × 10 <sup>-15</sup> –0.1 × 10 <sup>-7</sup> )	<0.001			0.1 × 10 <sup>-10</sup> (0.2 × 10 <sup>-15</sup> –0.4 × 10 <sup>-6</sup> )	<0.001					0.9 × 10 <sup>-8</sup> (0.2 × 10 <sup>-12</sup> –0.5 × 10 <sup>-3</sup> )	<b>0.001</b>
BAG	1.22 (1.12–1.34)	<0.001					1.37 (1.21–1.55)	<0.001	1.33 (1.15–1.54)	<0.001	1.30 (1.13–1.49)	<0.002
R <sup>2</sup> *				0.59		0.58		0.60		0.69		0.68

**Table 3.** Adjusted associations to dementia/non-dementia (a), and to dementia/SCD (b). Logistic regression with dementia (1) versus non-dementia (0) (a), and versus SCD (0) (b), as the dependent variables. \*Nagelkerke R<sup>2</sup>. SCD subjective cognitive decline, OR odds ratio, CI confidence interval, Forebrain pvc forebrain parenchyma, ICV intracranial volume, BAG brain age gap. Significant values are in bold.

	AD, n 76	FTD, n 6	DLB, n 12	$p^*$	AD versus non-AD	FTD versus non-FTD	DLB versus non-DLB
					$p^{**}$	$p^{**}$	$p^{**}$
Age, years	72.3 (7.5)	64.7 (6.1)	70.2 (8.5)	0.050	<b>0.048</b>	<b>0.023</b>	0.505
Female, n (%)	41 (53.9%)	4 (66.7%)	2 (16.7%)	<b>0.039</b>	0.106	0.399	<b>0.013</b>
Education, years (n 84)	12.9 (3.5)	16.4 (3.1)	14.4 (3.4)	0.055	<b>0.030</b>	<b>0.042</b>	0.285
BAG, years	2.7 (5.2)	8.7 (4.9)	2.6 (3.0)	<b>0.019</b>	0.159	<b>0.005</b>	0.574
Forebrain parenchyma/ ICV (%)	58.2 (3.2)	57.4 (2.2)	60.6 (2.2)	<b>0.034</b>	0.117	0.380	<b>0.012</b>
Hippocampi/ICV (%)	0.37 (0.07)	0.37 (0.04)	0.40 (0.06)	0.450	0.323	0.893	0.205
WMH/ICV (%)	0.61 (0.58)	0.30 (0.35)	0.53 (0.54)	0.419	0.295	0.215	0.574

**Table 4.** Characteristics of patients with various dementia etiologies, and comparisons of diagnostic groups. \*Three group comparison using ANOVA/ $\chi^2$ . \*\*t test. AD Alzheimer's disease, FTD frontotemporal dementia, DLB dementia with Lewy bodies, BAG brain age gap, ICV intracranial volume, WMH white matter hypointensities. Significant values are in bold.

highest Nagelkerke  $R^2$ . Finally, models including BAG and one of the NQ measures had the highest Nagelkerke  $R^2$ , indicating that NQ volumetrics and BAG provides complementary information for dementia prediction.

The clinical utility of a biomarker ultimately depends on its value for etiological diagnostic work-up. BAG was larger in patients with FTD and the discriminatory power for separating FTD from other etiologies was excellent, with sensitivity and specificity levels above 80%. Previous studies have reported increased brain age in severe mental disorders including schizophrenia, and a genome wide association study found an association between brain ageing and the MAPT gene which encodes for tau protein that is related to FTD<sup>17</sup>. The present sample is relatively small, and the explorative design does not allow for decisive conclusions. Based on previous findings of BAG being associated with diseases associated with frontal lobe pathologies<sup>17,18</sup> our findings encourage further studies on the association between BAG and FTD and other frontal lobe pathologies. However, the size and distribution of the affected brain regions are expected to influence brain age prediction and could introduce bias to the associations. Indeed, the frontal lobes occupy a relatively large proportion of the brain, accounting for two thirds of the total brain volume<sup>19</sup>. Atrophy of this region could possibly therefore affect brain age estimates to a larger extent than focal atrophy of a smaller brain region, e.g. the hippocampus. Another possibility is that age prediction was biased by age, i.e. that the accuracy of the prediction model varies with age, a phenomenon commonly seen in brain age models<sup>20</sup>. The younger age of the FTD patients may thus have influenced the results. Indeed, when adjusting for age the group differences were attenuated to the point where they no longer reached the threshold for statistical significance. Practically, it is difficult to correct for bias between groups with different age distributions since the true structure of the bias is unknown; correcting jointly in both groups based on independent data can have little to no effect, whereas an in-sample correction could reduce actual group differences. Thus, we performed a sensitivity analysis matching the groups on age, confirming higher BAG in FTD patients.

Hippocampus volume was not significantly smaller in patients with AD dementia compared to non-AD dementia, which might seem unexpected as hippocampal atrophy is known to be a marker of AD. This is however in line with a previous study based on a larger, yet partly overlapping, cohort where hippocampus volume reached an AUC of only 0.62 for discrimination of AD dementia versus non-AD dementia<sup>12</sup>. Further, previous studies from our group concluded that as much as 53% of patients with AD dementia lack atrophy of the hippocampi, and that atypical atrophy patterns are common<sup>21,22</sup>. Both these findings might explain why hippocampus was not able to separate patients with AD dementia from non-AD dementia at a clinically acceptable level, in that study.

There are limitations to the current study. The cross-sectional explorative design and the relatively low number of patients in the etiological comparisons and in the FTD sensitivity analysis limit the confidence and generalizability of the conclusions. Another limitation is that only clinical criteria without AD specific molecular imaging ( $A\beta$ -PET) or biofluid (CSF, plasma,  $A\beta$ /p-tau) biomarkers were used as the gold standard for the etiological diagnoses. Although the clinical diagnoses were made using the NIA/AA criteria by two experienced physicians, and while the main goal of this study was to examine whether BAG could serve as an additional diagnostic marker in a naturalistic clinical setting, future studies on the diagnostic properties of BAG should include specific biomarkers to substantiate the results. Further, information on comorbidity was not available in the current data set. Although our analyses revealed no substantial confounding effects of white matter cerebrovascular pathology as indexed using WMH from FreeSurfer, various comorbid clinical conditions may influence MRI based analyses and subsequent brain age prediction, and should be considered in future studies.

## Conclusions

Brain age estimation using clinically available MRI scans adds an interesting perspective to the association between brain ageing and neurodegenerative diseases. While NQ volumetrics outperformed BAG in terms of discriminatory power for patients with dementia versus those without dementia, the two measures provided complementary information and we did not find evidence to suggest that our findings were confounded by cerebrovascular comorbidity. The finding of increased brain age in FTD patients is of clinical interest as few biomarkers are available for this diagnosis. The causal direction of effects and prognostic properties remain to be further characterized, preferably in a longitudinal study.



While automated tools for individual-level brain phenotyping based on machine learning, such as brain age prediction, have potential to support clinical diagnostics in a memory clinic setting, further developments and validations of its etiology discriminating and prognostic properties are needed to characterize its clinical potential.

## Methods

**Participants.** All patients assessed for cognitive complaints at the memory clinic at Oslo University hospital (OUH), Norway, between June 2015 and January 2019 that met the criteria of SCD, MCI, or dementia (see below), and that had been examined with brain MRI at the same scanner at OUH +/- 6 months from the clinical assessment were eligible for inclusion. Referral to the research MRI scanner at OUH, and not to another MRI scanner, was done at random when an MRI scan was indicated as part of the clinical routine, and it was convenient for the patient due to geography to perform it at OUH. Among the 254 patients fulfilling these criteria, MRI scans of sufficient quality were available from 232 patients, which form the present study cohort.

All patients had consented to be part of a national quality and research register (The Norwegian registry of persons assessed for cognitive symptoms, NorCog). The inclusion and clinical assessments carried out at the memory clinic and data included in NorCog have been described previously<sup>23</sup>.

**Diagnoses and clinical assessments.** All patients were diagnosed retrospectively by two experienced physicians (KP and THE), using all available information from the extensive clinical assessments including information from patients and proxies on symptoms, cognitive test results, function in activities of daily living, and physical and psychiatric examinations<sup>23,24</sup>. The NIA/AA 2011 criteria were used to diagnose MCI and dementia<sup>25</sup>, and the Jessen criteria were used to diagnose SCD<sup>26</sup>. Among the patients with dementia, those fulfilling clinical criteria of AD according to NIA/AA 2011 criteria (probable AD and possible AD mixed with vascular pathology)<sup>25</sup>, frontotemporal dementia (FTD) according to the Rascovsky and Gorno-Tempini 2011 criteria<sup>27,28</sup>, and dementia with Lewy bodies (DLB) according to the 2017 McKeith criteria<sup>29</sup>, were included in etiology-based validity analyses. Other diagnoses were excluded due to few cases or other mixed etiologies (i.e., two patients with vascular cognitive impairment, one patient with Parkinson dementia, ten patients with various mixed etiologies, and three unspecific dementia diagnoses). Clinical radiology reports including information on structural pathologies of both cortical and subcortical regions were used to exclude etiologies not related to dementia (i.e. intracranial bleedings or tumors). Further, signs of vascular pathology indicative of cerebrovascular disease and frontal atrophy were used according to criteria of vascular cognitive impairment and FTD. Information on regional structural changes based the clinical report, or from the NQ report was not included in the diagnostic criteria.

The Norwegian version of the Mini Mental State Examination (MMSE) and the Clinical Dementia Rating scale-sum of boxes (CDR-SB) were used as measures of global cognitive and functional performance for descriptive purposes. MMSE gives a score between zero and 30, the higher score the better global cognitive function<sup>30,31</sup> and the CDR-SB is a global measure of cognitive and functional impairment including six items scored from zero to 3 and summed up to a score ranging from zero to 18, the higher the score the greater the impairment<sup>32</sup>. Two CDR-certified physicians scored the CDR-SB post hoc, based on all available information from the patient records (KP and THE). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) 10-word delayed recall test, with scores from zero to 10, the higher score the better the learning and retrieving capacity<sup>33</sup> was included as a descriptive measure of memory function.

**MRI acquisition and analysis.** All patients were assessed with brain MRI according to the same research protocol using a GE Discovery MR750 3T scanner (GE Healthcare, Milwaukee, WI, US). Whole brain T1-weighted structural MRI data was acquired using an inversion recovery-fast spoiled gradient echo sequence (BRAVO) with the following parameters: TR = 8.16 ms, TE = 3.18 ms, TI = 450 ms, flip angle = 12°, field of view = 256 mm, acquisition matrices = 256 × 256, 188 sagittal slices, slice thickness = 1.0 mm, voxel size = 1 × 1 × 1 mm<sup>3</sup>. The scans were analyzed following a previously established minimal processing pipeline and brain age prediction model<sup>11</sup>, and with NeuroQuant 3. version (NQ, CorTechs labs/University of California, San Diego, CA, USA)<sup>13</sup>.

Brain age was computed using a state-of-the-art deep convolutional neural network, trained on a large population dataset (N = 53,542 from 21 publicly accessible datasets) with a wide age range from a multitude of scanners<sup>11</sup>, not including the one used for the present study. The model is available online<sup>34</sup>. As input, the model used minimally processed imaging data linearly registered with six degrees of freedom to MNI152 space. BAG was calculated by subtracting chronological age from predicted brain age, such that a positive BAG reflects higher predicted age compared to chronological age and vice versa.

NQ produces valid and reliable volumetric measures of several brain regions<sup>13,35</sup>. NQ volumetry of hippocampus correlates well with visual ratings of the medial temporal lobe using the Scheltens scale<sup>36</sup>. In the present study, we included the volume of the hippocampus as atrophy of the hippocampus is one of the best-established diagnostic imaging biomarkers for AD<sup>15</sup>, constituting the majority of the patients with dementia in the present sample. Additionally, we included forebrain parenchyma volume, including all parenchymal brain volumes except the brainstem and cerebellum, as this volume was previously shown to have the best ability to discriminate between dementia and non-dementia<sup>12</sup>, and to include a measure that would represent more than the AD specific medial temporal region. Both structures were included as proportions of estimated intracranial volume, i.e. the sum of whole brain volume and CSF spaces, to adjust for head size.

In 229 of the 232 included patients, FreeSurfer data on white matter hypointensities (WMH) was available<sup>37</sup>. Although based on T1-weighted MRI scans, this measure has been found to correlate well with both state-of-the-art T2/FLAIR white matter hyperintensities and the visual rating scale of Fazekas<sup>38–40</sup>. To adjust for head

size, WMH was divided by the FreeSurfer measure of total intracranial volume. WMH was included post hoc to evaluate if cerebrovascular comorbidity could confound the association between BAG and dementia.

**Statistics.** Data were analyzed using IBM SPSS Statistics for Windows (version 27, Armonk, NY, USA). The significance level was set at 0.05. Diagnostic groups were compared using independent samples t-test and ANOVA for continuous measures and  $\chi^2$  tests for categorical measures. Age- and sex-adjusted linear models were performed for group-wise comparisons of the MRI measures. Medians and Mann–Whitney U test were used in a sensitivity analysis of the subgroup of patients 70 years of age and below.

To compare the validity of the two MRI methods, receiver operating characteristics (ROC) analyses were carried out for each method, calculating the area under the curve (AUC) as a measure of the performance of the classifiers to separate dementia from non-dementia and from SCD. The interpretation of the AUC depends on the clinical setting in which the test should be used, but generally an AUC of 0.5–0.7 is regarded poor, 0.7–0.8 acceptable, 0.8–0.9 excellent, > 0.9 outstanding<sup>41</sup>. For a biomarker to be clinically useful, the sensitivity and specificity should be at least 80%<sup>14</sup>. Thus, for each MRI measure, the sensitivity was set at 80% and the specificity was obtained from the ROC analysis.

Bivariate Pearson correlation of the three MRI measures (BAG, hippocampus volume and forebrain parenchyma volume) was performed to prepare the logistic regression analysis. Hippocampus volume and forebrain parenchyma volume were highly correlated ( $r=0.703$ ,  $p<0.001$ ). Thus, in the logistic regression analyses predicting dementia versus non-dementia and dementia versus SCD, the forebrain parenchyma, hippocampus volume, and BAG were included in separate models (models 1, 2, and 3), adjusting for demographic covariates (age, sex, and educational level). The Nagelkerke  $R^2$  was used as an estimate of the explained variance to compare the models. In models 4 and 5, BAG was added to each of the two volumetric measures to assess its additive value for the prediction of disease stage. Finally, to explore if cerebrovascular comorbidity could confound the association between BAG and diagnosis, WMH was added to model 3 (not in table).

**Ethics declarations.** All patients gave written informed consent to be included in NorCog. The Regional Committee of Medical Research Ethics of the South-East Norway approved the use of NorCog data in the present study (REC South-East number 29461). All methods and analyses were performed in accordance with the Declaration of Helsinki.

### Data availability

Data from this study are not made publicly available. The regional ethics committee has not approved data delivery outside of Europe, consent for publication of raw data was not obtained from the participants of the study, and complete anonymization is difficult to achieve. Data are available from the National Advisory Unit on Ageing and Health for researchers who meet the criteria for access to confidential data. Data requests can be addressed to post@aldringoghelse.no.

Received: 3 May 2023; Accepted: 8 September 2023

Published online: 11 September 2023

### References

- World Health Organization. *Dementia fact sheet*. <https://www.who.int/news-room/fact-sheets/detail/dementia> (2022).
- Karran, E. & De Strooper, B. The amyloid hypothesis in Alzheimer disease: New insights from new therapeutics. *Nat Rev Drug Discov* **21**, 306–318. <https://doi.org/10.1038/s41573-022-00391-w> (2022).
- Graham, S. A. *et al.* Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry Res.* **284**, 112732. <https://doi.org/10.1016/j.psychres.2019.112732> (2020).
- Helsedirektoratet. Tilrettelegging for bruk av kunstig intelligens i helsetjenesten. (2021).
- Wrigglesworth, J. *et al.* Factors associated with brain ageing—A systematic review. *BMC Neurol.* **21**, 312. <https://doi.org/10.1186/s12883-021-02331-4> (2021).
- Rokicki, J. *et al.* Multimodal imaging improves brain age prediction and reveals distinct abnormalities in patients with psychiatric and neurological disorders. *Hum. Brain Mapp.* **42**, 1714–1726. <https://doi.org/10.1002/hbm.25323> (2021).
- Kaufmann, T. *et al.* Common brain disorders are associated with heritable patterns of apparent aging of the brain. *Nat. Neurosci.* **22**, 1617–1623. <https://doi.org/10.1038/s41593-019-0471-7> (2019).
- Doan, N. T. *et al.* Distinguishing early and late brain aging from the Alzheimer's disease spectrum: Consistent morphological patterns across independent samples. *Neuroimage* **158**, 282–295. <https://doi.org/10.1016/j.neuroimage.2017.06.070> (2017).
- Wang, J. *et al.* Gray matter age prediction as a biomarker for risk of dementia. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21213–21218. <https://doi.org/10.1073/pnas.1902376116> (2019).
- Taylor, A. *et al.* Investigating the temporal pattern of neuroimaging-based brain age estimation as a biomarker for Alzheimer's Disease related neurodegeneration. *Neuroimage* **263**, 119621. <https://doi.org/10.1016/j.neuroimage.2022.119621> (2022).
- Leonardsen, E. H. *et al.* Deep neural networks learn general and clinically relevant representations of the ageing brain. *Neuroimage* **256**, 119210. <https://doi.org/10.1016/j.neuroimage.2022.119210> (2022).
- Persson, K. *et al.* Fully automated structural MRI of the brain in clinical dementia workup. *Acta Radiol.* **58**, 740–747. <https://doi.org/10.1177/0284185116669874> (2017).
- Brewer, J. B., Magda, S., Airriess, C. & Smith, M. E. Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease. *AJNR Am. J. Neuroradiol.* **30**, 578–580. <https://doi.org/10.3174/ajnr.A1402> (2009).
- The Ronald and Nancy Reagan research institute of the Alzheimer's association and the National institute on aging working group. Consensus report of the Working Group on: "Molecular and Biochemical Markers of Alzheimer's Disease". The Ronald and Nancy Reagan Research Institute of the Alzheimer's Association and the National Institute on Aging Working Group. *Neurobiol. Aging* **19**, 109–116 (1998).
- Teipel, S. J. *et al.* Relevance of magnetic resonance imaging for early detection and diagnosis of Alzheimer disease. *Med. Clin. N. Am.* **97**, 399–424. <https://doi.org/10.1016/j.mcna.2012.12.013> (2013).
- Stern, Y. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol.* **11**, 1006–1012. [https://doi.org/10.1016/s1474-4422\(12\)70191-6](https://doi.org/10.1016/s1474-4422(12)70191-6) (2012).

17. Jonsson, B. A. *et al.* Brain age prediction using deep learning uncovers associated sequence variants. *Nat. Commun.* **10**, 5409. <https://doi.org/10.1038/s41467-019-13163-9> (2019).
18. Howes, O. D., Cummings, C., Chapman, G. E. & Shatalina, E. Neuroimaging in schizophrenia: An overview of findings and their implications for synaptic changes. *Neuropsychopharmacology* **48**, 151–167. <https://doi.org/10.1038/s41386-022-01426-x> (2023).
19. Chayer, C. & Freedman, M. Frontal lobe functions. *Curr. Neurol. Neurosci. Rep.* **1**, 547–552. <https://doi.org/10.1007/s11910-001-0060-4> (2001).
20. Liang, H., Zhang, F. & Niu, X. Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders. *Hum. Brain Mapp.* **40**, 3143–3152. <https://doi.org/10.1002/hbm.24588> (2019).
21. Persson, K. *et al.* MRI-assessed atrophy subtypes in Alzheimer's disease and the cognitive reserve hypothesis. *PLoS ONE* **12**, e0186595. <https://doi.org/10.1371/journal.pone.0186595> (2017).
22. Persson, K. *et al.* Hippocampal atrophy subtypes of Alzheimer's disease using automatic MRI in a memory clinic cohort: clinical implications. *Dement. Geriatr. Cogn. Disord.* **51**, 80–89. <https://doi.org/10.1159/000522382> (2022).
23. Medbøen, I. T. *et al.* Cohort profile: the Norwegian Registry of Persons Assessed for Cognitive Symptoms (NorCog)—A national research and quality registry with a biomaterial collection. *BMJ Open* **12**, e058810. <https://doi.org/10.1136/bmjopen-2021-058810> (2022).
24. Braekhus, A., Ulstein, I., Wyller, T. B. & Engedal, K. The Memory Clinic—outpatient assessment when dementia is suspected. *Tidsskr. Nor. Laegeforen.* **131**, 2254–2257. <https://doi.org/10.4045/tidsskr.11.0786> (2011).
25. Albert, M. S. *et al.* The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* **7**, 270–279. <https://doi.org/10.1016/j.jalz.2011.03.008> (2011).
26. Jessen, F. *et al.* A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement.* **10**, 844–852. <https://doi.org/10.1016/j.jalz.2014.01.001> (2014).
27. Gorno-Tempini, M. L. *et al.* Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006–1014. <https://doi.org/10.1212/WNL.0b013e31821103e6> (2011).
28. Rascofsky, K. *et al.* Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **134**, 2456–2477. <https://doi.org/10.1093/brain/awr179> (2011).
29. McKeith, I. G. *et al.* Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology* **89**, 88–100. <https://doi.org/10.1212/wnl.0000000000004058> (2017).
30. Engedal, K., Haugen, P., Gilje, K. & Laake, P. Efficacy of short mental tests in the detection of mental impairment in old age. *Compr. Gerontol. A* **2**, 87–93 (1988).
31. Folstein, M. F., Folstein, S. E. & McHugh, P. R. “Mini-mental state.” A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
32. Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A. & Martin, R. L. A new clinical scale for the staging of dementia. *Br. J. Psychiatry* **140**, 566–572 (1982).
33. Morris, J. C. *et al.* The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159–1165 (1989).
34. Leonardsen, E. H. <https://github.com/estenh1/pyment-public> (2021).
35. Albright, J., Leyden, K. & Airriess, C. *CorTechs Labs white paper "The Importance of Quantitative Volumetric Analysis for Brain MRI 10 Years of Clinical Practice"*. <http://www.cortechslabs.com/whitepapers> (2015).
36. Persson, K. *et al.* Comparison of automated volumetry of the hippocampus using NeuroQuant(R) and visual assessment of the medial temporal lobe in Alzheimer's disease. *Acta Radiol.* **59**, 997–1001. <https://doi.org/10.1177/0284185117743778> (2018).
37. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781. <https://doi.org/10.1016/j.neuroimage.2012.01.021> (2012).
38. Wei, K. *et al.* White matter hypointensities and hyperintensities have equivalent correlations with age and CSF  $\beta$ -amyloid in the nondemented elderly. *Brain Behav* **9**, e01457. <https://doi.org/10.1002/brb3.1457> (2019).
39. Cedres, N. *et al.* Predicting Fazekas scores from automatic segmentations of white matter signal abnormalities. *Aging Albany* **12**, 894–901. <https://doi.org/10.18632/aging.102662> (2020).
40. Fazekas, F., Chawluk, J. B., Alavi, A., Hurtig, H. I. & Zimmerman, R. A. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am. J. Roentgenol.* **149**, 351–356. <https://doi.org/10.2214/ajr.149.2.351> (1987).
41. Hosmer, D. W. Jr., Lemeshow, S. & Sturdivant, R. X. *Applied logistic regression* Vol. 398 (Wiley, 2013).

## Acknowledgements

We want to acknowledge the Norwegian registry of persons assessed for cognitive symptoms (NorCog), for providing access to patient data.

## Author contributions

K.P., E.H.L., T.W. designed the study; K.P. did the analyses and drafted the work; K.P., K.E., T.H.E., G.G.T., A.B.K., G.S., E.H.L., T.W., L.W. made substantial contributions to the conception, and interpretation; E.H.L., T.W., L.W. created the B.A.G. method. All authors read, revised, and approved the final manuscript.

## Funding

The Southern and Eastern Norway Regional Health Authority supported the study through funding for Karin Persson but was not involved in conducting the study or in preparation of the manuscript. The study was also supported by the the UiO:Life Science Convergence environment (4MENT), University of Oslo, the Research Council of Norway (RCN, 298646), and the European Research Council (ERC) under the European Union's Horizon 2020 research and Innovation program (802998).

## Competing interests

Dr. Persson, Dr. Edwin, and Dr. Knapskog report work with Roche BN29553 trial; Dr. Persson and Dr. Knapskog report work with Novo Nordisk NN6535-4730 trial; Dr. Edwin, and Dr. Knapskog report work with Boehringer-Ingelheim 1346.0023 trial; outside the submitted work. Esten H. Leonardsen is chairman of the board and major shareholder of Biometrical AS. Dr. Selbæk reports participation in a Biogen advisory board. The other authors declare that they have no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to K.P.



**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023