Brain-age gap in neuromyelitis optica spectrum disorders and multiple sclerosis

Ren Wei,[1] Xiaolu Xu,[1] Yunyun Duan,[1] Ningnannan Zhang,[2] Jie Sun,[2] Haiqing Li,[3] Yuxin Li,[3] Yongmei Li,[4] Chun Zeng,[4] Xuemei Han,[5] Fuqing Zhou,[6] Muhua Huang,[6] Runzhi Li,[7]Zhizheng Zhuo,[1] Frederik Barkhof,[8,9] James H. Cole,[9,10] Yaou Liu,[1] Alzheimer's Disease Neuroimaging Initiative*.

Affiliations

1. Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China.

2. Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin Medical University General Hospital, Tianjin, China.

3. Radiology department, Huashan Hospital, Fudan University, Shanghai, China.

4. Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.

5. No.1 Department of Neurology, China-Japan Union Hospital of Jilin University, Changchun, Jilin, China

6. Department of Radiology, the First Affiliated Hospital, Nanchang University, Nanchang, Jiangxi Province, China.

7. Department of Neurology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China.

8. Department of Radiology and Nuclear Medicine, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam 1007 MB, The Netherlands

9. Centre for Medical Image Computing, University College London, UK

10. Dementia Research Centre, Queen Square Institute of Neurology, University College London, UK.

Correspondence to Dr Yaou Liu, Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China;liuyaou@bjtth.org

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI

investigators can be found at:http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Abstract: 195 words

Word count: 2991 words

Reference count: 27

Tables: 3

Figures: 2

Commented [MOU1]: 更改后最后 update

Key messages

What is already known on this topic:

The deep-learning-derived brain age gap (BAG) is associated with various clinical risk factors and can be used for risk stratification of various neurological and psychiatric diseases, including multiple sclerosis (MS). The clinical significance of BAG prediction in neuromyelitis optica spectrum disorder (NMOSD) relative to RRMS is not known.

What this study adds:

A deep learning model was able to estimate the brain age gap (BAG) from 3D structural MRI scans and is robust across multiple centers and multiple scanners. A significant BAG was found in NMOSD patients compared with healthy controls, although it was less marked than in RRMS patients. Higher disability and advanced atrophy were associated with a larger BAG in both NMOSD and RRMS. BAG was a predictive biomarker of EDSS worsening in NMOSD and RRMS.

How this study might affect research, practice or policy:

The BAG is a comprehensive and relevant disease marker in NMOSD and RRMS.

ABSTRACT

Objective

To evaluate the clinical significance of deep-learning-derived brain age prediction in neuromyelitis optica spectrum disorder (NMOSD) relative to relapsing-remitting MS(RRMS).

Methods

This cohort study used data retrospectively collected from 6 tertiary neurological centers in China between 2009 and 2018. In total, 199 NMOSD and 200 RRMS patients were studied alongside 269 healthy controls. Clinical follow-up was available in 85 NMOSD and 124 RRMS patients (mean duration NMOSD = 5.8±1.9 [1.9-9.9] years, RRMS = 5.2±1.7 [1.5-9.2] years). Deep learning was used to learn 'brain age' from MRI scans in the healthy controls and estimate the brain age gap (BAG) in patients.

Results

A significantly higher BAG was found in the NMOSD (5.4±8.2 years) and RRMS (13.0±14.7 years) groups compared with healthy controls. A higher baseline disability score and advanced brain-volume loss were associated with increased BAG in both patient groups. A longer disease duration was associated with increased BAG in RRMS. BAG significantly predicted EDSS worsening in NMOSD and RRMS patients.

Conclusions

There is a clear brain age gap in NMOSD, although smaller than in RRMS. The brain age gap is a clinically relevant MRI marker in NMOSD and RRMS.

INTRODUCTION

Age is an independent marker for disease progression in neuromyelitis optica spectrum disorder (NMOSD)[1] and multiple sclerosis (MS)[2], two major inflammatory demyelinating diseases of the central nervous system[1,3]. However, aging does not affect everyone in the same way, so researchers have sought biological markers of aging processes that may explain some of these individual differences and are more reflective of age-related disease processes. The so-called "brain age" paradigm has been designed to determine the brain's biological age[4], which can be estimated from anatomical brain MRI scans. By analyzing the similarity of a given brain scan with scans from a range of healthy individuals, machine-learning techniques can predict a person's brain age from neuroimaging features, providing a novel way of indexing deviations from normal brain aging. Compared to calendar age, brain age may provide more comprehensive information for understanding disease impact in NMOSD and RRMS.

The brain age gap (BAG) is the difference between calendar age and predicted brain age. BAG thus represents the deviation from an expected healthy aging trajectory. This MRI biomarker integrates structural alterations across the brain associated with the aging process.[5,6] Previous studies have suggested that BAG is associated with various clinical risk factors and can be used for risk stratification of various neurological and psychiatric diseases including MS[7]. However, no one has investigated BAG in NMOSD patients and its ability to understand and predict EDSS worsening.

In this study, we used a novel deep-learning brain age model to investigate the utility of BAG as a neuroimaging biomarker to predict EDSS worsening in NMOSD compared to RRMS in a large multicenter dataset.

METHODS

Participants

Data from NMOSD and RRMS patients were retrospectively collected from six tertiary neurological centers in China covering the period between Nov 2009 and Apr 2018. Patients who fulfilled the following criteria were included: (a) confirmed diagnosis of NMOSD according to 2015 revised diagnostic criteria [8] or relapsing-remitting MS(RRMS) according to 2015 McDonald criteria; (b) complete demographic and clinical information, including baseline Expanded Disability Status Scale (EDSS) score and disease duration; and (c) good quality baseline 3D T1-weighted structural images. Clinical evaluation, diagnosis, treatment, and follow-up assessments of the participants were conducted at each center by local neurologists with expertise in demyelinating diseases. EDSS worsening was defined as an increase in EDSS score ≥1.0 for baseline EDSS ≤5.5 or an increase in EDSS score ≥0.5 for baseline EDSS >5.5, as previously published [9].

Data for deep-learning model training

Training data for our deep-learning-derived brain age included MRI scans from healthy controls (HCs, n=9794) from publicly available datasets, including ADNI, AIBL[10], GSP[11] and SLIM[12], as well as a group of healthy people scanned at Beijing Tiantan Hospital from January-December 2019 (**Supplementary Table 1, Supplementary Figure 1**). After training, the model was tested on two further independent datasets. Internal validation data comprised another group of healthy participants (n=462) scanned at Beijing Tiantan Hospital from January-April 2020 on two different scanners (see **Supplementary Table 1**). The external validation dataset included healthy controls from the multicenter NMOSD and MS cohorts (n=267).

Image acquisition and data preprocessing

All the MRI images of participants as well as the validation dataset were acquired on 3.0T machines at near 1.0 mm isotropic resolution by MP-RAGE or similar sequences. Noncontrast 3D T1WI scans were affinely registered to MNI space. Skull stripping was performed by HD-BET on the registered scans[13]. The signal intensity of the resulting images was normalized by dividing by the mean intensity within the cerebral mask. Scans were then resampled to 1 mm isotropic resolution using linear interpolation and served as the input of the proposed convolutional neural network (CNN).

Age at each scan was determined by either of two methods: 1) the demographic metadata (in years) provided by owners of the dataset; 2) calculated from the difference between date of birth and image acquisition date recorded in DICOM metadata, which was done in days and converted to years. Inconsistent data were omitted from the study.

Brain volume measurement

Brain volume segmentation was performed using the automated recon-all procedure in FreeSurfer package (version 6.0.0) as described by Fischl et al.[14]. The total brain volume was calculated and

normalized by dividing by the estimated total intracranial volume[15].

Model construction, training and prediction

We built a 3D CNN called the 3D SFCN network as per the work of Peng and colleagues[16]. We modified the output structure so that the network could predict age across a larger range of 6-90 years. Model training and mathematical details are described in the supplementary material.

BAG was calculated by subtracting chronological age from predicted brain age, with a positive BAG indicating an older-looking brain. To investigate the possible influence of brain lesions on age prediction, we performed a correlation analysis between raw and lesion-filled 3D T1WI images. Lesion filling was performed by default pipeline of Lesion Segmentation Tool (LST, version 3.0.0, https://www.applied-statistics.de/lst.html).

Statistical analysis

Statistical analyses were conducted using R version 3.6.3. Graphs were plotted with ggplot2 package. Intergroup comparison was conducted using the Chi-square test (for categorical variables), Wilcoxon signed-rank test (for EDSS) and student's t-test or ANOVA with Tukey's range test as post hoc analysis (for continuous variable). Survival analysis with Kaplan-Meier curve and Cox-proportional hazards model were used to analyze time-to-progression data. Other details are described in the supplementary material. All statistical tests were 2-sided, and $p<0.05$ was considered statistically significant.

RESULTS

Participants

In total, 199 NMOSD patients, 200 RRMS patients, and 269 age- and sex-matched HC subjects were included (**Table 1**). Patients with NMOSD were older at baseline (41.0±13.0 yr versus 37.1±11.4 yr, p=0.005), had a longer disease duration (4.5±5.1 yr versus 3.2±4.4 yr, p=0.006), and had less severe disability measured by EDSS at baseline (2.0 versus 3.5, p<0.001) than patients with RRMS. Of the NMOSD patients included, 52(26.1%) patients received disease-modifying therapy (DMT), others received immunosuppressants including cyclophosphamide, azathioprine, etc. In the RRMS group, 86(43.0%) patients received DMT, others received the above other treatment.

Follow-up data were available for 85 NMOSD and 124 RRMS patients (median follow-up duration: 5.8±1.9 years and 5.2±1.7 years, respectively). During follow-up, 31 NMOSD patients and 42 RRMS patients experienced EDSS worsening.

Table 1. Demographic characteristics, baseline status and deep-learning-derived brain age of participants.

	NMOSD	**RRMS**	**Healthy Controls**	**p value**
Baseline				
N	199	200	269	
Age at baseline, yr [min-max]	41.0±13.0[16.9-66.0]	37.1±11.4[16.6-66.9]	38.5±12.7[17.0-69.0]	NMOSD vs HC 0.071 RRMS vs HC 0.468 NMOSD vs RRMS 0.005
Female, n (%)	176/199 (88.4)	128/200 (64.0)	152/269 (56.5)	<0.001
Seropositive for AQP4-IgG, n (%)	84/132 (63.6)	-	-	-
First onset to diagnosis, yr [min-max]	4.5±5.1[0.0-35.0]	3.2±4.4[0.0-21.0]	-	0.006
Baseline treatment with DMT, n (%)	52(26.1%)	86(43.0%)	-	-
EDSS at baseline, median (IQR) [min, max]	2.0 (2.0) [0.0-9.0]	3.5 (3.0) [0.0-9.0]	-	<0.001
Brain segmentation volume without ventricles, ml [min-max]	1058.9±94.4[798.7-1390.1]	1080.1±121.5[742.6-1484.5]	1154.6±98.5[910.7-1434.0]	NMOSD vs HC <0.001 MS vs HC <0.001 NMOSD vs RRMS 0.108
Normalized brain volume, [min-max]	0.750±0.038[0.647-0.891]	0.731±0.045[0.590-0.858]	0.765±0.030[0.700-0.894]	<0.001†

Total volume of lesion, ml [min-max]	4.9±8.1[0.0-43.9]	12.7±17.9[0.0-134.0]	-	<0.001†
Deep-learning derived brain age				
Predicted brain age, yr (min-max)	46.4±16.0[18.8-77.5]	49.8±17.5[19.5-77.8]	39.3±13.7[14.8-73.8]	<0.001†
Brain age gap, yr (95% CI)	5.4±8.2 (4.3, 6.5)	13.0±14.7 (10.9, 15.0)	0.8±6.2 (0.1, 1.6)	<0.001†
Predicted brain age standard deviation, yr (95% CI)	6.0±3.0 (5.6, 6.5)	7.2±4.2 (6.6, 7.8)	4.8±1.1 (4.7, 4.9)	<0.001†
Follow up				
N with follow-up data, n (%)	85 (42.7)	124 (62.0)	-	
Mean follow-up time, yr [min-max]	5.8±1.9[1.9-9.9]	5.2±1.7[1.5-9.2]	-	0.020
EDSS worsening, n (%)	31 (36.5)	42 (33.9)	-	0.764

Continuous variables other than EDSS are reported as the mean ± standard deviation. EDSS are reported as the median (IQR).
† For all pairwise comparisons, i.e. for NMOSD vs HC, RRMS vs HC and NMOSD vs RRMS.
NMOSD: neuromyelitis optica spectrum disorder; MS: multiple sclerosis; EDSS: Expanded Disability Status Scale; IQR: Interquartile range; vs: versus. DMT: disease-modifying therapy.

Brain morphometry of the participants

Both the NMOSD and RRMS groups had lower brain volumes than the healthy controls (1080.1±121.5 ml and 1058.9±94.4 ml versus 1154.6±98.5 ml, both p<0.001). While the NMOSD and RRMS groups were not significantly different in raw brain volume (p=0.108), normalized brain volumes revealed less pronounced atrophy in NMOSD patients (0.750±0.038 versus 0.731±0.045, p<0.001). The lesion load in the NMOSD group was lower than that in the RRMS group (4.9±8.1 ml versus 12.7±17.9 ml, p<0.001) **(Table 1)**.

Performance of the brain age prediction model

Model training (using 9794 HCs) was terminated at epoch 108. The mean absolute error (MAE) before inverse linear bias correction was 2.63 years in the developmental validation set, and this model was used as the final model for further analysis.

The model was then tested using 462 images for internal (across-scanner) validation and 267 images for external validation (across-center). In the internal validation dataset, the MAE was 2.9±3.1 years, with no significant difference across scanner types (p=0.581, n=2). The Pearson's correlation coefficient (r) between age and brain age was 0.957. In the external validation set, the MAE was 4.5±3.9 years, and the Pearson's r was 0.890. The MAE was not significantly different across different centers (p=0.660, n=5; **Supplementary Table 2**).

Increased BAG in NMOSD and RRMS compared with healthy controls.

The difference in BAG among NMOSD patients, RRMS patients, and HCs was relatively

consistent across baseline chronological ages (Figure 1A). At baseline, patients with NMOSD had a significantly higher BAG than HCs (NMOSD - HC = 4.6 years, 95% CI 2.4-6.9, p<0.001), but patients with RRMS had a markedly higher BAG than HCs (MS - HC = 12.1 years, 95% CI 9.9-14.3, p<0.001). BAG was lower in NMOSD than in RRMS (NMOSD -RRMS = -7.5 years, 95% CI 5.2-9.9, p<0.001). (**Table 1, Figure 1B**).

Furthermore, we performed subgroup analysis of BAG in AQP4 seropositive versus seronegative NMOSD patients, as well as in NMOSD patients with versus without brain lesions. We demonstrated that there was no significant difference of BAG between the AQP-4 seropositive and seronegative subgroups (5.8±8.8 versus 4.2±6.9 years, p=0.256). However, BAG of patients with brain lesions was significantly higher than patients without brain lesions (7.1±8.5 versus 3.4±7.2 years, p=0.001) (Supplementary Table 5).

A significant difference in BAG across centers (p<0.001) was noted, although post hoc analysis revealed consistent trends in disease effects on BAG in all six centers (Figure 1C). Sample images and the corresponding output from both the NMOSD and RRMS groups were provided for better understanding (**Figure 1D-G**).

The correlation between raw and lesion-filled 3D T1WI images was very high (R^2=0.984, p<0.001, Supplementary Figure 3A). A Bland–Altman plot showed that the mean difference between raw and lesion-filled brain age was 0.28±2.11 years with no apparent systematic bias (**Supplementary Figure 3B**), indicating that the lesion filling process did not have a particular impact on the model.

Correlation of BAG with clinical variables

At baseline, univariate linear regression analysis demonstrated that BAG was positively associated with EDSS in both the NMOSD and RRMS groups (NMOSD r=0.217, β=0.86, p=0.002; RRMS r=0.268, β=2.31, p<0.001; **Figure 2A**). Normalized brain volume was inversely associated with BAG in both NMOSD and RRMS (NMOSD r=-0.202, β=-48.5, p<0.001; RRMS r=-0.384, β=-126.9, p<0.001; Figure 2B). Multivariable linear regression found that BAG was positively predictive of baseline EDSS independent of normalized brain volume and disease duration (NMOSD p=0.030; RRMS p=0.009; **Supplementary Table 3**).

We performed 1:1 nearest neighbor propensity score matching (PSM)[17] to exclude the possible confounding influence of clinical variables on BAG. This matching yielded adequate balance for all included coefficients. The mean BAG was 5.0±7.1 years in NMOSD and 11.1±12.7 years in RRMS after adjustment for sex, age at diagnosis, baseline EDSS, and normalized brain volume, with an estimated difference of -6.1 years [95% CI -8.7 to -3.4] years between NMOSD and RRMS (**Table 2**).

Table 2. NMOSD patients exhibits lower brain age gap over RRMS adjusted for sex, age at diagnosis, baseline EDSS and normalized brain volume with propensity score matching.

	NMOSD	RRMS	**p value**

N	119	119	-
Age at diagnosis, yr	39.6±13.2	39.9±11.7	0.855
Female, n (%)	96 (80.7)	97 (81.5)	1.000
First onset to diagnosis, yr	3.8±4.0	3.5±5.1	0.661
EDSS at baseline, median (IQR)	2.5 (2.0)	2.5 (2.0)	0.300
Normalized brain volume	0.745±0.038	0.742±0.042	0.538
Predicted brain age	44.5±15.5	50.0±16.9	0.008
Brain age gap	5.0±7.1	11.1±12.7	<0.001

Continuous variables other than EDSS are reported as the mean ± standard deviation. EDSS are reported as the median (IQR).
NMOSD: neuromyelitis optica spectrum disorder; MS: multiple sclerosis; EDSS: Expanded Disability Status Scale; IQR: Interquartile range.

The area under the curve (AUC) of the ROC for BAG in predicting progression was 0.599 in NMOSD and 0.670 in RRMS. The optimal cutoff of BAG was 6.1 (sensitivity 38.7%, specificity 81.5%) for NMOSD and 24 (sensitivity 50.0%, specificity 80.5%) for RRMS (Supplementary Figure 4). Kaplan–Meier survival analysis indicated that BAG was predictive of progression in both groups. For NMOSD patients, the median time to progression for BAG > 6.1 years was 5.79 years versus 7.99 years for BAG ≤ 6.1 years (p=0.003, Figure 2C). The median time to progression for BAG > 24.0 years was 5.36 years versus 8.95 years for BAG ≤ 24.0 years in RRMS patients (p=0.002, Figure 2D).

We utilized the Cox proportional hazards model to investigate whether BAG could be used to predict time to EDSS worsening independent of age at diagnosis, sex, disease duration, baseline EDSS and normalized brain volume. In univariate models, normalized brain volume and BAG were significantly associated with EDSS worsening in both NMOSD and RRMS patients (**Table 3**, univariate model). In a multivariable model, BAG was associated with EDSS worsening in NMOSD patients (HR=1.02 [1.00, 1.04]). p=0.027, **Table 3**), independent of normalized brain volume (p=0.158). However, neither normalized brain volume nor BAG was significant in theRRMS group in multivariable analysis. Interestingly, we found baseline EDSS to be negatively associated with EDSS worsening in NMOSD (multivariable Model p=0.001, **Table 3**).

Table 3. Univariable and multivariable Cox proportional hazards model analysis for predicting EDSS worsening by BAG, age at diagnosis, sex, duration between first onset to diagnosis, baseline EDSS and normalized brain volume.

	Univariable				**Multivariable**			
	NMOSD		**RRMS**		**NMOSD**		**RRMS**	
	Hazard Ratio	**p value**	**Hazard Ratio**	**p value**	**Hazard Ratio**	**p value**	**Hazard Ratio**	**p value**
N	85		124		85		124	
Number of events,	31 (36.5)		42 (33.9)		31 (36.5)		42 (33.9)	

n(%)								
Age at diagnosis, yr	1.04[1.00, 1.08]	0.032	0.99[0.96, 1.02]	0.540	1.02[0.98, 1.06]	0.398	-	
Sex=Male	0.52[0.07, 3.90]	0.527	1.23[0.65, 2.34]	0.523	-		-	
First onset to diagnosis, yr	0.97[0.89, 1.05]	0.416	1.00[0.91, 1.11]	0.890	-		-	
EDSS at baseline	0.68[0.54, 0.85]	<0.001	0.91[0.73, 1.12]	0.364	0.65[0.50, 0.83]	0.001	-	
Normalized brain volume (%)	0.90[0.81, 1.00]	0.049	0.91[0.85, 0.98]	0.009	0.92[0.81, 1.04]	0.158	0.92 [0.85, 1.02]	0.107
Brain age gap, yr	1.06[1.00, 1.13]	0.031	1.02[1.00, 1.04]	0.029	1.07[1.01, 1.14]	0.027	1.02 [0.98, 1.04]	0.633

NMOSD: neuromyelitis optica spectrum disorder; MS: multiple sclerosis; EDSS: Expanded Disability Status Scale.

Analysis of predicted standard deviation in brain age prediction.

The predicted standard deviation was positively associated with BAG in all three groups (linear model $p<0.001$ in HC and NMOSD, $p=0.011$ in RRMS, Supplementary Figure 5A). The mean standard deviation in NMOSD was higher than HC and lower than RRMS (Supplementary Figure 5B), which was consistent with the trend seen in BAG, indicating a higher model uncertainty in those images with greater discrepancy between apparent and chronological age. We examined scans with high model uncertainty and found that some of them could be attributed to low image quality or incomplete anatomical coverage (Supplementary Figure 5C), while others were not visually distinguishable from those with lower model uncertainty (Supplementary Figure 5D). To analyze whether the difference in BAG was driven by the difference in predicted standard deviation, we performed PSM with predicted standard deviation added as a covariate. The difference in BAG between NMOSD and RRMS, as well as NMOSD and RRMS versus HC, remained statistically significant after PSM adjusted for age, sex, duration to diagnosis, baseline EDSS, normalized brain volume and predicted standard deviation ($p<0.001$, **Supplementary Table 4**).

DISCUSSION

In this study, we developed a deep learning model to accurately predict age from 3D structural MRI scans and demonstrated its robustness in the context of multiple centers and MRI scanners. Using this model, the BAG was estimated to be approximately +5 years in NMOSD and +13 years in RRMS. Baseline BAG was independently predictive of EDSS worsening in both NMOSD and RRMS, suggesting its additional clinical value as a noninvasive biomarker for early triage, stratified follow-up management and clinical trial enrollment.

Previous nondeep learning studies on age prediction tasks reported 2.9- to 5.0-year MAEs on their validation sets[7,18,19] (some of which included multimodality-derived features, including fMRI and DTI), while deep learning studies reported validation MAEs as low as 2.14 years, such as in the original SFCN study[16]. We reached similar performance levels of MAE = 2.5 years in the developmental validation set, and the performance was maintained in an internal test set, demonstrating the usefulness of our model and highlighting the versatility and potential of deep-learning-based methods. We have also shown that the whole-brain CNN-based model was robust within scanners and centers, supporting the clinical use of the brain-age paradigm.

BAG has been investigated extensively as a comprehensive biomarker for accelerated aging. Increased BAG has been observed in dementia[20], epilepsy[21] and traumatic brain injury[22]. We report for the first time the meaningfulness of BAG in NMOSD as well as the difference between NMOSD and RRMS. We found a BAG of 5.4 [95% CI 4.3 to 6.5] years in NMOSD patients, which, although lower than RRMS, is still marked compared to healthy controls. The degree of BAG increase in NMOSD is similar to what has been reported in epilepsy (4.5 years)[21] and traumatic brain injury (4.7 years)[22].

BAG in NMOSD was positively associated with baseline EDSS score and whole-brain atrophy, with associations comparable to those in RRMS but with a generally less steep slope. BAG was also predictive of EDSS worsening in NMOSD, which is in line with the idea that BAG is a composite marker of abnormal aging and a disease-related brain. Furthermore, subgroup analysis of NMOSD patients demonstrated that BAG of patients with brain lesions was significantly higher than patients without brain lesions. This indicates that brain involvement may accelate brain aging in NMOSD patients. Future studies with larger sample size are required to validate this finding.

In a recent study of brain age using Gaussian Processes (GP) regression on multiple sclerosis, the authors reported 11.9 [95% CI 10.3 to 13.4] years BAG in RRMS patients in the European MAGNIMS cohort[7], which is consistent with our result of 13.0 [95% CI 10.9 to 15.0] years BAG in Chinese MS patients. Furthermore, increased BAG was predictive of EDSS worsening in MS, also consistent with previous work[7]. Even though we used a fundamentally different methodology and datasets, these results provide additional evidence for the usefulness of BAG in the evaluation of MS patients. Moreover, using deep learning can substantially shorten the runtime of the analysis pipeline. This acceleration in computation time is potentially of great benefit for widespread application in a clinical setting.

Comparing NMOSD and MS is difficult given the difference in confounding factors that may influence BAG. It has been reported previously that the atrophy patterns in NMOSD and MS are different. NMOSD exhibits more atrophy in the spinal cord but less atrophy in the brain[23], which can partially explain the lower BAG in NMOSD given the strong association between BAG and brain atrophy. To address the influence of confounding effects such as demographics and brain volume, we used propensity score matching to sample a subset with matched baseline confounding factors. In this matched subset, the difference in BAG between NMOSD and MS was still significantly different even when matched for normalized brain volume. This finding indicates that the brains of RRMS patients appear older than those of NMOSD patients even at the same level of atrophy, implying that BAG can be seen as a global estimation that integrates information beyond simple brain volumetry while being more accessible and informative than tables of volumetric measurements.

The uncertainty and distributional pattern of predicted brain age is an important field of research that has attracted little attention. A recent study modeled brain-age uncertainty with a single-layer neural network that addressed aleatoric uncertainty with quantile regression and epistemic uncertainty with the Monte Carlo drop-out technique[24]. In contrast to other studies that utilize quantile regression, the novel method in our study renders aleatoric uncertainty a natural derivative since the model output itself is a distribution instead of the point estimate used in previous studies[4]. Epistemic uncertainty was not derived in this study due to computational cost. Although the uncertainty correlated positively with BAG, the PSM analysis indicated that the BAG difference between NMOSD and RRMS remained statistically significant even after adjustment for predicted standard deviation. We observed that the predicted standard deviations were higher in those scans without enough information for brain age inference (i.e., low image quality, etc., and in those with a greater discrepancy between predicted and actual age. This observation suggests a potential use case for the predicted standard deviation. The quantification of individual-level uncertainty in this way could provide an integrated, intuitive metric for image quality control, especially in healthy people, as well as provide a measure of 'confidence' for applications in clinical contexts.

Our study has a few limitations. First, the follow-up duration was relatively short, and the sample size of patients with follow-up was small, which may have introduced selection bias. Second, although previous studies have suggested the longitudinal utility of brain age in healthy cohorts[6] and accelerated aging measured by BAG has been observed in MS cohorts[7], our cohort lacked sufficient follow-up assessments for this type of analysis. Finally, the interpretability of the results needs to be further improved; specifically, the anatomical meaning of brain age remains ill-defined. Deep-learning-based methods have been cast as 'black boxes'; however, tools such as class activation mapping, guided backpropagation and occlusion analysis are emerging that aim to extract mechanistic information from the network.[25] However, the translation of these methods to three-dimensional data is complex, and they have yet to be validated for use in interpreting medical imaging data. Additionally, our study relied on 3D T1WI MRI, which is not always available in clinical contexts. Future work will take advantage of brain-age models developed to work on routine clinical 2D scans[26].

In conclusion, NMOSD demonstrated a significant BAG compared with healthy controls, although less marked than RRMS. BAG is a predictive biomarker of EDSS worsening in both NMOSD and RRMS.

Acknowledgements Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.;Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Contributors RW and XX conception and design of the study, acquisition and analysis of data, drafting the manuscript. YL acts as the guarantor of the study and takes full responsibility for the work. YD, NZ, JS, HL, YL, FB, JH: conception and design of the study, acquisition and analysis of data. YL, CZ, XH, FZ, MH, RL, ZZ: acquisition and analysis of data. All authors revised the manuscript and approved the final draft.

Funding This work was supported by the National Science Foundation of China (No.81870958), the Beijing Municipal Natural Science Foundation for Distinguished Young Scholars (No. JQ20035), Beijing Young Scholarship, and the Capital's Funds for Health Improvement and Research (CFH2022-1-2042). FB is supported by the NIHR biomedical research center at UCLH.

Competing interests F.B acts as a consultant for Combinostics, Biogen-Idec, Janssen, Ixico, Merck-Serono, Novartis and Roche. He has received grants, or grants are pending, from the Amyloid Imaging to Prevent Alzheimer's Disease (AMYPAD) initiative, the Biomedical Research Centre at University College London Hospitals, the Dutch MS Society, ECTRIMS–MAGNIMS, EU-H2020, the Dutch Research Council (NWO), the UK MS Society, and the National Institute for Health Research, University College London. He has received payments for the development of educational presentations from Ixico and his institution from Biogen-Idec and Merck. He is con-founder of to Queen Square Analytics. He is on the editorial board of Radiology, Neuroradiology, Multiple Sclerosis Journal and Neurology. J.H.C. is a scientific consultant to and shareholder in BrainKey and Claritas Healthcare, as has worked as a consultant to Queen Square Analytics.

Ethics approval The study was approved by ethics committee of Beijing Tiantan Hospital,

Capital Medical University (No. KY-2019-050-02). Participants gave informed consent to participate in the study before taking part.

REFERENCES

1. Papais-Alvarenga RM, Vasconcelos CC, Carra A, et al. Central Nervous System Idiopathic Inflammatory Demyelinating Disorders in South Americans: A Descriptive, Multicenter, Cross-Sectional Study. *PLoS One*. 2015;10(7):e0127757. doi:10.1371/journal.pone.0127757

2. Trojano M, Avolio C, Manzari C, et al. Multivariate analysis of predictive factors of multiple sclerosis course with a validated method to assess clinical events. J Neurol Neurosurg Psychiatry. 1995;58(3):300-306. doi:10.1136/jnnp.58.3.300

3. Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*. 2019;18(5):459-480. doi:10.1016/S1474-4422(18)30499-X

4. Cole JH, Franke K. Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends Neurosci*. 2017;40(12):681-690. doi:10.1016/j.tins.2017.10.001

5. Zhai J, Li K. Predicting Brain Age Based on Spatial and Temporal Features of Human Brain Functional Networks. *Front Hum Neurosci*. 2019;13:62. doi:10.3389/fnhum.2019.00062

6. Cole JH, Poudel R, Tsagkrasoulis D, et al. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *Neuroimage*. 2017;163:115-124. doi:10.1016/j.neuroimage.2017.07.059

7. Cole PhD JH, Raffel MD J, Friede PhD T, et al. Longitudinal Assessment of Multiple Sclerosis with the Brain-Age Paradigm. *Ann Neurol*. 2020;88(1):93-105. doi:10.1002/ana.25746

8. Wingerchuk DM, Banwell B, Bennett JL, et al. International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology*. 2015;85(2):177-189. doi:10.1212/WNL.0000000000001729

9. Costelloe L, O'Rourke K, Kearney H, et al. The patient knows best: significant change in the physical component of the Multiple Sclerosis Impact Scale (MSIS-29 physical). *J Neurol Neurosurg Psychiatry*. 2007;78(8):841-844. doi:10.1136/jnnp.2006.105759

10. Ellis KA, Bush AI, Darby D, et al. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr*. 2009;21(4):672-687. doi:10.1017/S1041610209009405

11. Buckner RL, Roffman JL, Smoller JW. Brain Genomics Superstruct Project (GSP). Neuroinformatics Research Group, ed. Published online 2014. doi:10.7910/DVN/25833

12. Liu W, Wei D, Chen Q, et al. Longitudinal test-retest neuroimaging data from healthy young adults in southwest China. *Sci Data*. 2017;4:170017. doi:10.1038/sdata.2017.17

13. Isensee F, Schell M, Tursunova I, Brugnara G, Bonekamp D, Neuberger U, Wick A, Schlemmer HP, Heiland S, Wick W, Bendszus M, Maier-Hein KH, Kickingereder P. Automated brain extraction of multi-sequence MRI using artificial neural networks. *Hum Brain Mapp.* 2019; 1 – 13.

14. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron.* 2002; 33: 341-355.

15. Whitwell JL, Crum WR, Watt HC, Fox NC. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJNR Am J Neuroradiol*. 2001;22(8):1483-1489.

16. Peng H, Gong W, Beckmann CF, Vedaldi A, Smith SM. Accurate brain age prediction with lightweight deep neural networks. *Med Image Anal*. 2020;68:101871. doi:10.1016/j.media.2020.101871

17. Wang Y, Cai H, Li C, et al. Optimal caliper width for propensity score matching of three treatment groups: a Monte Carlo study. *PloS One*. 2013;8(12):e81045. doi:10.1371/journal.pone.0081045

18. Elliott ML, Belsky DW, Knodt AR, et al. Brain-age in midlife is associated with accelerated biological aging and cognitive decline in a longitudinal birth cohort. *Mol Psychiatry*. Published online 2019. doi:10.1038/s41380-019-0626-7

19. Smith SM, Elliott LT, Alfaro-Almagro F, et al. Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. Peelle JE, de Lange FP, Madan C, Nyberg L, eds. *eLife*. 2020;9:e52677. doi:10.7554/eLife.52677

20. Wang J, Knol MJ, Tiulpin A, et al. Gray Matter Age Prediction as a Biomarker for Risk of Dementia. *Proc Natl Acad Sci - PNAS*. 2019;116(42):21213-21218. doi:10.1073/pnas.1902376116

21. Pardoe HR, Cole JH, Blackmon K, Thesen T, Kuzniecky R, Human Epilepsy Project Investigators. Structural brain changes in medically refractory focal epilepsy resemble premature brain aging. *Epilepsy Res*. 2017;133:28-32. doi:10.1016/j.eplepsyres.2017.03.007

22. Cole JH, Leech R, Sharp DJ. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann Neurol*. 2015;77(4):571-581. doi:10.1002/ana.24367

23. Liu Y, Wang J, Daams M, et al. Differential patterns of spinal cord and brain atrophy in NMO and MS. *Neurology*. 2015;84(14):1465-1472. doi:10.1212/WNL.0000000000001441

24. Hahn T, Ernsting J, Winter NR, et al. An uncertainty-aware, shareable, and transparent neural network architecture for brain-age modeling. *Sci Adv*. 2022;8(1):eabg9471. doi:10.1126/sciadv.abg9471

25. Reyes M, Meier R, Pereira S, et al. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiol Artif Intell*. 2020;2(3):e190043. doi:10.1148/ryai.2020190043

26. Wood DA, Kafiabadi S, Busaidi AA, et al. Accurate brain-age models for routine clinical MRI examinations. *NeuroImage*. 2022;249:118871. doi:10.1016/j.neuroimage.2022.118871

Figure 1. Deep learning derived brain age vs chronological age in neuromyelitis optica spectrum disease (NMOSD), multiple sclerosis (MS), and healthy control (HC). (A) Deep learning derived brain age vs chronological age in NMOSD, MS and HC group. Predicted brain age is consistently higher in NMOSD and MS group compared with HC group. (B) NMOSD patients exhibits lower brain age gap (BAG) over MS and lower BAG over HCs. (C) The difference of BAG across centers in NMOSD, MS and HC group. The tendency that MS-BAG > NMOSD-BAG > HC-BAG remains consistent even if there are significant differences across centers. (D, E, F, G) A sample input and prediction result of NMOSD and MS patient. Solid line indicates brain age estimation and dashed lines indicate standard deviation of prediction. The predicted brain age was 42.0±5.1 years for (D) and 70.0±7.2 years for (E), yielding BAG of 6.0 years and 42.0 years namely. Both (D) and (E) experienced disability progression in follow-up sessions. Predicted brain age for (F) and (G) was 39.2±5.6 years and 66.1±4.4 years yielding BAG of -1.8 years and 12.1 years namely. These patients with lower BAG didn't experience disability progression within follow up period.

Figure 2. Correlation of brain age gap (BAG) with clinical variables and its prognostic value. (A) Increased BAG was associated with more severe baseline disability status in both neuromyelitis optica spectrum disease (NMOSD), multiple sclerosis (MS), which was more prominent in MS patients. (B) Normalized brain volume was strongly negatively associated with BAG both in NMOSD and MS indicating possible contribution of atrophy in increased BAG. (C, D) Survival curve of BAG predicting disability progression in NMOSD and MS patients. Cutoff point was determined by 80% specificity. Operating cutoff point for for NMOSD is set to BAG > 6.1 (sensitivity 38.7%, specificity 81.5%),MS is set to BAG > 24.0(sensitivity 50.0%, specificity 80.5%).