Original research

Brain age gap in neuromyelitis optica spectrum disorders and multiple sclerosis

Ren Wei,[1] Xiaolu Xu [iD],[1] Yunyun Duan [iD],[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 [iD],[9,10] Yaou Liu [iD][1]

Check for updates

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 and RRMS in a large multicentre dataset.

METHODS
Participants
Data from patients with NMOSD and RRMS were retrospectively collected from six tertiary neurological centres in China covering the period between November 2009 and April 2018. Patients who fulfilled the following criteria were included: (a) confirmed diagnosis of NMOSD according to 2015 revised diagnostic criteria[8] or RRMS according to 2017 McDonald criteria[9]; (b) complete demographic and clinical information, including baseline EDSS score and disease duration and (c) good quality baseline three-dimensional (3D) T1-weighted structural images (T1WI). Clinical evaluation, diagnosis, treatment and follow-up assessments of the participants were conducted at each centre by local neurologists with expertise in demyelinating diseases.

baseline EDSS >5.5, as previously published.[10]

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 Alzheimer's Disease Neuroimaging Initiative (ADNI), The Australian Imaging, Biomarkers and Lifestyle (AIBL),[11] Brain Genomics Superstruct Project (GSP)[12] and Southwest University Longitudinal Imaging Multimodal (SLIM),[13] as well as a group of healthy people scanned at Beijing Tiantan Hospital from January to December 2019 (online supplemental table 1, online supplemental 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 to April 2020 on two different scanners (see online supplemental table 1). The external validation dataset included HCs from the multicentre NMOSD and MS cohorts (n=267).

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

	NMOSD	RRMS	HCs	P value
Baseline				
N	199	200	269	
Age at baseline, year (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 versus HC 0.071 RRMS versus HC 0.468 NMOSD versus 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, year (min–max)	4.5±5.1 (0.0–35.0)	3.2±4.4 (0.0–21.0)	–	0.006
Baseline use of 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 versus HC <0.001 MS versus HC <0.001 NMOSD versus RRMS 0.108
Normalised 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, year (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, year (95% CI)	5.4±8.2 (4.3 to 6.5)	13.0±14.7 (10.9 to 15.0)	0.8±6.2 (0.1 to 1.6)	<0.001*
Predicted brain age SD, year (95% CI)	6.0±3.0 (5.6 to 6.5)	7.2±4.2 (6.6 to 7.8)	4.8±1.1 (4.7 to 4.9)	<0.001*
Follow-up				
N with follow-up data, n (%)	85 (42.7)	124 (62.0)	–	
Mean follow-up time, year (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±SD. EDSS is reported as the median (IQR).
*For all pairwise comparisons, that is, for NMOSD versus HC, RRMS versus HC and NMOSD versus RRMS.
DMT, disease-modifying therapy; EDSS, Expanded Disability Status Scale; HC, healthy control; MS, multiple sclerosis; NMOSD, neuromyelitis optica spectrum disorder; RRMS, relapsing-remitting multiple sclerosis.

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-centre). 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 centres (p=0.660, n=5; online supplemental table 2).

Increased BAG in NMOSD and RRMS compared with healthy controls

The difference in BAG among patients with NMOSD, patients with RRMS and HCs was relatively consistent across baseline chronological ages (figure 1A). At baseline, patients with

HC=4.6 years, 95% CI 2.4 to 6.9, p<0.001), but patients with

years, 95% CI 9.9 to 14.3, p<0.001). BAG was lower in

CI 5.2 to 9.9, p<0.001) (table 1, figure 1B).

Furthermore, we performed subgroup analyses of BAG in AQP4 seropositive versus seronegative patients with NMOSD, as well as in patients with NMOSD with versus without brain lesions. We observed that there was no significant difference in BAG between the AQP4 seropositive and seronegative subgroups

(5.8±8.8 vs 4.2±6.9 years, p=0.256). However, the BAG in patients with brain lesions was significantly higher than those without (7.1±8.5 vs 3.4±7.2 years, p=0.001) (online supplemental table 5).

A significant difference in BAG across centres (p<0.001) was noted, although post hoc analysis revealed consistent trends in disease effects on BAG in all six centres (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 (R²=0.984, p<0.001, online supplemental 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 (online supplemental 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). Normalised brain volume was inversely associated with BAG

β   β

figure 2B). Multivariable linear regression found that BAG was positively predictive of baseline EDSS independent of normalised brain volume and disease duration (NMOSD p=0.030; RRMS p=0.009; online supplemental table 3).

).

patients with RRMS (p=0.002, figure 2D).

We used the Cox proportional hazards model to investigate v 0 s I v

We performed 1:1 nearest neighbour propensity score matching (PSM)[18] 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 normalised

table 2).

The area under the curve of the receiver operating characteristic for BAG in predicting progression was 0.599 in NMOSD and 0.670 in RRMS. The optimal cut-off of BAG was 6.1 (sensitivity 38.7%, specificity 81.5%) for NMOSD and 24 (sensitivity 50.0%, specificity 80.5%) for RRMS (online supplemental figure 4). Kaplan-Meier survival analysis indicated that BAG was predictive of progression in both groups. For patients with NMOSD, the median time to progression for BAG >6.1 years

figure 2C). The median time to progression for BAG >24.0

Analysis of predicted SD in brain age prediction

The predicted SD was positively associated with BAG in all three groups (linear model p<0.001 in HC and NMOSD, p=0.011 in RRMS, online supplemental figure 5A). The mean SD in NMOSD was higher than HC and lower than RRMS (online supplemental 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 (online supplemental figure 5C), while others were not visually distinguishable from those with lower model uncertainty (online supplemental figure 5D). To analyse whether the difference in BAG was driven by the difference in predicted SD, we performed PSM with predicted SD 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, normalised brain volume and predicted SD (p<0.001, online supplemental 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 centres 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 non-invasive biomarker for early triage, stratified follow-up management and clinical trial enrolment.

Previous non-deep learning studies on age prediction tasks reported 2.9-year to 5.0-year MAEs on their validation sets[7 19 20] (some of which included multimodality-derived features, including functional MRI and diffusion tensor imaging), while deep learning studies reported validation MAEs as low as 2.14 years, such as in the original SFCN study.[17] 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 centres, supporting the clinical use of the brain age paradigm.

BAG has been investigated extensively as a comprehensive biomarker for accelerated ageing. Increased BAG has been observed in dementia,[21] epilepsy[22] and traumatic brain injury.[23] 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 patients with NMOSD, which, although lower than RRMS, is still marked

indicates that the brains of patients with RRMS appear older than those of patients with NMOSD 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 modelled brain age uncertainty with a single-layer neural network that addressed aleatoric uncertainty with quantile regression and epistemic uncertainty with the Monte Carlo dropout technique.[25] In contrast to other studies that use 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 SD. We observed that the predicted SD were higher in those scans without enough information for brain age inference (ie, 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 SD. 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 ageing 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.[26] However, the translation of these methods to 3D 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 two-dimensional scans.[27]

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

Author affiliations
[1]Department of Radiology, Beijing Tiantan Hospital, Beijing, China
[2]Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin Medical University General Hospital, Tianjin, China
[3]Department of Radiology, Huashan Hospital Fudan University, Shanghai, China
[4]Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[5]Department of Neurology, China-Japan Union Hospital of Jilin University, Changchun, China
[6]Department of Radiology, The First Affiliated Hospital of Nanchang University, Nanchang, China
[7]Department of Neurology, Beijing Tiantan Hospital, Beijing, China
[8]Department of Radiology and Nuclear Medicine, Neuroscience Campus Amsterdam, VU University Medical Centre Amsterdam, Amsterdam, The Netherlands
[9]Centre for Medical Image Computing, Department of Computer Science, University College London, London, UK
[10]Dementia Research Centre, Queen Square Institute of Neurology, University College London, London, UK

Twitter James H Cole @JamesCole_Neuro

ORCID iDs
Xiaolu Xu http://orcid.org/0000-0001-9963-529X
Yunyun Duan http://orcid.org/0000-0002-8753-0260
James H Cole http://orcid.org/0000-0003-1908-5588
Yaou Liu http://orcid.org/0000-0002-9930-0331

REFERENCES