

Gray Matter Age Prediction as a Biomarker for Risk of Dementia

Wang, Johnny; Knol, Maria J.; Tiulpin, Aleksei; Dubost, Florian; de Bruijne, Marleen; Vernooij, Meike W.; Adams, Hieab H.H.; Ikram, M. Arfan; Niessen, Wiro J.; Roshchupkin, Gennady V.

DOI

[10.1073/pnas.1902376116](https://doi.org/10.1073/pnas.1902376116)

Publication date

2019

Document Version

Final published version

Published in

Proceedings of the National Academy of Sciences of the United States of America

Citation (APA)

Wang, J., Knol, M. J., Tiulpin, A., Dubost, F., de Bruijne, M., Vernooij, M. W., Adams, H. H. H., Ikram, M. A., Niessen, W. J., & Roshchupkin, G. V. (2019). Gray Matter Age Prediction as a Biomarker for Risk of Dementia. *Proceedings of the National Academy of Sciences of the United States of America*, 116(42), 21213-21218. <https://doi.org/10.1073/pnas.1902376116>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Gray Matter Age Prediction as a Biomarker for Risk of Dementia

Johnny Wang^{a,b,1}, Maria J. Knol^{c,1}, Aleksei Tiulpin^d, Florian Dubost^a, Marleen de Bruijne^{a,e}, Meike W. Vernooij^{c,f}, Heiab H. H. Adams^{c,f}, M. Arfan Ikram^c, Wiro J. Niessen^{a,f,g,2}, and Gennady V. Roshchupkin^{a,c,f,2,3}

^aDepartment of Medical Informatics, Erasmus MC University Medical Center, 3015 CE, Rotterdam, The Netherlands; ^bMechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, 2628 CJ, Delft, The Netherlands; ^cDepartment of Epidemiology, Erasmus MC University Medical Center, 3015 CE, Rotterdam, The Netherlands; ^dResearch Unit of Medical Imaging, Physics and Technology, University of Oulu, 90014, Oulu, Finland; ^eDepartment of Computer Science, University of Copenhagen, 1165 Copenhagen, Denmark; ^fDepartment of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, 3015 CE, Rotterdam, The Netherlands; and ^gImaging Physics, Faculty of Applied Sciences, Delft University of Technology, 2628 CJ, Delft, The Netherlands

Edited by Marcus E. Raichle, Washington University in St. Louis, St. Louis, MO, and approved September 10, 2019 (received for review February 13, 2019)

The gap between predicted brain age using magnetic resonance imaging (MRI) and chronological age may serve as a biomarker for early-stage neurodegeneration. However, owing to the lack of large longitudinal studies, it has been challenging to validate this link. We aimed to investigate the utility of such a gap as a risk biomarker for incident dementia using a deep learning approach for predicting brain age based on MRI-derived gray matter (GM). We built a convolutional neural network (CNN) model to predict brain age trained on 3,688 dementia-free participants of the Rotterdam Study (mean age 66 ± 11 y, 55% women). Logistic regressions and Cox proportional hazards were used to assess the association of the age gap with incident dementia, adjusted for age, sex, intracranial volume, GM volume, hippocampal volume, white matter hyperintensities, years of education, and APOE $\epsilon 4$ allele carriership. Additionally, we computed the attention maps, which shows which regions are important for age prediction. Logistic regression and Cox proportional hazard models showed that the age gap was significantly related to incident dementia (odds ratio [OR] = 1.11 and 95% confidence intervals [CI] = 1.05–1.16; hazard ratio [HR] = 1.11, and 95% CI = 1.06–1.15, respectively). Attention maps indicated that GM density around the amygdala and hippocampi primarily drove the age estimation. We showed that the gap between predicted and chronological brain age is a biomarker, complementary to those that are known, associated with risk of dementia, and could possibly be used for early-stage dementia risk screening.

deep learning | dementia | age prediction | magnetic resonance imaging | voxel-based morphometry

The human brain continuously changes throughout the entire lifespan. These changes partially reflect a normal aging process and are not necessarily pathological (1). However, neurodegenerative diseases, including dementia, also affect brain structure and function (2, 3). Therefore, a better understanding and modeling of normal brain aging can help to disentangle these two processes and improve the detection of early-stage neurodegeneration.

Age prediction models based on brain MRI are a popular trend in neuroscience (4–7). The difference between predicted and chronological age is thought to serve as an important biomarker reflecting pathological processes in the brain. Several recent studies showed the relation between accelerated brain aging and various disorders, such as Alzheimer’s disease (8), schizophrenia, epilepsy, or diabetes (7, 9, 10).

In recent years, CNNs have become the methodology of choice for analyzing medical images. These models are able to learn complex relations between input data and desired outcomes. Recent studies (11, 12) were able to demonstrate that CNN models can be successfully applied in brain MRI-based age prediction (5, 6).

Although cross-sectional studies have suggested that the gap between predicted and chronological age may serve as a biomarker for dementia diagnosis, it remains unclear whether this is also the

case for the years preceding dementia diagnosis (5, 7). It was shown in recent research that the brain age gap is associated with mortality risk (13). Longitudinal studies examining the link between such a gap and incident dementia are lacking and are crucial for validation of this biomarker for early-stage neurodegeneration detection. Using a deep learning (DL) model, we investigated the association of the GM age gap with incident dementia in a large population-based sample of middle-aged and elderly subjects.

Methods

Study Population. Data were acquired from the Rotterdam Study, an ongoing population-based cohort study among the inhabitants of Ommoord, a suburb of Rotterdam, the Netherlands (14, 15). More details of the study design and population are described in *SI Appendix, Methods 1*.

Data from the Rotterdam Study are not publicly available due to informed consent and legal restrictions (e.g., General Data Protection Regulation law in the European Union). However, specific requests for access to the data can be addressed to the Rotterdam Study Management Team that assesses the proposals and adjudicates access—in line with national and international regulations—on a case-by-case basis.

Significance

The difference between brain age estimated from MRI and chronological age is thought to serve as an important biomarker reflecting pathological processes in the brain. Several recent studies showed the relation between accelerated brain aging and various disorders. However, until now, the utility of such an age difference for preclinical screening using longitudinal studies was absent. To fill this gap, we first built a deep learning model using brain MRI from a population-based study including 5,496 participants. And then, using follow-up information, we observed that this age difference was significantly associated with the risk of dementia. Therefore, our study shows that the difference between MRI-brain predicted and chronological age is potentially a biomarker for early dementia risk screening.

Author contributions: J.W., M.J.K., M.W.V., M.A.I., W.J.N., and G.V.R. designed research; J.W., M.J.K., A.T., F.D., M.d.B., H.H.H.A., and G.V.R. performed research; J.W., M.J.K., A.T., F.D., and G.V.R. analyzed data; and J.W., M.J.K., M.W.V., M.A.I., W.J.N., and G.V.R. wrote the paper.

Competing interest statement: The authors declare no competing interest. W.J.N. is co-founder, scientific lead, and shareholder of Quantib BV. Other authors report no biomedical financial interests.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹J.W. and M.J.K. contributed equally to this work.

²W.J.N. and G.V.R. contributed equally to this work.

³To whom correspondence may be addressed. Email: g.roshchupkin@erasmusmc.nl.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1902376116/-DCSupplemental.

First published October 1, 2019.

images, we provide information about the sex of the subject. This allows the network to adjust for GM differences between male and female subjects.

The dataset, excluding subjects with incident dementia, was randomly split into 3 sets: training (3,688 subjects), validation (1,099 subjects), and test (550 subjects). Subjects with incident dementia (159 subjects) were put in a fourth independent dataset. The CNN was trained using the training set as described in *SI Appendix, Methods 4*. For training, we used all available scans for each subject. Prediction accuracy was assessed on the test set. Model accuracy was measured based on the absolute gap, or mean absolute error (MAE) of prediction, i.e., the difference between model output and real chronological age ($gap = age_{\text{brain,predicted}} - age_{\text{chronological}}$). Given the design of the Rotterdam Study, several follow-up scans were available for some subjects. For training, we used all available scans for each subject. These training methods allowed us to increase the number of training images thereby introducing a natural type of data augmentation. **Attention mapping.** We retrieved attention maps from the trained networks using gradient-weighted class activation mapping (24). Attention maps show which areas on the subject GM image are more important for age prediction. More details about implementation of attention maps can be found in *SI Appendix, Methods 5*.

Statistical Analysis. Reproducibility of the CNN age prediction was quantified using the intraclass correlation coefficient (ICC[3,1]), computed on a subset of 80 persons out of the test set who were scanned twice with a time interval of 1–9 wk (25).

In order to be able to compare our findings with previous studies, logistic regression models and Cox proportional hazard models were used to assess the association between the age gap and the incidence of dementia. We adjusted the regression models for biomarkers, which are known for their relation with dementia: age and sex (model I); additional GM volume, ICV, hippocampal volume, and WMHs (model II); and years of education and *APOE ε4* carriership (model III) (26, 27). The logistic regression model used the occurrence of dementia development during follow-up as output. The proportional hazards and linearity assumption were met for the Cox proportional hazard models. Python and R were used to perform the statistical analyses (28–31).

Results

The study population characteristics are described in Table 1. The algorithm was trained and validated on random subsets of

Table 2. Quantitative analysis of the attention map per brain region. Mean and fifth quintiles (lower boundary) of attention map intensity per brain region are listed. Brain regions are grouped by lobes

Brain region	Size (voxels)	Attention map intensity	
		Mean	Fifth quartile
Temporal lobe			
Amygdala	4,398	0.71	0.98
Hippocampus	6,687	0.61	0.80
Anterior temporal lobe medial part	22,842	0.54	0.78
Superior temporal gyrus, anterior part	14,369	0.54	0.74
Lateral occipitotemporal gyrus (gyrus fusiformis)	12,908	0.53	0.62
Posterior temporal lobe	143,237	0.52	0.68
Superior temporal gyrus, central part	42,794	0.52	0.68
Gyri parahippocampalis and ambiens	13,767	0.51	0.63
Medial and inferior temporal gyri	55,102	0.50	0.68
Anterior temporal lobe lateral part	11,999	0.49	0.65
Insula and cingulate gyri			
Cingulate gyrus anterior part (supragenual)	24,751	0.53	0.63
Cingulate gyrus posterior part	24,235	0.52	0.64
Insula	44,328	0.51	0.64
Frontal lobe			
Subcallosal area	788	0.70	0.98
Posterior orbital gyrus	15,061	0.54	0.72
Straight gyrus (gyrus rectus)	11,826	0.54	0.67
Inferior frontal gyrus	55,754	0.53	0.72
Superior frontal gyrus	166,766	0.52	0.77
Precentral gyrus	106,145	0.52	0.77
Medial orbital gyrus	18,554	0.52	0.77
Presubgenual anterior cingulate gyrus	2,451	0.52	0.61
Middle frontal gyrus	161,999	0.51	0.74
Anterior orbital gyrus	19,514	0.51	0.73
Lateral orbital gyrus	11,112	0.51	0.77
Subgenual anterior cingulate gyrus	4,287	0.50	0.71
Occipital lobe			
Cuneus	28,209	0.57	0.67
Lingual gyrus	36,627	0.55	0.65
Lateral remainder of occipital lobe	131,852	0.54	0.73
Parietal lobe			
Superior parietal gyrus	130,908	0.54	0.74
Remainder of parietal lobe (including supramarginal and angular gyri)	131,972	0.52	0.75
Postcentral gyrus	89,087	0.52	0.74
Central structures			
Nucleus accumbens	888	0.89	0.99
Thalamus	20,953	0.61	0.79
Putamen	14,502	0.60	0.74
Pallidum (globus pallidus)	3,835	0.58	0.69
Caudate nucleus	12,229	0.56	0.67

subjects with mean age 66.09 ± 10.76 y and 55% females; and mean age 64.84 ± 9.69 y and 54% females, respectively. The following results are reported for the test set (mean age 64.85 ± 10.82 y and 55% females).

Network Performance. The overall performance measured on the test set was $MAE = 4.45 \pm 3.59$ y (Fig. 1) with a correlation between chronological and predicted brain age of 0.85 (P value = 4.76×10^{-156}). A reproducibility score of $ICC = 0.97$ (95% CI 0.96–0.98) was achieved. No significant difference in prediction was found between male and female subjects (P value = 0.34), and detailed numbers are provided in *SI Appendix, Text 1*.

Attention map. *SI Appendix, Fig. S5* shows the global attention map of the test set, indicating the areas contributing to age prediction in bright color, as well as the increase in attention map values over age. We found that the amygdala and hippocampus are not only important for predicting brain age, but also that these regions grow more important with increasing chronological age, which is shown in *SI Appendix, Fig. S5B*. A quantitative analysis per brain region is presented in Table 2, which shows that highest mean intensities were computed for the nucleus accumbens (0.89) and amygdala (0.71). Highest intensity quintiles were computed for the nucleus accumbens (0.99), amygdala (0.98), and subcallosal area (0.98).

Logistic Regression. We computed a logistic regression for the three models as shown in Table 3. The age gap was significantly associated with dementia incidence while age, sex, GM volume, ICV volume, hippocampal volume, WMH volume, years of education, and the *APOE* $\epsilon 4$ allele carriership were included in the model with model III: OR = 1.09 (95% CI 1.04–1.14) per year age gap.

Survival Analysis. As shown in Table 3 and Fig. 2, the age gap was significantly associated with the incidence of dementia with model III, HR = 1.09 (95% CI 1.04–1.14) per year age gap. These associations were similar in a subsample with a follow-up time for indecent dementia of more than 5 y, model III, HR = 1.09 (95% CI 1.01–1.16) per year age gap.

Gap-Associated Features. *SI Appendix, Table S1* and *Fig. S7* show a list of features that can affect the brain pathology and may be associated with the gap (10). Significantly lower values were found for GM volume, hippocampal volume, and WMH volume in the highest quintile.

Discussion

In a large sample of community-dwelling middle-aged and older adults, using a DL model for brain age prediction on MRI-derived

GM tissue density, we found that the gap between predicted brain age and chronological age was related to an increased risk of dementia, independent of standard established risk factors for dementia.

Our trained CNN model showed a similar MAE value in age prediction compared to previous studies that use a multimodal data model (5) and DL-based approach (6), which achieved performances of $MAE = 4.29$ and $MAE = 4.16$, respectively. Previous studies looked cross sectionally (5, 6) at the association of the age gap and dementia occurrences, while in the current study, we evaluated associations in longitudinal data. As non-reversible pathological changes already occur years prior to diagnosis, identifying early-stage biomarkers for dementia is of importance. The age gap has the potential to be utilized alongside other clinical risk factors and biomarkers to separate the population into categories with sufficiently distinct degrees of risk to drive clinical or personal decision-making, e.g., dementia screening and informed life planning.

Moreover, we retrieved attention maps from the model, showing the relative importance of different brain regions for age prediction. While the network looks at the entire GM (*SI Appendix, Fig. S6*), the attention pattern is quite complex, which suggests that the gap holds more specific information than global measures of GM volume when predicting brain age. This was further established by the association found between the gap and the incident dementia, which remained significant after adjusting for total GM volume. Interestingly, based on the attention maps, the amygdala and hippocampus, in particular, are relatively more important for age prediction, also increasing in attention and map intensity with older subjects (*SI Appendix, Fig. S5B*). This is in accordance to literature where significant negative associations between GM volume and age have been reported for these regions (2, 26). Atrophy of these two structures has also shown to be more prevalent in dementia patients, including years before diagnosis (32, 33). Yet, even after adjusting for hippocampal volume, the association between the age gap and the risk of dementia remained significant. This shows that the features which the neural network extracts from images go beyond just global or local volumetric measurements. A more in-depth evaluation of the attention map can be found in *SI Appendix, Text 2*.

Limitations. We were not able to perfectly predict the age for healthy subjects based only on MRI. We assume that, due to biological similarity of the brain within a range of several years, there will always be an according level of uncertainty in the age prediction.

Table 3. Association of gap between brain age and chronological age with incident dementia assessed by logistic regression and Cox proportional hazards models, both in the total study sample and in a subsample with a minimum follow-up time of 5 years

Model	Logistic regression			Cox regression		
	n/N	OR (95% CI)	P value	n/N	HR (95% CI)	P value
Total sample						
Model I	159/1808	1.15 (1.10–1.20)	2.66×10^{-10}	159/1808	1.15 (1.11–1.20)	1.02×10^{-12}
Model II	154/1790	1.09 (1.04–1.14)	4.77×10^{-4}	154/1790	1.09 (1.05–1.15)	2.27×10^{-5}
Model III	150/1714	1.09 (1.04–1.14)	7.97×10^{-4}	150/1714	1.09 (1.04–1.14)	9.62×10^{-5}
Sample follow-up time >5 y						
Model I	62/1366	1.11 (1.04–1.18)	1.26×10^{-3}	62/1366	1.13 (1.06–1.20)	1.38×10^{-4}
Model II	60/1352	1.08 (1.01–1.16)	2.48×10^{-2}	60/1352	1.10 (1.02–1.17)	7.58×10^{-3}
Model III	58/1305	1.08 (1.01–1.16)	3.38×10^{-2}	58/1305	1.09 (1.01–1.16)	1.78×10^{-2}

Model I: age + sex.
 Model II: model I + GM volume + ICV + hippocampal volume + WMH volume.
 Model III: model II + years of education + *APOE* $\epsilon 4$ carrier status.
 CI; OR; HR; number of cases (n); total number of participants (N).

