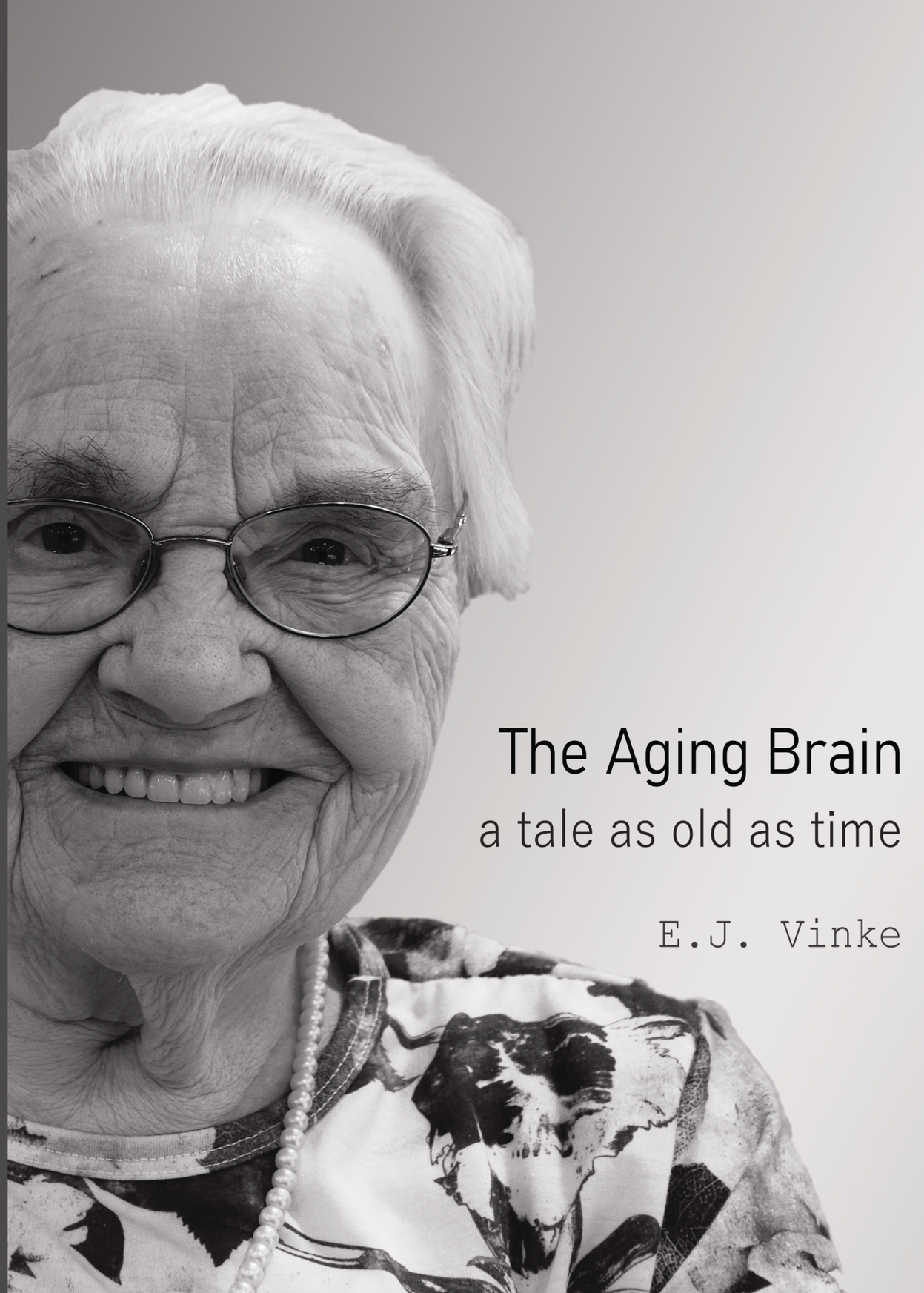
The Aging Brain

E.J. Vinke



The Aging Brain

a tale as old as time

E.J. Vinke

The Aging Brain

a tale as old as time

EJ Vinke

Acknowledgements

The work presented in this thesis was conducted at the Department of Radiology and Nuclear Medicine and the Department of Epidemiology of the Erasmus MC - University Medical Center, Rotterdam, The Netherlands.

The studies described in this thesis used data from the Rotterdam Study, which is supported by: Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Netherlands Genomics Initiative (NGI), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The work in this thesis was further supported by the European Research Council (ECR) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 666992 and grant agreement No. 678543).

The contributions of inhabitants of the Ommoord district and the general practitioners and pharmacists who took part in the Rotterdam Study are gratefully acknowledged.

Financial support for the publication of this thesis was kindly provided by Erasmus Medical Center Rotterdam, Erasmus University Rotterdam and Stichting Alzheimer Nederland.

ISBN/EAN: 978-94-6419-675-7
Cover design and lay-out: EJ Vinke & RS Eijgelaar
Printing: Gildeprint - Enschede

The Aging Brain

a tale as old as time

Het verouderende brein

een eeuwenoud verhaal

Proefschrift

Ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof. dr. A.L. Bredenoord

en volgens besluit van het college voor promoties.

De openbare verdediging zal plaatsvinden op

vrijdag 3 februari 2023 om 13.00 uur

door

Elisabeth Janine Vinke

geboren te Arnhem

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof. dr. M.W. Vernooij
	Prof. dr. M.A. Ikram
Overige leden:	Prof. dr. E.H.G. Oei
	Prof. dr. P.E. Slagboom
	Prof. dr. H.J. Grabe

Table of Content

Prologue

When I think about the aging process, the allusion “tale as old as time”, for the fellow Disney-enthusiasts known as the opening line from a song sung by Mrs. Potts, the enchanted teapot, comes to mind. A tale as old as time, meaning a tale that is ancient, timeless and its form is highly dependent on the context of that specific time.

Understanding the aging process has been a quest that dates back to the beginning of philosophical and medical thinking. Like it still is today, the perspectives on aging could be described as a balancing act between the ideas of illness and of natural decline. The definition of what is normal or healthy and what is disease is so intertwined, that one cannot exist without the other. Though the focus of research in the field of medicine has mostly focussed on understanding diseases rather than natural decline. I hope that after reading this thesis, you agree with me that the field of aging and brain aging is a fascinating and crucial one.

1

General Introduction

"Research on aging has emphasized losses"

- John W. Rowe and Robert L. Kahn, 1987 in Human Aging: Usual and Successful -

In the nineteenth and twentieth century, in general the 'normal' aging process was considered the age-related physiologic, cognitive and sociologic changes that occurred, in the absence of disease. Examples of age-related changes, or as Rowe and Kahn described them: *losses*, were: loss of hearing, vision, bone density, lung function, renal function, increase in systolic blood pressure and decline in cognitive and motor function.

Though this concept of normality has provided an important foundation for our understanding of the aging process today, the emphasis on normality had major limitations, as was pointed out by Rowe and Kahn[1] in 1987:

"In short, the emphasis on "normal" aging focuses attention on learning what most older people do and do not do, what physiologic and psychologic states are typical. It tends to create a gerontology of the usual".

They stated that by defining normal aging with merely absence of (identifiable) disease, the heterogeneity among the non-diseased aging individuals is neglected. The division between diseased and non-diseased fails to recognize this variance. They urged to focus on this heterogeneity and the development of an additional conceptual distinction within the normal category. To avoid creating a "gerontology of the usual", they suggested to make a distinction between usual and successful aging. With the concept of successful aging they hoped to stimulate research on the criteria and determinants of successful aging, and identify proper targets for interventions with "normal" elderly. Though there is still no consensual definition of healthy or successful aging[2], Rowe and Kahn's view on aging and their concept of successful aging has been a major contributor to moving the aging field to where it is today.

In aging and brain aging research it all boils down to the question what is considered to be the "norm" (e.g. normal, physiological, average, successful, healthy), for the aging process that you are researching. What someone considers to be "normal" is

highly dependent on a person's values, background, culture, environment and experiences. However, the lack of a universal definition of "normal" brain aging should not stop us from trying to understand brain aging, to ultimately unravel when, how and where in the brain deviations from the physiological aging take place towards pathophysiological degeneration and ultimately clinical disease.

The Aging Brain

The aging process is biologically complex and there is a large variation in the effects of aging on the human body between as well as within individuals. Similarly, functional and structural changes in the brain that occur with advancing age show large intra- and inter-individual variability. Irrespective of the definition of "normal" brain aging and the interpretation of its variability, what we generally observe with advancing age is a shrinking brain, called brain atrophy. Brain atrophy, also known as cerebral atrophy, is the macroscopic manifestation of the loss of neurons, synapses, glia and demyelination that occurs on a microscopic level. On structural brain imaging, cerebral atrophy is reflected in the loss of whole brain or regional brain tissue volume. Zooming in on the white matter of the brain, which embodies the connections and communication of cortical regions with each other and with subcortical regions, loss of microstructural integrity and ultimately the formation of lesions are associated with advancing age. Age-related changes in cognition vary considerable across cognitive domains and across individuals.[3] Basic cognitive functions that are most affected by age are memory and attention. These fundamental cognitive functions are thought to account for a large part of the variance observed in higher-level cognitive processes.

Similar to aging in general, conceptualizing brain aging comes with many challenges. The large variability in the effects of aging on brain structure and function leads to a broad spectrum of age-related changes. An important challenge is to define the part(s) of the spectrum where "normal" brain aging ends and (pre)clinical disease starts. Taking Alzheimer's disease as an example, macrostructural brain changes in Alzheimer's disease are typically characterized by severe medial temporal lobe atrophy.[4,5] Furthermore, cognitive decline, often first recognized as memory impairment, is a typical feature of Alzheimer's disease.[5] With brain atrophy being a hallmark of structural brain aging and loss of memory function a hallmark of cognitive aging, the challenges that arise in distinguishing age-related changes from (undetected) disease are evident.[6] Though challenging, this overlap also underlines the importance of understanding brain aging to ultimately improve our understanding of neurodegenerative disease.

I believe that by estimating brain aging trajectories, the brain aging process can be captured and inform us on the timing and sequence of changes in the brain. The definition of an aging trajectory in the context of this thesis is the average course of a marker with age. Using mixed models, both participants with a single measurement and those with multiple measurements, with unstructured time intervals, are used to estimate the average aging trajectory. This trajectory represents between-subject differences at different ages, rather than within-subject differences.

The aging trajectories are estimated based on longitudinal data from study participants without neurodegenerative diseases. It goes without saying that different definitions of 'normal' aging, based on the same data, could result in different trajectories. By using these trajectories to estimate brain aging patterns rather than focussing on the trajectories themselves, I believe that the variation in normal aging definitions will not retain us from better understanding the brain aging process. Furthermore, the descriptive nature of the work presented in this thesis allows for translation to different normal aging definitions.

With brain aging trajectories and brain aging patterns, we can subsequently investigate the effects of different determinants on brain aging. This could bring us one step closer towards understanding the sources of variability, and their implications. Furthermore, improving our understanding of the timing and sequence of changes in the brain with aging could provide a new perspective on the health-disease continuum and ultimately improve our understanding of disease.

Overall aim and outline of this thesis

The overall aim of this thesis is to gain insight in the brain aging patterns, to ultimately improve our understanding of the continuum of aging and age-related diseases.

The specific objectives of this thesis are:

1. Quantification of brain aging trajectories of structural imaging markers, cognitive function and motor function, and the brain aging patterns based on these markers.
2. Determining the association between risk factors and the brain aging trajectories.
3. Determining how brain aging can inform disease assessment and prediction.

Chapter 2, devoted to the first objective, focusses on the quantification of brain aging trajectories and the corresponding brain aging pattern, in a population free of overt neurological disease. In **Chapter 2.1** I estimated brain aging trajectories of

structural imaging markers and based on these markers I describe the average pattern of brain aging. In **Chapter 2.2** brain aging trajectories of cognitive and motor function were estimated and, similar to **Chapter 2.1**, I determined the average pattern of changes in cognitive and motor function with aging. In **Chapter 2.3** I assessed the relation between cortical gyrification and age and cognition in older adult. In **Chapter 2.4** the focus has shifted from the average brain aging pattern, to identifying potential different brain aging patterns present within the population. I used a data-driven disease progression model to identify different brain aging subtypes based on structural imaging markers. In **Chapter 2.5** I discuss the importance and potential of using longitudinal data for unravelling the brain aging process.

Chapter 3 focusses on determinants of brain aging in the general population and is devoted to objective 2. In **Chapter 3.1** I studied the relation between intracranial arteriosclerosis and brain aging trajectories of structural imaging markers. To deal with variating time between the measurement of intracranial calcification (using computed tomography), and the repeated structural imaging markers (using MRI), I determined calcification percentile curves to estimate the age-specific calcification burden. In **Chapter 3.2** I assessed the relation between cardiovascular health factors and APOE ε4 carriership and the brain aging trajectories of structural imaging markers. In **Chapter 3.3** I investigated the relation between hearing loss and cognitive decline. Hearing loss is considered as a promising modifiable risk factor for cognitive decline and dementia. Since both hearing loss and cognitive decline are highly related to aging, this chapter is focussed on investigating whether hearing loss accelerates cognitive decline with aging.

Chapter 4 focusses on how brain aging can inform disease assessment and prediction, and is devoted to objective 3. In **Chapter 4.1** I studied the impact of using normative brain volumetry derived from different reference population. I compared percentile curves of subcortical brain structure volumes between three different reference populations. Furthermore I evaluated the impact of these differences on single-subject diagnostic assessment in dementia. In **Chapter 4.2** I assessed the use of a data-driven disease progression model for the prediction of Alzheimer's disease in a population-based cohort. I used a recently developed data-driven disease progression model, called co-initialized Discriminative Event-Based Model (co-init DEBM), to construct APOE-specific disease timelines based on structural imaging markers in a case-control setting. Furthermore, I evaluated the generalizability from these disease timelines to a population-based setting. Finally, I investigated if progression along these disease timelines is predictive of Alzheimer's disease.

2

Quantifying brain aging trajectories

2.1

Trajectories of imaging markers in brain aging: The Rotterdam Study

EJ Vinke, M de Groot,V Venkatraghavan, S Klein,
WJ Niessen, MA Ikram, MW Vernooij

Neurobiology of Aging. 2018;71:32-40

With aging, the brain undergoes several structural changes. These changes reflect the normal aging process and are therefore not necessarily pathologic. In fact, better understanding of these normal changes is an important cornerstone to also disentangle pathologic changes. Several studies have investigated normal brain aging, both cross-sectional and longitudinal, and focused on a broad range of magnetic resonance imaging (MRI) markers. This study aims to comprise the different aspects in brain aging, by performing a comprehensive longitudinal assessment of brain aging, providing trajectories of volumetric (global and lobar; subcortical and cortical), microstructural, and focal (presence of microbleeds, lacunar or cortical infarcts) brain imaging markers in aging and the sequence in which these markers change in aging. Trajectories were calculated on 10,755 MRI scans that were acquired between 2005 and 2016 among 5286 persons aged 45 years and older from the population-based Rotterdam Study. The average number of MRI scans per participant was 2 scans (ranging from 1 to 4 scans), with a mean interval between MRI scans of 3.3 years (ranging from 0.2 to 9.5 years) and an average follow-up time of 5.2 years (ranging from 0.3 to 9.8 years). We found that trajectories of the different volumetric, microstructural, and focal markers show nonlinear curves, with accelerating change with advancing age. We found earlier acceleration of change in global and lobar volumetric and microstructural markers in men compared with women. For subcortical and cortical volumes, results show a mix of more linear and nonlinear trajectories, either increasing, decreasing, or stable over age for the subcortical and cortical volume and thickness. Differences between men and women are visible in several parcellations; however, the direction of these differences is mixed. The presence of focal markers show a nonlinear increase with age, with men having a higher probability for cortical or lacunar infarcts. The data presented in this study provide insight into the normal aging process in the brain, and its variability.

Introduction

The aging brain undergoes various structural changes, which can manifest themselves clinically in corresponding functional changes. Much research has been dedicated to understanding these brain changes because these do not only inform about healthy brain aging, but also provide a reference point against which pathologic changes can be contrasted.

The development of noninvasive imaging techniques has fueled research into the aging brain in healthy individuals. Since magnetic resonance imaging (MRI) was

first introduced in biomedical research in the 1980s, several pioneers performed small studies using this novel technique to assess macrostructural brain changes in aging.[7–12] After approximately one decade, large cross-sectional studies and population-based studies followed to inform about, for example, sex differences and brain changes in a large sample of healthy volunteers, instead of specific control subjects.[13–15] Simultaneous developments in MRI scanners and software increased the accuracy of structural (volumetric) measurements and enabled measuring microstructural (white matter organization) changes in aging.[16–22] In the last 15 years, more and more longitudinal studies have been performed to estimate the rate of brain changes in aging or investigating possible causes and effects of these changes.[23–35] Overall, these studies show that the vulnerability of the brain to aging is heterogeneous. Furthermore, some studies show sex differences in the effect of age on the imaging markers[31,33–35], whereas others do not[31,35].

Against the background, we aimed to comprise these different aspects in brain aging, by performing a comprehensive longitudinal assessment of brain aging in a middle- and old-aged population. We examined trajectories of volumetric, microstructural, and focal MRI markers in aging across a wide age range (45–95 years) in men and women based on a large prospective population-based cohort study with over 10,000 MRI scans. Furthermore, we analyzed the sequence in which MRI markers change in aging, so as to provide an overview of the concurrency of the changing imaging markers.

Methods and Materials

Study population

This study is embedded within the Rotterdam Study, an ongoing prospective population-based study designed to investigate causes and consequences of age-related diseases. The design of the Rotterdam Study has been described previously.[36] Since 2005, brain MRI was implemented in the core Rotterdam Study protocol, and participants are invited every 3–4 years for repeat imaging. In Figure 1, a flowchart of the inclusion of the MRI scans is shown. In this study, all available MRI scans from the Rotterdam Study that were acquired since August 2005 (date of installment of the MRI scanner in the research center) were included (n = 12,023 scans). Scans from participants with dementia or Parkinson's disease that were performed after clinical diagnosis were excluded (n = 110 MRI scans from 94 participants). Scans from participants with a symptomatic stroke that were performed after the event were excluded (n = 385 scans from 235 participants). Furthermore, MRI scans with incomplete acquisition, scans with artifacts hampering automated processing,

unreliable tissue segmentation (or unreliable intracranial volume segmentation in case for the focal marker analysis), and incomplete ratings of microbleeds, cortical, and lacunar infarcts were excluded (volumetric and microstructural markers analysis: n = 543 MRI scans, 4.7%; focal markers analysis: n = 42, 0.4%). Finally, all scans with MRI-defined cortical infarcts were excluded from volumetric and microstructural analysis (n = 230 MRI scans), but not for the focal marker analysis. In total, 10,755 MRI scans from 5286 participants were available for analysis of the volumetric and microstructural imaging markers, and 11,486 MRI scans from 5522 participants were available for analysis of focal markers.

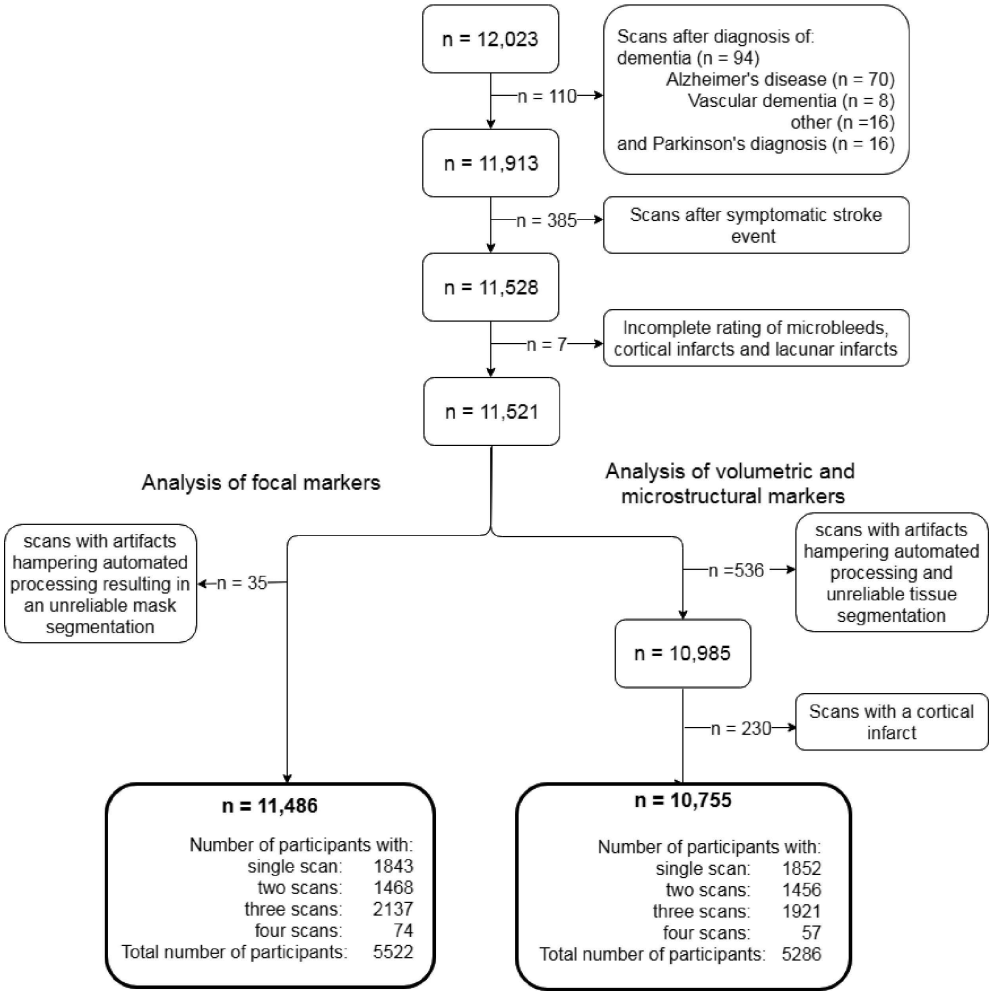


Figure 1. A flowchart of the inclusion of MRI scans for both the analysis of volumetric and microstructural markers and the analysis of focal markers is shown.

MRI acquisition and processing

Brain MRI scanning was performed in all participants during the entire study period on the same single 1.5-Tesla MRI scanner (GE Signa Excite; GE Healthcare, Milwaukee, USA), keeping hardware and software setup unchanged over the entire study period. The scan protocol and sequence details have been described elsewhere.[37]

For brain volumetry, T1-weighted (voxel size 0.49 × 0.49 × 1.6 mm3), proton density–weighted (voxel size 0.6 × 0.98 × 1.6 mm3), and the fluid-attenuated inversion recovery (FLAIR) (voxel size 0.78 × 1.12 × 2.5 mm3) scans were used for automated segmentation of supratentorial gray matter, white matter, cerebrospinal fluid (CSF), and white matter lesions.[38,39] All scans were transformed to the high-resolution data set (256 × 256 × 128) using tri-linear interpolation. Automated processing tools from the Brain Imaging Center, Montreal Neurological Institute and McGill University (www.bic.mni.mcgill.ca) were used to coregister the MRI data (based on mutual information) and subsequently normalize the intensities for each feature image volume using N3.[40] All segmentations were visually inspected and manually corrected if needed. Total brain volume was the sum of gray matter, normal-appearing white matter, and white matter lesion volume. Supratentorial intracranial volume was estimated by summing gray and white matter (consisting of the sum of normal-appearing white matter and white matter lesion volume) and CSF volumes.[38] For measurement of lobar volumes, an atlas was created in which the lobes were labeled according to a slightly modified version of the segmentation protocol, as described by Bokde et al.[41,42] Subsequently, nonrigid transformation was used to transform this atlas to each brain as described previously[42], to obtain volume for each lobe. Furthermore, T1-weighted MR images were processed using FreeSurfer[43] (version 5.1) to obtain cortical parcellations (n = 33) and subcortical structure volume of the hippocampus, putamen, amygdala, pallidum, and caudate nucleus.

For microstructural measures, diffusion tensor imaging (voxel size 3.3 × 2.2 × 3.5 mm3) was used. A single shot, diffusion-weighted spin echo echo-planar imaging sequence was performed with maximum b value of 1000 s/mm2 in 25 noncollinear directions; 3 b0 volumes were acquired without diffusion weighting. Diffusion data were preprocessed using a standardized processing pipeline[196], yielding (in combination with the tissue segmentation) global mean fractional anisotropy and mean diffusivity (MD) in the normal-appearing white matter (voxels of white matter excluding white matter lesions).

For (dichotomous) focal lesions, infarcts showing involvement of cortical gray matter were classified as cortical infarcts. Lacunar infarcts were defined by focal loss of noncortical tissue (size ≥ 3 and <15 mm) with signal intensity similar to CSF on all sequences, and when located supratentorially with a hyperintense rim on the FLAIR sequence.[44] To differentiate lacunar infarcts from dilated perivascular spaces, symmetry of the lesions, sharp demarcation, and absence of a hyperintense rim on the FLAIR sequence supported presence of a dilated perivascular space.[45] Cerebral microbleeds were rated on a 3-dimensional, T2*-weighted gradient-recalled echo MRI scan (voxel size 0.78 × 1.12 × 1.6 mm3) as focal areas of very low signal intensity.[37] Note that although white matter lesions can be considered focal markers as well, we measured these continuously in the present study, and therefore for the purpose of visualization and comparison with other volumetric markers, we categorized white matter lesion volume as a volumetric marker.

Statistical analysis

Trajectories of global and lobar volumetric, cortical thickness, cortical volume and subcortical structure volume MRI markers, and microstructural measures were assessed using linear mixed models with random intercepts and slopes. Cortical and subcortical structure volumes and cortical thickness was the average of the left and right hemisphere. The global volumetric, microstructural, and subcortical structure volume markers that were modeled were white matter volume, white matter lesion volume, normal-appearing white matter volume, gray matter volume, CSF volume, total brain volume, hippocampus, putamen, amygdala, pallidum, and caudate nucleus volume, global fractional anisotropy, and global MD. The linear mixed models were performed using the "lme" function from the R-package "nlme".[46] In each model, age of the participant at each measurement was used as the time variable. To account for possible nonlinear trajectories, exploratory analysis was performed to assess whether splines of age (with increasing degrees of freedom) improved the model compared with the linear age term. As a result from these analyses, splines of age with 1 knot were used in all models. Other covariates in the model were intracranial volume and sex. The interaction of sex and the spline coefficients of age was integrated into the model to test for possible slope differences between men and women. White matter lesion volume was natural log-transformed to account for the skewness of the measure distribution.

Next, trajectories of global and lobar volumetric, hippocampus volume, and microstructural MRI markers were Z-transformed to compare the temporal course of MRI marker change in aging. We selected these MRI markers for this analysis as these are currently the most extensively studied in aging and neurodegeneration, with relevant clinical interpretation. The transformation from absolute values to Z-

scores was performed by subtracting the predicted curve value at age 45 years from the predicted curve at a certain age and dividing it by the standard deviation of the residuals of the linear mixed model. In case of an increasing trajectory (e.g., in white matter lesion volume), the Z-score trajectory was multiplied with −1 to orient all trajectories to the same direction as those of markers that decrease with age. The sequence with which imaging markers change in aging after age 45 years was determined by calculating the age at which a −2SD change compared with the mean population value at age 45 years was reached (Z-score of −2). We assessed the sequence of change for men and women separately.

For focal lesions, the probability of having one or more microbleeds, cortical infarcts, or lacunar infarcts was assessed using generalized estimating equations (GEEs). The GEE was performed using the "geeglm" function from the R-package "geepack".[47] In the GEE, natural cubic splines of age with 1 knot were used as the time variable. The covariates in the model were the same as in the linear mixed model, namely intracranial volume, gender, and the interaction of the splines of age and gender.

As sensitivity analysis, all analyses were performed after exclusion of scans before dementia and Parkinson's diagnosis and scans before a stroke event to investigate the effect of including preclinical scans. Furthermore, all analysis were performed after exclusion of participants with only a single MRI scan, to investigate the effect of possible population differences between participants with a single scan and those with multiple scans.

Results

Characteristics at first MRI scan of the participants included for studying the volumetric and microstructural markers and the focal markers are presented in Table 1. In Supplemental Table 1, the presence of several cardiovascular risk factors in our study population is shown. For studying the volumetric and microstructural markers, 5286 participants were included. The mean age at first scan was 64.4 years (range 45.7–97.9 years), and 2962 (56.0%) participants were women. The mean intracranial volume was 1138.2 mL (range 813.6–1699.4 mL). Of 5286 participants, 1852 participants had a single brain MRI scan, 1456 participants had 2 MRI scans, 1921 participants had 3 MRI scans, and 57 participants had 4 MRI scans available for analysis. The average follow-up time was 5.2 years (range 0.3–9.8). The mean scan interval between the first and second, second and third, and third and fourth MRI scan was, respectively, 4.0 (0.66), 1.9 (0.71), and 4.4 years (0.18). For studying the

focal markers, 5522 participants were included. Of these 5522 participants, 1055, 375, and 144 participants had, respectively, one or more microbleeds, lacunar infarcts, and cortical infarcts at their first scan.

Trajectories of global volumetric and microstructural markers

Figure 2 depicts the estimated trajectories of global volumetric, microstructural, and subcortical volume markers, for men and women separately. Furthermore, Supplemental Table 2 shows the trajectory values corresponding to the trajectories at age 45 and 95 years. White matter volume, normal-appearing white matter volume, gray matter volume, total brain volume, global fractional anisotropy, hippocampus volume, amygdala volume, and pallidum volume all showed a nonlinear decrease with age, whereas white matter lesion volume, CSF volume, and global MD increased (nonlinearly) with age. After approximately age 50–55 years, most trajectories showed an accelerated change, which was most pronounced for total brain volume, CSF volume, hippocampus volume, and MD (Figure 2). Gray matter volume and putamen volume showed a more linear decrease with age compared with the other imaging markers. Caudate nucleus volume shows a more U-shaped curve with an increasing volume at increasing age, with the deflection point around age 65 years.

In Figure 3A and B, the trajectories of the MRI markers are shown in Z-scores for men and women separately. In Supplemental Table 2, the corresponding Z-values at age 45 and 95 years are given. For both sexes, CSF and total brain volume reached the largest total change in Z-score (men −6.1; women −4.8), compared with the other MRI markers. There was significant interaction between age and sex for all markers, except for global FA, global MD, and amygdala volume (p-values of the age and sex interaction is given in Supplemental Table 2). In general, the trajectories show an earlier acceleration of changing markers in men compared with women (Figure 3C).

Lobe-specific trajectories of volumetric markers and cortical parcellations

The lobe-specific trajectories of volumetric markers for men and women are shown in Supplemental Figure 2 (absolute values) and Supplemental Figure 3 (Z-scores). After taking into account differences in intracranial volume, the frontal lobe showed the largest relative change in each of the MRI markers, except for gray matter volume in men, where temporal gray matter showed the largest relative change. The occipital lobe showed the smallest change in each of the MRI markers, except for gray matter volume and for total lobar volume in women, where the parietal lobe showed the smallest relative change.

Table 1. Characteristics of the study population.

Characteristic	Analysis of volumetric and microstructural markers (N = 5,286)	Analysis of focal markers (N = 5,522)
Age at first scan, years	64.4 (10.7)	64.6 (10.8)
Age in men, years	64.2 (10.5)	64.4 (10.6)
Age in women, years	64.6 (10.8)	64.8 (10.9)
Sex, women	2962 (56.0)	3079 (55.8)
Intra-cranial volume, ml	1138.2 (115.7)	1138.4 (116.5)
Scan interval ,years	3.3 (1.2)	3.3 (1.2)
Follow-up time, years	5.2 (1.1)	5.2 (1.0)
Availability of MRI scans with acceptable segmentation		
Number of participants with a single MRI scan	1852 (35.0)	1843 (33.4)
Number of participants with two MRI scans	1456 (27.5)	1468 (26.6)
Number of participants with three MRI scans	1921 (36.3)	2137 (38.7)
Number of participants with four MRI scans	57 (1.1)	74 (1.3)
Number of available MRI scans with acceptable segmentation	10755	11486
Baseline global FA[a]	0.34 (0.015)	
Baseline global MD, 10^{-3} mm^2/sec[a]	0.74 (0.029)	
Baseline total white matter volume, ml	407.3 (59.8)	
Baseline normal appearing white matter volume, ml	401.3 (61.3)	
Baseline white matter lesion volume[b], ml	2.9 (1.6-6.0)	
Baseline grey matter volume, ml	529.7 (54.9)	
Baseline cerebrospinal fluid volume, ml	198.9 (54.9)	
Baseline total brain volume, ml	937.0 (100.1)	
Baseline hippocampus volume[a], ml	3.9 (0.52)	
Baseline putamen volume[a], ml	4.6 (0.61)	
Baseline amygdala volume[a], ml	1.4 (0.19)	
Baseline pallidum volume[a], ml	1.5 (0.23)	
Baseline caudate nucleus volume[a], ml	3.4 (0.52)	
Microbleed(s) at baseline scan	1003 (19.0)	1055 (19.1)
Lacunar infarct(s) at baseline scan	336 (6.4)	375 (6.8)
Cortical infarct(s) at baseline scan	*Not applicable*	144 (2.6)

Continuous variables are presented as means (standard deviations) and categorical variables as number (percentages).

Abbreviations: N: number of participants, sec: seconds, FA: fractional anisotropy, MD: mean diffusivity.

[a]Data on the subcortical structure volumes was the average of the left and right volume. Subcortical structure volume was missing in 23 participants due to failed FreeSurfer segmentation. Data on global FA and global MD in the normal-appearing white matter were missing in 295 participants, due to failed segmentation of the diffusion tensor images.

[b]White matter lesion volume are presented as median (inter-quartile range).

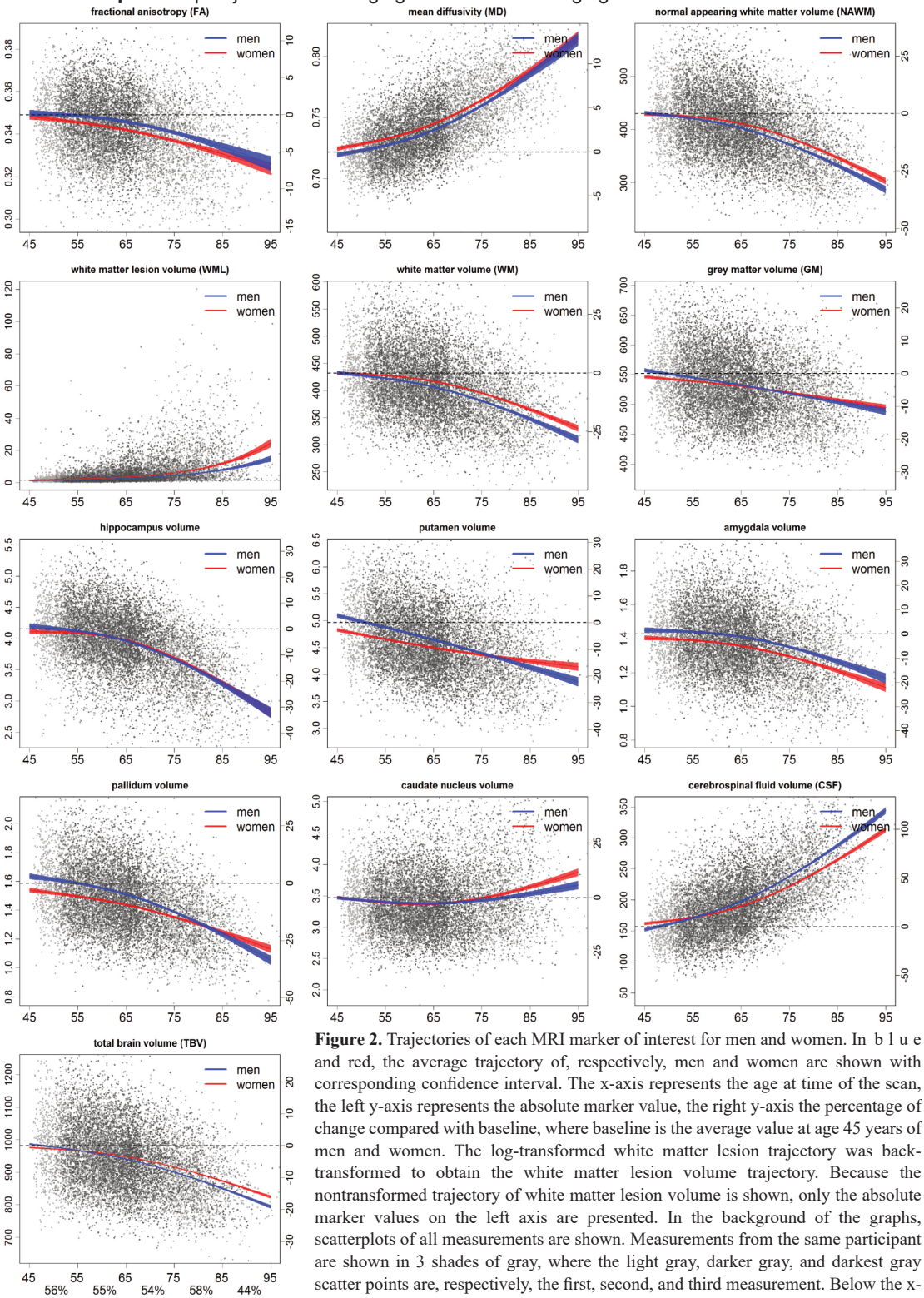


Figure 2. Trajectories of each MRI marker of interest for men and women. In blue and red, the average trajectory of, respectively, men and women are shown with corresponding confidence interval. The x-axis represents the age at time of the scan, the left y-axis represents the absolute marker value, the right y-axis the percentage of change compared with baseline, where baseline is the average value at age 45 years of men and women. The log-transformed white matter lesion trajectory was back-transformed to obtain the white matter lesion volume trajectory. Because the nontransformed trajectory of white matter lesion volume is shown, only the absolute marker values on the left axis are presented. In the background of the graphs, scatterplots of all measurements are shown. Measurements from the same participant are shown in 3 shades of gray, where the light gray, darker gray, and darkest gray scatter points are, respectively, the first, second, and third measurement. Below the x-axis in the graph of total brain volume, the proportion scans from women are shown for each 10-year age bin.

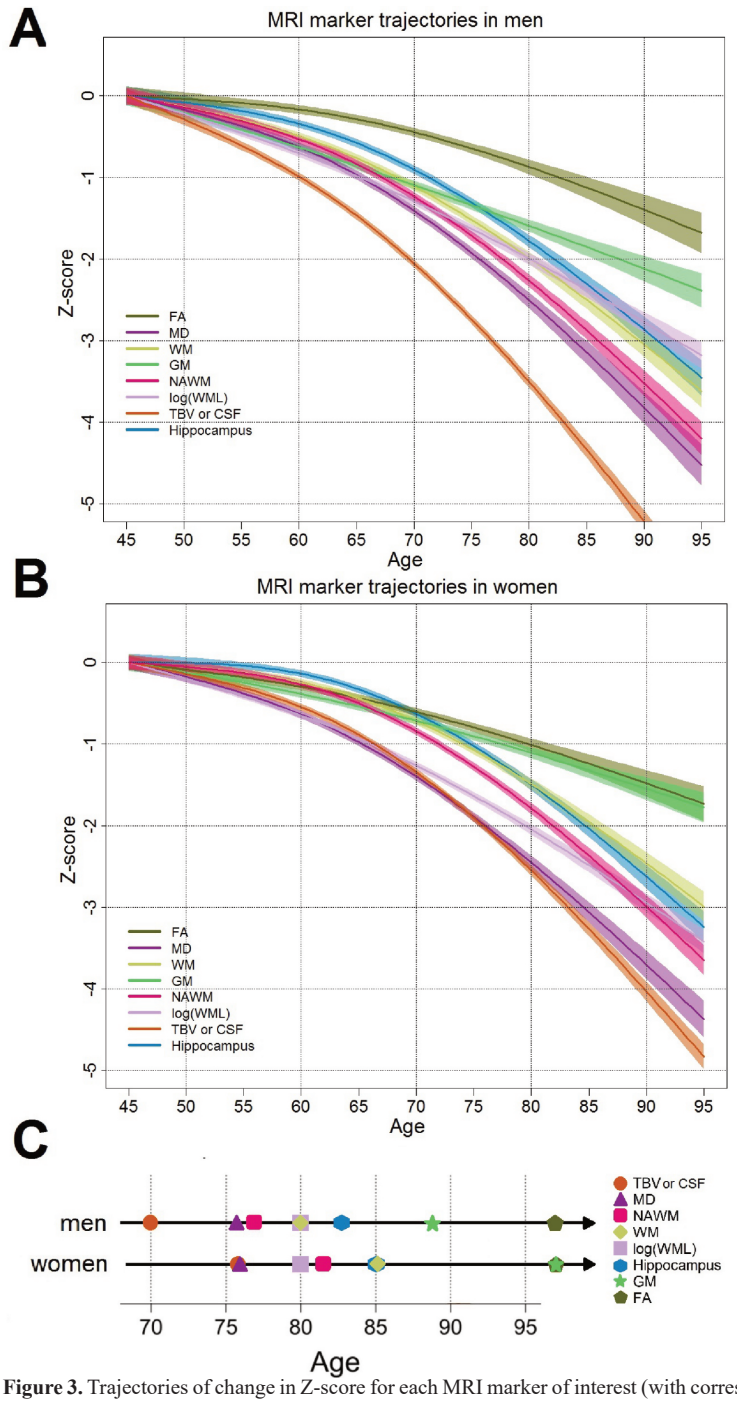


Figure 3. Trajectories of change in Z-score for each MRI marker of interest (with corresponding confidence interval) for (A) men and (B) women. (C) Representation of the age at which Z-score of each MRI marker reached a 2SD change compared with the value at age 45 years. Trajectories are shown for global FA, global MD, WM, GM volume, NAWM, natural log transform of white matter lesion volume (log [WML]), TBV, CSF volume, and hippocampus volume (Hippocampus). The trajectories of MRI markers which increase with age were multiplied with −1 to orient all lines in the same direction. The trajectory of CSF and total brain volume overlap; therefore, this trajectory represents both TBV and CSF multiplied with −1. In men the age at which a 2SD change occurred is the same for white matter and the natural log transform of white matter lesion; therefore, at age 80 years the symbols for these two markers are overlapping. The same holds for fractional anisotropy and gray matter volume in women.

For all volumetric markers of the lobes, there was a significant interaction between age and sex, except for the log-transformed white matter lesion volume in the frontal, temporal, and occipital lobe, and gray matter volume in the occipital lobe. Similar to the trajectories of the MRI markers in the whole brain, the lobe-specific biomarkers in men showed earlier changing markers compared with women. In addition, the change expressed in Z-score was larger in men compared with women, except for white matter lesion volume.

The trajectories of the volume and thickness of the cortical parcellations in men and women are shown in Supplemental Figures 3–6. These figures show a mix of more linear and nonlinear trajectories, either increasing, decreasing, or stable over age for the volume and thickness of cortical parcellations. Differences between men and women are visible in several parcellations; however, the direction of these differences is mixed, where in some parcellations, there seems to be only an intercept difference, and in other parcellations, men show an earlier decrease in volume or thickness and vice versa. Overall, the amount of change in cortical volume was different across and within lobes, were certain regions in the frontal, temporal, and parietal lobe decreased ~25% at age 95 years compared with age 45 years, whereas other regions showed less vulnerability to age.

Sequence of changing volumetric and microstructural MRI markers

The sequence in which volumetric and microstructural MRI markers reach a 2SD change after age 45 years is shown in Figure 3C. For both men and women, total brain volume and global MD were the first 2 markers to change after age 45 years, although in men, changes started 6 years before women (70 years vs. 76 years). Other differences between men and women are primarily a later change in women of normal-appearing white matter volume and normal-appearing white matter volume (Figure 3C). Global gray matter volume change occurs relatively late in both sexes, with the largest age difference between men and women (89 in men, 97 in women). The last marker in the sequence was global fractional anisotropy, occurring at approximately the same age in both sexes (age 97 years). The sequence of reaching a 2SD change after age 45 years of total lobe volume is shown in Supplemental Figure 2A. The first lobe that changes after age 45 years is the frontal lobe, which is then followed by the temporal lobe. There is a difference in sequence between sexes, where in men, after the temporal lobe, the parietal lobe changes and finally the occipital lobe, whereas in women, this is the other way around. Overall, the changes in total lobe volume of the frontal, temporal, parietal, and occipital lobe occurs, respectively, 6, 7, 11, and 4 years later in women.

Focal imaging markers in aging

Figure 4 shows the probability curves in aging for microbleeds, cortical infarcts, and lacunar infarcts on MRI, for men and women separately. The probability of having one or more microbleeds ranged from 4.7% (age 45 years) to 54.8% (age 95 years) and was overall higher than lacunar infarcts (0.9%–23.8%, respectively) and cortical infarcts (1.0%–15.2%, respectively). Men had higher prevalence of lacunar infarcts and cortical infarcts than women, respectively, for the ages 64.7 to 83.7 years, and 54.9 to 90.7 years. The interaction between age and sex was not significant in any of the focal markers.

Sensitivity analyses

We performed a sensitivity analysis in which scans before a dementia or Parkinson's diagnosis, or scans before a stroke event were excluded (n = 488 scans). Furthermore, we performed a sensitivity analysis in which only participants with multiple scans were included. Both sensitivity analyses showed similar results compared with the original results.

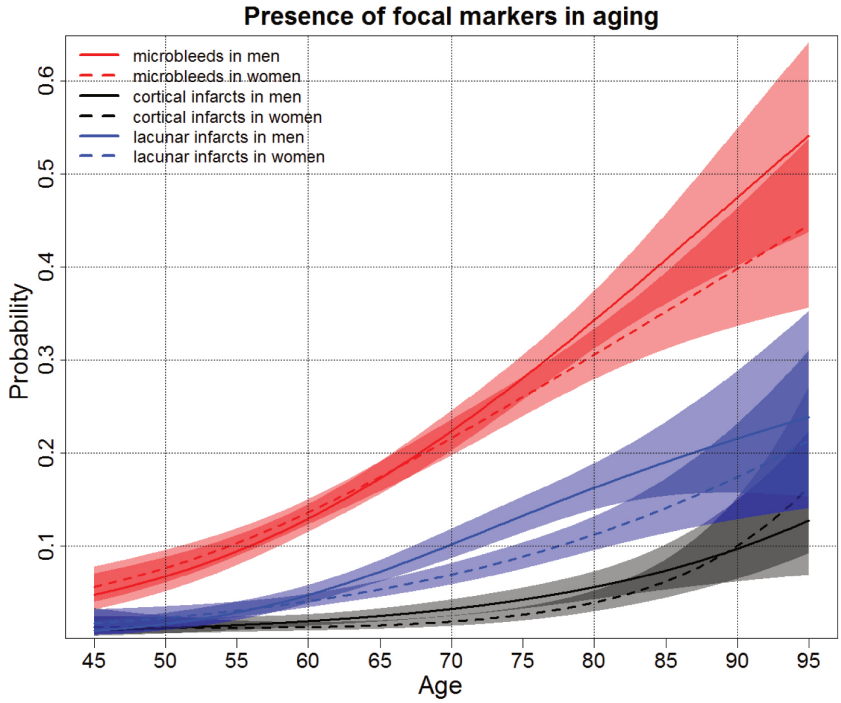


Figure 4. Probability curves of the presence of the focal markers: one or more microbleeds, cortical infarct, or lacunar infarct in men and women in aging. The solid and dotted lines represent the trajectories of, respectively, men and women with corresponding 95% confidence interval. The red, black, and blue prediction lines represent the probability of presence of respectively one or more microbleeds, cortical infarct, and lacunar infarct.

Discussion

In this study, we present a comprehensive longitudinal assessment of brain aging, providing an overview of the concurrency of changing imaging markers. We show trajectories of volumetric, microstructural, and focal imaging markers in a large aging population, based on longitudinal MRI data. The trajectories of the different global, cortical, subcortical, and lobar MRI markers follow a nonlinear curve, with accelerating change with advancing age. Regarding temporal patterns, the change in MRI markers generally occurs earlier in men than in women. Among focal lesions, microbleeds show the highest prevalence and steepest increase across the age range, up to 54.8% in age 95 years. Overall, men tend to have higher prevalence of focal lesions (microbleeds, lacunes, and cortical infarcts) compared with women.

A major strength of this study is its longitudinal design, which increases sensitivity to detect rates of change compared with a more often used cross-sectional design. In combination with the large sample size from a population-based setting, this increases the generalizability of our results. Furthermore, because of the availability of many different MRI markers within this large longitudinal sample, we were able to simultaneously analyze and compare the aging effect of these markers. However, some limitations also need to be considered. Owing to a relatively short time between the first and last scan of participants, and relatively sparse proportion of older participants, the estimated trajectories may not be representing the longitudinal effect at the older ages reliably. Furthermore, although scans from participants after diagnosis of Alzheimer's disease, Parkinson's disease, and stroke were excluded, it is likely that a proportion of the participants are prodromal, which could influence the trajectories. However, our sensitivity analysis in which we also excluded scans before diagnosis showed similar results. Therefore, we believe that the effect of prodromal participants on these trajectories is minimal. Another limitation of this study is the possible selection bias, due to differences in people participating in the scan study and participants that refuse.[48] Furthermore, the same could hold for participants with only a single scan compared with participants with multiple scans. In 35.0% of the participants, there was only one MRI scan available for analysis. An important advantage of including persons with both single and multiple brain scans is that we include all available information and avoid the risk of including relatively healthier survivors with only longitudinal information. However, in this study, results remained similar after excluding participants with only a single scan.

The trajectories of whole brain, cortical, subcortical, and lobar volumetric markers we assessed are largely in accordance with previous literature.[16,26,33,49,50] Our study confirms previous findings that showed that white matter volume changes in aging are nonlinear, with a more rapid change with advancing age, whereas gray matter

shows a smaller and more linear decrease with advancing age.[30,35,49,51,52] Previous studies have shown different curves for subcortical structures, for example, hippocampus volume atrophy accelerates at increasing age, whereas the caudate nucleus follows a more stable curve with increasing age;[16,30] our study also shows a nonlinear decrease in volume for hippocampus, amygdala, and pallidum in aging and a more linear volume decrease in putamen volume and a U-shaped curve for the caudate nucleus volume.

Our study adds to existing literature in showing the longitudinal trajectories of microstructural white matter changes on top of these macrostructural changes. As understanding normal aging is essential to better understand or detect abnormal aging, we examined the sequence with which these imaging markers change with age and found among volumetric markers total brain volume to change first, which is likely due to the fact that this reflects a summation of changes in other tissues as well. This was followed by total white matter, with total gray matter being one of the last markers to change. An important limitation of our method determining the sequence is that it depends on the amount of change relative to the variation, meaning that differences in variation and measurement error directly influences the sequence. In accordance with literature, lobar trajectories of volumetric imaging markers and cortical volume and thickness in this study confirm regional differences in the amount of atrophy within and across lobes.[6,50,53] Trajectories of white matter and total lobe volume show that the frontal lobe is most affected, which has also been described in other studies.[24,35,49,54,55] Furthermore, our study shows that within frontal, temporal, and parietal lobe, there are certain cortical regions highly affected by aging, reaching up to ~25% decrease in volume at age 95 years compared with age 45 years. In addition, we determined the sequence with which the lobes change with age, showing that total frontal lobe volume changes first, followed by the temporal lobe, parietal lobe, and lastly the occipital lobe. Several hypotheses have been proposed to explain this selective vulnerability of the brain regions, including the "retrogenesis hypothesis" which states that late maturing regions are most vulnerable to aging and the hypothesis of an anterior-to-posterior gradient of age vulnerability. However, the underlying biological mechanism remains unknown.[6,56]

Although several studies have looked at sex differences in volumetric imaging markers in the aging brain, they have not yielded consistent results. Some studies showed that women have overall a proportionally larger gray matter volume and less white matter volume than men,[57,58] whereas others showed the opposite[14,52]. A recent cross-sectional study showed absolute volume differences between men and women for several markers, but the shape of the trajectories across the life span was equal between men and women.[33] Discrepancies in these findings could most likely be explained by cross-sectional study design, small sample sizes and limited spread of

the MRI data over the total age range. In this large longitudinal study, we show sex differences in the trajectories of all volumetric MRI markers in normal aging, after correcting for head size differences. We show an earlier acceleration and a larger amount of change in men compared with women for both whole brain, subcortical, cortical, and lobar volumes, but also for focal lesions. These differences are also reflected in a difference in the sequence in which imaging markers change between men and women, where especially total brain volume, white matter, normal-appearing white matter, and gray matter volume changes in women appear later in the sequence than in men. These sex-specific differences are important to take into account when normative reference values on a lobar level, for example, in a memory clinic, would be applied in a clinical setting to assess pathology in individual patients.

Less is known on global microstructural changes in aging, but the microstructural MRI trajectories we describe are in accordance with published literature, in which decreased global fractional anisotropy and increased global MD with age have been suggested to reflect a reduced microstructural integrity.[59] Our results suggest global MD to be a more sensitive marker of reduced microstructural organization in aging, as it showed an earlier and more accelerated change in comparison with global fractional anisotropy. In fact, we found that global MD was the second marker to change (after total brain volume) in the sequence of changing MRI markers, before change in volumetric white matter markers, indicating that microstructural changes precede volumetric white matter changes. This is in agreement with a previous study by our group showing that microstructural changes in white matter appear before development of white matter lesions.[60]

Similar to volumetric and microstructural MRI markers, focal markers also show a nonlinear relationship with age. The probability of having one or more microbleeds was higher than having ischemic lesions (lacunes and cortical infarcts) at all ages. Over practically the entire age range, the probability of having one or more lacunar infarcts (up to 23.8%) was higher than for cortical infarcts (maximum 15.2%). This corresponds to a previous study from our own group, which showed in a cross-sectional sample of the same study population that prevalence of microbleeds is higher than infarcts, and that the prevalence of lacunar infarcts is higher than cortical infarcts.[61] Age trajectories for these focal lesions, especially derived from longitudinal data, were lacking so far, and the present study provides important information that may also be useful as a background in a clinical setting. The overall prevalence of cortical and lacunar infarcts is comparable with the prevalence of silent infarcts in the different age categories described in a previous cross-sectional study in Rotterdam Study participants (not included in the present sample).[62]

Between the age range 65–85 years, the probability of having one or more lacunar and cortical infarcts was higher in men. This is in contrast to what previously has been shown in a study within other subjects from the same community-dwelling population, where the prevalence of silent infarcts was higher in women than in men,[22] whereas in other studies no significant difference between the prevalence of infarcts were found between men and women.[62] A possible reason for the differences of our results with other reports is that we made no distinction between silent and symptomatic infarcts, while it has been shown that the prevalence of symptomatic infarcts is higher in men.[63] Another explanation of our contradictory findings could be the larger sample size in combination with a longitudinal design, which may be more sensitive to find gender differences.

In this study, we focused on aggregated measures derived from the images; however, data-driven methods using the complete scan information, such as machine learning, could provide new insights in the effect of aging. Another interesting next step is to assess the effect of different determinants on the aging trajectories.

Overall, the trajectories of imaging markers in brain aging that we describe are essential background information for studies into age-related neurological diseases, or for clinical translation, for example, use of reference values. Especially in studies looking at differences between age-related pathology and normal aging, it is important to take into account the nonlinear age effects we found for all markers, as well as the interaction of age and sex for several markers.

2.2

Trajectories of cognitive and motor function between ages 45 and 90 years: a population-based study

S Licher*, KD van der Willik*, EJ Vinke, MJ Knol, SKL Darweesh, JN van der Geest, SB Schagen, MK Ikram, AI Luik, MA Ikram

The Journals of gerontonlogy: Series A. 2021; 76(2):297:306

The objective of this study was to establish trajectories of cognitive and motor function, and to determine the sequence of change across individual tests in community-dwelling individuals aged 45–90 years. Between 1997 and 2016, we repeatedly assessed cognitive function with 5 tests in 9514 participants aged 45–90 years from the population-based Rotterdam Study. Between 1999 and 2016, we measured motor function with 3 tests in 8297 participants. All participants were free from dementia, stroke, and parkinsonism. We assessed overall and education-specific cognitive and motor trajectories using linear mixed models with age as time scale. Next, we determined the sequence of change across individual tests. The number of assessments per participant ranged between 1 and 6 (mean interval, years [SD]: 5.1 [1.4]) for cognitive function, and 1 and 4 (5.4 [1.4]) for motor function. Cognitive and motor trajectories declined linearly between ages 45 and 65 years, followed by steeper declines after ages 65–70 years. Lower educated participants had lower cognitive function at age 45 years (baseline), and declined faster on most cognitive, but not on motor tests than higher educated participants. Up to a 25-year age difference between the fastest and slowest declining test scores was observed.On a population-level, cognitive and motor function decline similarly. Compared to higher educated individuals, lower educated individuals had lower cognitive function at baseline, and a faster rate of decline thereafter. These educational-effects were not seen for motor function. These findings benefit the understanding of the natural course of cognitive and motor function during aging, and highlight the role of education in the preservation of cognitive but not motor function.

Introduction

Understanding the natural course of cognitive and motor function during brain aging is pivotal to determine deviations in function that may signal early stages of clinical neurodegenerative diseases.[64,65] Decline in both cognitive and motor function has been associated with an increased risk of dementia, Parkinson's disease, and stroke.[64–66] In addition, we recently showed that individuals in whom decline in motor function precedes decline in cognitive function are at an increased risk of dementia.[66] Numerous studies have quantified the temporal relation of cognitive and motor function with advancing age[67–80], yet little is known about the sequence of individual cognitive and motor tests in a population free from neurodegenerative diseases and stroke.

Comparing trajectories of cognitive and motor tests in the general population reveals whether decline in motor function precedes decline in cognitive function. In addition, it identifies the specific individual tests that have the earliest signs of

decline. Such findings could inform clinicians about which cognitive and motor tests are most sensitive to detect change in cognitive or motor function. These trajectories can also be used to signal vulnerable patient groups that deviate from their expected course based on several key characteristics, such as age, sex, educational level, or genes. These characteristics significantly influence cognitive function and the rate of cognitive decline, but their effects on motor function beyond gait speed are less understood.[81,82]

Alike changes in brain structure, we hypothesize that change in cognitive and motor function accelerates with advancing age.[83] To model this nonlinear change, we present trajectories of cognitive and motor function. In addition, we assess the effects of key determinants of cognitive and motor function, namely age, sex, education, and apolipoprotein E (APOE) genotype on these trajectories. Finally, we determine the sequence of change of individual cognitive and motor function tests.

Materials and methods

Study Design

This study was embedded within the Rotterdam Study, a prospective population-based cohort designed to study the occurrence and determinants of age-related diseases in the general population.[36] In 1989, all inhabitants aged 55 years and older from Ommoord, a well-defined district in Rotterdam, the Netherlands received an invitation to participate. This initial cohort comprised 7983 participants. In 2000, 3011 participants who had become 55 years of age or moved into the study district since the start of the study were additionally included in the cohort. In 2006, a further extension of the cohort was initiated in which 3932 participants aged 45 years and older participated. In total, the Rotterdam Study comprises 14 926 participants aged 45 years and older. The overall response rate across all 3 recruitment waves was 72%.

Standard Protocol Approvals, Registrations, and Patient Consents

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

Study Population

Of a total of 14 926 participants, we excluded those with a history of dementia (n = 907), stroke (n = 846), Parkinson's disease (n = 300), or parkinsonism (n = 20) at time of their first cognitive or motor assessment. Next, we excluded participants with insufficient data to determine whether they had a history of one or multiple of these diseases (n = 1800). Baseline and follow-up ascertainment methods for dementia, stroke, Parkinson's disease, and parkinsonism have previously been described in detail.[84] In addition, 5 participants were excluded because they did not provide informed consent to access medical records and hospital discharge letters during follow-up. From the remaining 11 048 participants, 1494 participants were excluded because they did not have data available on any cognitive or motor test. Finally, we excluded assessments from participants after they had reached age 90 years in order to minimize the influence of leverage points on the trajectories of cognitive and motor function. This resulted in an additional exclusion of 33 participants who did not have any cognitive or motor function assessment at all before the age of 90 years, leaving 9521 participants with at least one cognitive or motor assessment. During follow-up, we excluded assessments of participants after the age of 90 years (n = 1266) and of participants after a dementia, stroke, or Parkinson's disease diagnosis (n = 3175). All of the included participants were thus free from neurodegenerative diseases and stroke at time of their test assessments. In total, 155 347 cognitive function assessments from 9514 participants and 62 545 motor function assessments from 8297 participants were available for analyses.

Assessment of Cognitive Function

Between 1997 and 2016, participants underwent cognitive assessments at the research center using a neuropsychological test battery every 3–5 years.[36,73] This battery included the Word Fluency Test[85], Letter-Digit Substitution Test[86], and Stroop Test (Reading, Naming, and Interference subtask)[87]. In 2002, the 15-Word Learning Test (Immediate recall, Delayed recall, and Recognition) was added to the test protocol[88]. This test protocol was further expanded with the Design Organization Test in 2006.[89] Assessments of these cognitive tests have previously been validated and have a reasonable to good test–retest reliability.[90–93]

Word Fluency Test

In the Word Fluency Test, participants were asked to mention as many animals as possible within 60 seconds, thereby measuring semantic fluency.[85] The total number of correct answers was used as test score, with a maximum score of 30 in our study protocol.

Letter-Digit Substitution Test

The Letter-Digit Substitution Test is a modified version of the Symbol Digit Modalities Test for which participants were asked to write down as many numbers underneath the corresponding letters as possible in 60 seconds, following a key that shows correct combinations.[86] This test captures both information processing speed and aspects of executive function. The total number of correct answers was used as test score with a maximum attainable score of 125.

Stroop Test

The Stroop Test consists of 3 different subtasks, that is, Reading, Naming, and Interference.[87] In the Stroop Reading subtask, participants were asked to read the printed color names. For the Stroop Naming subtask, participants were asked to name the printed color blocks. In the Interference subtask, participants were asked to name the ink color of color names printed in incongruous ink colors (information processing on an interference subtask). The time taken to complete the subtask was used as the outcome for each subtask separately and was adjusted for failures, that is, total time plus for each failure the total time divided by the number of items, multiplied with 1.5.[94] Thus, a higher score indicates a worse performance. The Stroop Test assesses information processing speed and executive function.

Word Learning Test

The Word Learning Test comprises 3 subtasks: Immediate recall, Delayed recall, and Recognition.[88] For Immediate recall, participants were 3 times visually presented with a sequence of 15 words and were subsequently asked to recall as many of these words as possible, measuring verbal learning. Free Delayed recall was tested approximately 10 minutes after visual presentation, evaluating retrieval from verbal memory. Recognition was tested by visually presenting the participants a sequence of 45 words, followed by correctly recognizing the 15 words presented during the Immediate recall while mixed with 30 other words. Outcome variables were the mean number of words of 3 trials immediately recalled (as a summary score for Immediate recall), after the delay of 10 minutes (as a score for free Delayed recall), and the mean number of correctly recognized words during the recognition trial (as a score for Recognition), with a maximum score of 15 per subtask.

Design Organization Test

The Design Organization Test consists of square black-and-white grids with visual patterns, of which participants were asked to reproduce as many designs as possible in 2 minutes using a numerical code key. It measures visuospatial abilities and is

based on and highly correlated to WAIS-III block design[89], but is less dependent on motor skills. Test score on the Design Organization Test has a range from 0 to 56 points for each individual, with higher scores indicating better performance.

Assessment of Motor Function

Participants repeatedly underwent motor tests every 3–5 years at the research center between 1999 and 2016. This motor test battery included 2 tests to assess fine motor function and a quantitative gait assessment to assess gross motor function. From 1999 onwards, the Purdue Pegboard Test was implemented into the study protocol to assess manual dexterity. Assessment of fine motor function was further expanded in 2008 with the implementation of the Spiral Archimedes Test to assess manual precision. In 2009, a quantitative gait assessment using an electronic walkway at the research center was implemented in the core study protocol.

Purdue Pegboard Test

For the Purdue Pegboard Test, participants were asked to place as many as possible cylindrical metal pegs into one of the 25 holes in a pegboard in 30 seconds in 3 separate trials, using their left hand only, right hand only, and both hands simultaneously, measuring fine motor function.[95] The test–retest reliability of assessments has been established previously. The outcome variable was the sum score of Purdue Pegboard Test score of these 3 trials, with a maximum of 75 points.

Archimedes Spiral Test

The Archimedes Spiral Test measures fine motor function by requiring participants to trace a picture of a spiral template that was printed on paper attached to an electronic drawing board (WACOM Graphire Wireless Pen Tablet, model CTE-630BT).[74] Participants were instructed to trace the spiral as accurately and as fast as possible using a special pen with their dominant hand, starting in the middle (Supplementary Figure 1). Automatic quantitative analyses were done using custom-made software written in MatLab (version 8.1; The Mathworks, Natick, MA), and processed and visually inspected by 2 trained physicians (S.L. and S.K.L.D.) for analyses (intraclass correlation coefficient [ICC] for interrater reliability for all test components >0.95). A smoothly drawn spiral would have a length of drawing about 56 cm (the length of the template) with little deviation from the template, a low variability in speed, and no crossings (Supplementary Figure 1). The mean amplitude in deviation from the template to spiral drawing (cm) was used as outcome, since it is sensitive to capture small differences in fine motor function.[74] A higher deviation indicates worse performance.

Gait Assessment

Gait was evaluated using a 5.79 m long walkway (GAITRite Platinum; CIR systems, Sparta, NJ: 4.88m active area; 120-Hz sampling rate).[36] The reliability and validity of assessments obtained with this device have previously been established.[96] The standardized gait protocol comprises 3 walking conditions: normal walk, turning, and tandem walk. In the normal walk, participants walked at their usual pace across the walkway. In turning, participants walked at their usual pace, turned halfway, and returned to the starting position. In the tandem walk, participants walked heel-to-toe on a line across the walkway. Based on the recorded footfalls, the walkway software calculated 30 parameters, including 25 from the normal walk, 2 from turning, and 3 from the tandem walk. In Supplementary Table 1, we provide descriptions of the 30 gait parameters.

To summarize these 30 gait parameters into several independent domains, we log-transformed skewed gait parameters to obtain a normal distribution, and subsequently standardized all continuous gait parameters. Next, we conducted a principal component analysis with Varimax rotation to derive gait domains, as previously described.[97] This yielded 7 gait domains with an eigenvalue >1, which we labeled in accordance with the gait parameter that had the highest correlation coefficient with the corresponding domain: rhythm (step time), variability (standardized step length), phases (double support), pace (velocity), tandem (sum of step distance), turning (turning time), and base of support (stride width).[97] These gait domains are illustrated in Supplementary Figure 2. Higher values of the gait domains except “pace”, represent worse gait performance. Based on these 7 gait domains, the Purdue Pegboard Test, and the Archimedes Spiral Test, a total of 9 different facets of motor function were available for analysis.

Assessment of Study Population Characteristics

During home interviews, educational level was assessed and categorized as primary education (“primary”), lower/intermediate general education or lower vocational education (“lower”), intermediate vocational education or higher general education (“intermediate”), and higher vocational education or university (“higher”). Smoking and alcohol habits were assessed during the same home interviews. Participants were categorized as current, former, or never smokers. Alcohol habits were classified into any use or no use of alcohol. At the research center, height and weight were measured from which the body mass index (BMI; kg/m2) was computed. Blood pressure was measured twice in sitting position on the right arm using a random-zero sphygmomanometer, and the average of 2 measurements was used. In addition, non-fasting blood samples were collected and glucose levels were determined. In the

initial subcohort, type 2 diabetes was defined as a random or post-load serum glucose concentration ≥11.1 mmol/L, or the use of drugs to lower blood glucose. In the first and second extension subcohorts, type 2 diabetes was defined as a fasting serum glucose concentration ≥7.0 mmol/L, a non-fasting serum glucose concentration ≥11.1 mmol/L (only if fasting serum was unavailable), or usage of blood glucose lowering drugs. APOE genotype was determined using polymerase chain reaction on coded DNA samples in the initial cohort and with a bi-allelic TaqMan assay in the 2 extensions.[98,99] APOE ε4 carrier status was defined as carrier of one or 2 APOE ε4 alleles.

Statistical Analysis

We assessed trajectories of cognitive and motor function using linear mixed models with random intercepts and slopes. If models did not converge with both random intercepts and slopes, only a random intercept was used. Age of the participant at time of cognitive or motor function assessment was used as underlying time scale. To capture possible nonlinearity, we included natural cubic splines of age with 1, 2, or 3 knots, depending on model performance determined by a likelihood ratio test ($p < .05$). Knots were defined at the median, tertiles, or quartiles for, respectively, 1, 2, or 3 knots. We only reported p-values for each of the age intervals, since appropriate interpretation of effect estimates is hindered by the inclusion of natural cubic splines in the models. Skewed test outcomes (ie, Stroop Tests, Word Learning Test Recognition subtask, Archimedes Spiral Test, and gait domains "variability" and "tandem") were natural log-transformed to reach an approximate normal distribution, and were back-transformed for visualization. In addition, we visualized trajectories of cognitive and motor function by sex, education, or both, using interaction terms on the additive scale between age and sex, age and educational level, and age with sex and educational level. Missing data on education level (1.1%) were imputed by chained equations with 5 iterations. We generated one imputed dataset based on age at baseline and sex. Furthermore, we assessed trajectories for APOE ε4 carriers and non-carriers separately by including an interaction term between age and APOE ε4 status. This analysis was limited to the participants with known APOE ε4 status (N participants = 8986 for cognitive tests and N participants = 7835 for motor tests).

Next, we repeated these analyses by standardizing the cognitive and motor test results to the test performance of the age of 45 years (study baseline) to investigate the temporal course of change across tests with aging. Skewed test outcomes were natural log-transformed before standardization. We depicted a threshold of decline in performance of 0.5 and 1.0 SD compared to the test score at age 45 years. We subsequently assessed the age at which the test score had reached a decline of 0.5

and 1.0 SD compared to the test result at age 45 years. Data were analyzed with SPSS Statistics version 24.0 (IBM Corp., Armonk, NY) and R, CRAN version 3.4.3 "mice" and "nlme" packages.[46,100]

Table 1. Characteristics of the Study Population.

Characteristic	Analysis of Cognitive Function (*N* = 9514)	Analysis of Motor Function (*N* = 8297)
Age at study entry, years, mean (*SD*)	62.0 (7.9)	60.9 (7.4)
Age at first assessment, years, mean (*SD*)	64.7 (9.5)	64.6 (10.0)
Sex, women, *n* (%)	5442 (57.2)	4737 (57.1)
Educational level, *n* (%)		
Primary	1160 (12.2)	886 (10.7)
Lower	3889 (40.9)	3375 (40.7)
Intermediate	2751 (28.9)	2422 (29.2)
Higher	1714 (18.0)	1614 (19.5)
Number of assessments[a], *n* (%)		
1	2058 (21.6)	2136 (25.7)
2	4362 (45.8)	4192 (50.5)
3	1174 (12.3)	1091 (13.1)
≥4	1920 (20.2)	878 (10.6)
Median number of assessments (range)	2 (1–6)	2 (1–4)
Test assessment interval, years, mean (*SD*)	5.1 (1.4)	5.4 (1.4)
Body mass index, kg/m^2, mean (*SD*)	27.0 (4.1)	27.1 (4.2)
Smoking, *n* (%)		
Never	2941 (30.9)	2522 (30.4)
Past	4558 (47.9)	4063 (49.0)
Current	1944 (20.4)	1663 (20.0)
Alcohol use, *n* (%)	7760 (81.6)	6928 (83.5)
Systolic blood pressure, mm Hg, mean (*SD*)	136 (20.8)	136 (20.6)
Type 2 diabetes, *n* (%)	865 (9.1)	735 (8.9)
APOE ε4 carrier, *n* (%)	2539 (26.7)	2217 (26.7)

Notes: *APOE* = apolipoprotein E; *N* = number of participants. Characteristics were measured at study entry except for age at first assessment. Missing values for all characteristics but educational level are not imputed and therefore numbers do not always sum up to 100%.

[a]Gait was considered as one assessment, because virtually all participants (95%) with an available gait assessment had complete values for all underlying gait parameters. Therefore, the presented number of motor assessments is independent from the number of underlying available gait parameters that were used to compute 7 gait domains.

Results

Characteristics of the study population at time of study entry are presented in Table 1. A total of 9514 participants contributed to the cognitive function assessments. The mean (SD) age at first cognitive assessment was 64.7 years (9.5 years) and 5442 (57.2%) of the participants were women. Of all participants, 2058 (21.6%) had a single cognitive assessment, 4362 (45.8%) had 2, 1174 (12.3%) had 3, and 1920 (20.2%) had at least 4 cognitive assessments. The mean interval between tests was 5.1 years (1.4 years). During follow-up up to January 1, 2016, 2977 of 9514 participants (31.3%) died.

A total of 8297 participants contributed to the motor function assessments with a mean (SD) age at first assessment of 64.6 years (10.0 years), of whom 4737 (57.1%) were women (Table 1). Out of these participants, 2136 (25.7%) had a single motor function assessment, 4192 (50.5%) had 2, 1091 (13.1%) had 3, and 878 (10.6%) had 4 motor assessments with a mean (SD) test interval of 5.4 years (1.4 years). Out of 8297 participants, 1903 died (22.9%) during follow-up. The number of participants per cognitive and motor test is shown in Supplementary Table 2. Supplementary Table 3 shows the characteristics of the excluded participants. Overall, excluded participants were older at study entry, attained more often a lower level of education, and had a higher mean systolic blood pressure than included participants.

Trajectories of Cognitive Function

Performance on the cognitive tests declined with advancing age. Decline on cognitive tests was generally linear between ages 45 and 65 years, followed by a steeper, nonlinear decline. Men had higher scores on most cognitive tests and generally declined less fast than women ($p = .003$ for Letter-Digit Substitution Test, $p = .02$ for Word Learning Test: Immediate recall, $p = 0.05$ for Word Learning Test: Delayed recall). These differences between men and women disappeared after assessing the trajectories per educational level, suggesting that this sex difference was largely attributable to differences in the level of attained education between men and women. As such, results from here onwards are presented per educational level for men and women combined.

For each higher level of attained education, participants showed better performance on all cognitive tests at age 45 years (Figure 1 and Supplementary Figure 3). Differences in trajectories of cognitive function between participants with “primary” educational level and participants with other educational levels became larger with advancing age, albeit not statistically significant. Furthermore, participants with “higher” education declined slower than those with “primary” education over time

on the Interference subtask of the Stroop Test (p = .002, Figure 1) and the Word Learning Test Recognition subtask (p = .017, Supplementary Figure 3). However, they declined faster than participants with "primary" education on the Word Fluency Test (p = .048, Figure 1) and the Word Learning Test Delayed recall subtask (p = .007, Supplementary Figure 3).

Regarding APOE ε4 carrier status, carriers declined faster on all cognitive tests than non-carriers (p for interaction between age and APOE ε4 carrier status <.005), except on the Design Organization Test that showed similar trajectories for carriers and non-carriers (Supplementary Figure 4).

Trajectories of Motor Function

Trajectories of decline in motor function varied across different motor tests (Figure 2 and Supplementary Figure 3) with the gait domain "phases" and the Purdue Pegboard Test declining first at the age of 56 and 60 years, respectively. Performance on the gait domains "rhythm," "tandem," and "base of support" remained largely stable over time. Significant differences between men and women were only found for trajectories of the domain "tandem" and "phases," with women performing increasingly worse over age than men (p = .005 for "tandem" and p < .001 for "phases").

In contrast to the effects of education on cognitive function, motor function trajectories were not associated with educational level (Figure 2 and Supplementary Figure 3), but those with a "primary" educational level performed better over time on the Purdue Pegboard Test than participants with other educational levels (p < .016, Figure 2). In addition, they decreased less fast on the gait domains "rhythm," "phases," and "turning" than participants with higher education levels (p for all tests <.039, Figure 2 and Supplementary Figure 3).

APOE ε4 carriers performed worse with advancing age than non-carriers on the Purdue Pegboard Test and on the gait parameters "phases," and "turning" (p for all tests <.034, Supplementary Figure 4).

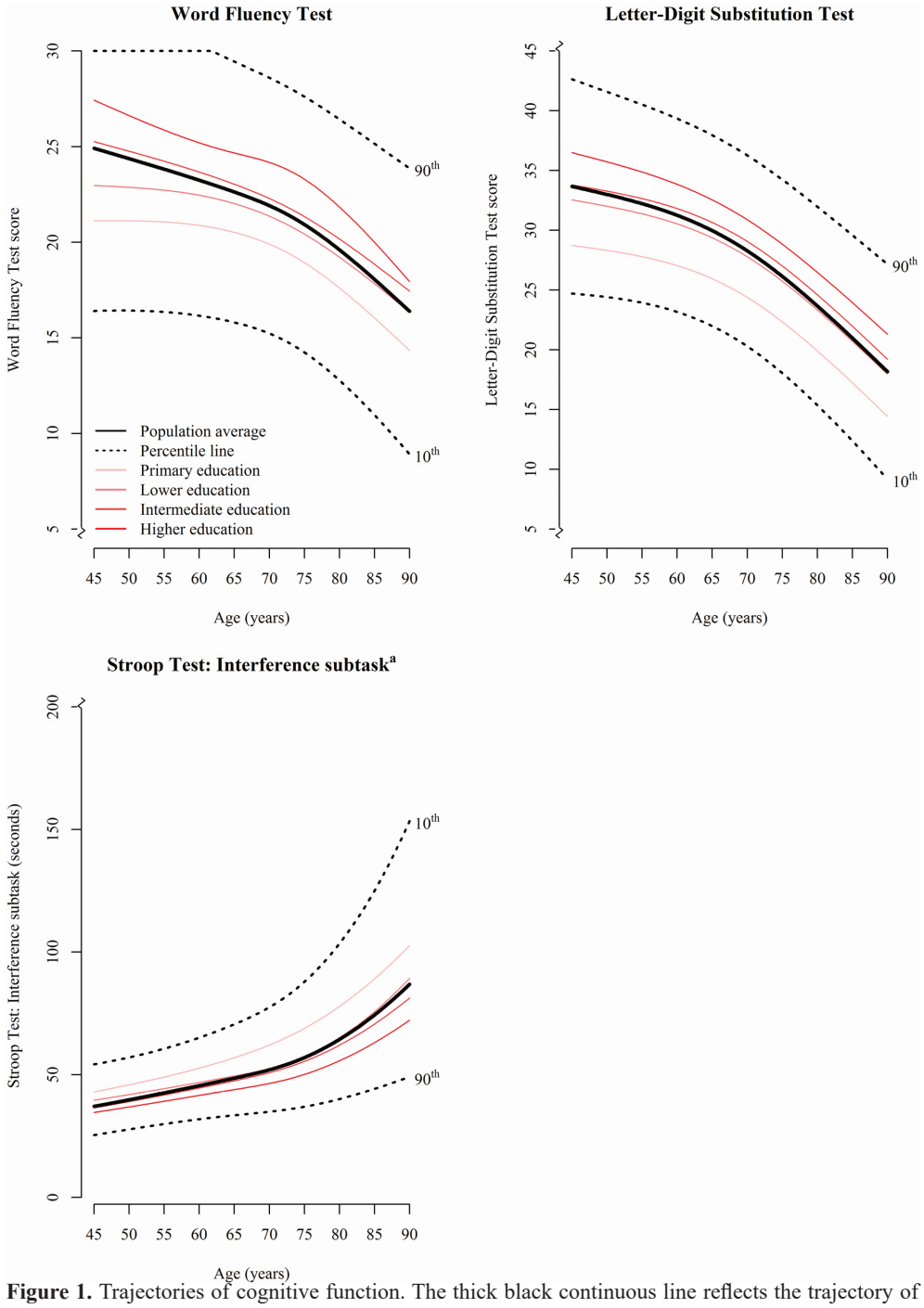


Figure 1. Trajectories of cognitive function. The thick black continuous line reflects the trajectory of performance for the total study population based on the results of the linear mixed model; the black dotted lines represent the 10th and 90th percentile curves. Furthermore, test performance was visualized per educational level in red. Only cognitive tests most commonly studied in studies of cognitive aging are presented in this Figure. The remaining cognitive tests are shown in Supplementary Figure 3. [a]Higher scores indicate worse performance.

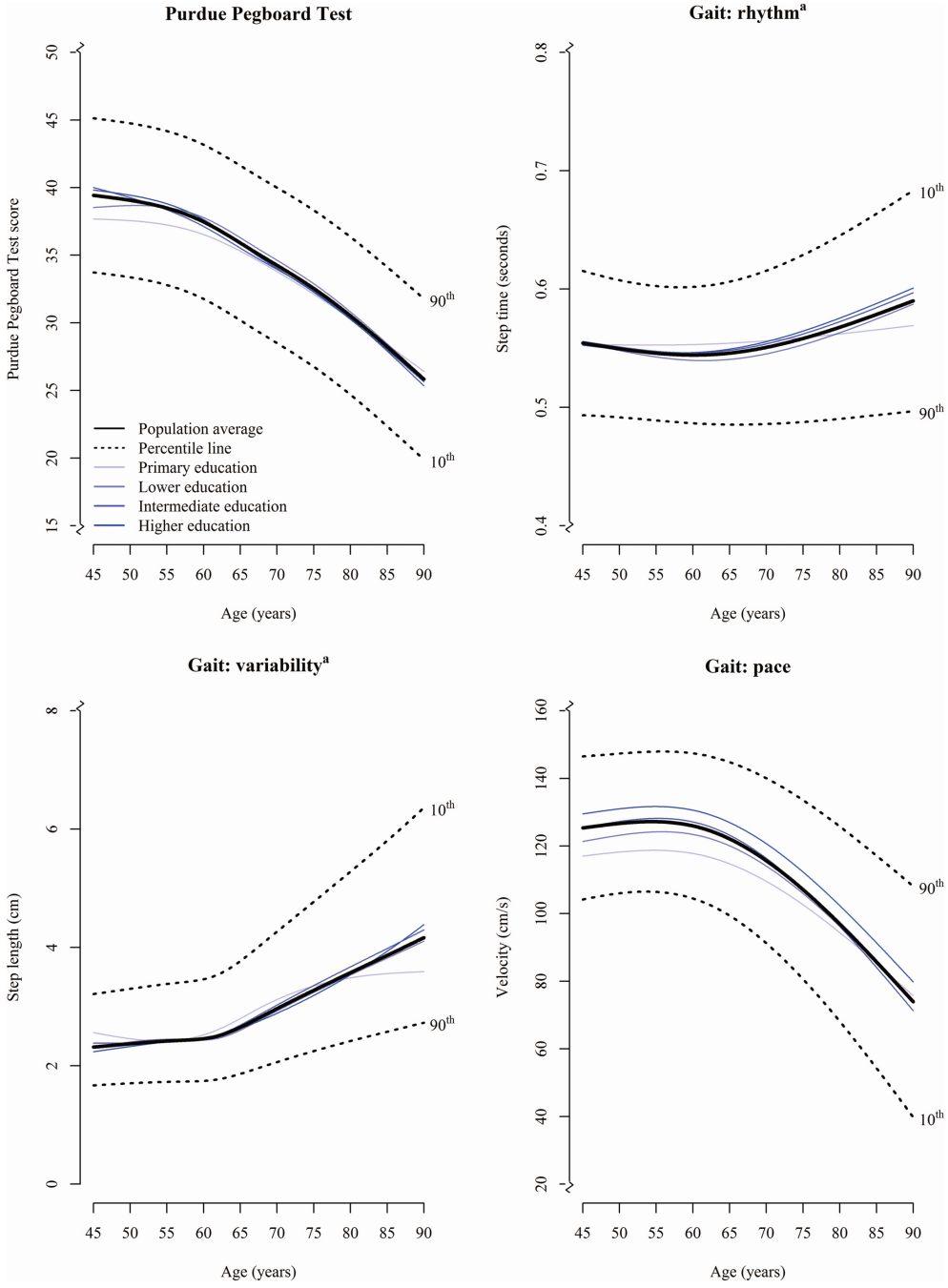


Figure 2. Trajectories of motor function. The thick black continuous line reflects the trajectory of performance for the total study population based on the results of the linear mixed model; the black dotted lines represent the 10th and 90th percentile curves. Furthermore, test performance was visualized per educational level in blue. Only gait domains most strongly related to are presented in this figure. The Archimedes Spiral Test and remaining gait domains are shown in Supplementary Figure 3. [a]Higher scores indicate worse performance.

Sequence of Change in Cognitive and Motor Function

Before the age of 75 years, most cognitive and motor test scores had reached a decline of 0.5 SD in standardized test score compared to test scores at age 45 years (Figure 3). Considering both cognitive and motor tests, the decline of 0.5 SD was first reached for the Stroop Test Interference subtask at the age of 58 years. This was followed by the Design Organization Test at the age of 59 years and the Stroop Test Naming subtask at the age of 64 years. Of all motor tests, the gait domain "phases" showed the fastest decline, reaching a 0.5 SD decrease in test score at the age of 58 years. Across all tests, the average time between the age of 45 years and the age at which 0.5 SD decrease in test score was reached, was shorter for cognitive tests than for motor tests (20.0 vs 24.7 years, respectively, p = .039). By contrast, the time between 0.5 SD and 1.0 SD decrease in test scores was longer for cognitive compared to motor tests (11.2 years vs 8.9 years, respectively, p < .001).

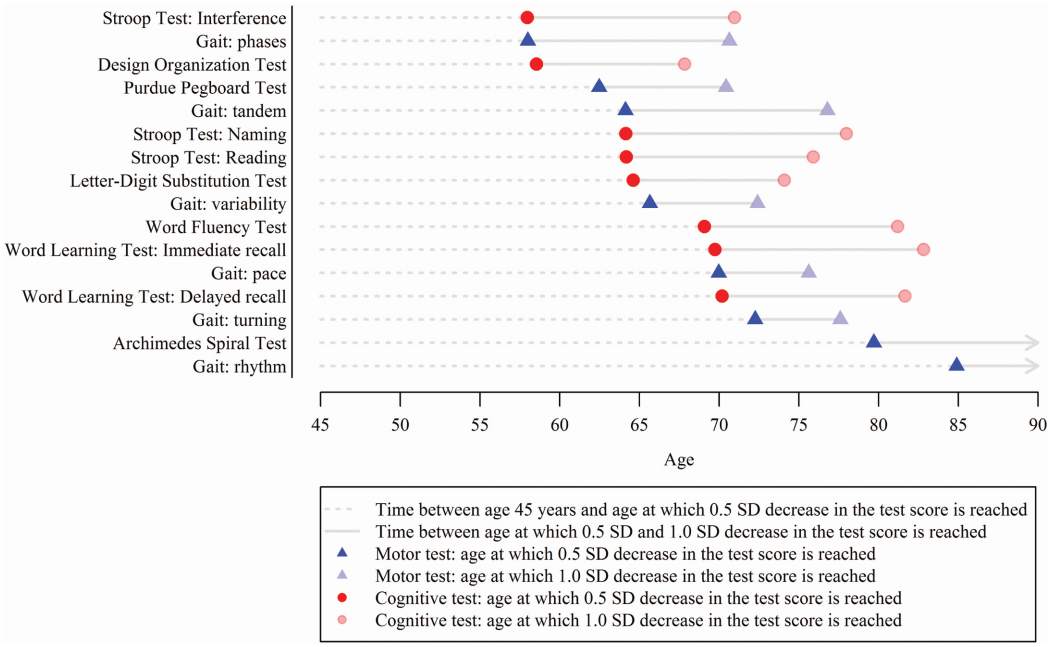


Figure 3. Sequence of decline of cognitive and motor function. Decline was defined as reaching an average population test score of 0.5 or 1.0 SD below the population mean of the test score at age 45 years. The circle or triangle is displayed at the age at which 0.5 SD (opaque) or 1.0 SD (transparent) lower score was reached with cognitive tests depicted in red circles and motor tests in blue triangles. The dotted line represents time between mean population test score at age 45 years and the age at which 0.5 SD decrease in that test score is reached. The continuous line denotes time between the age at which 0.5 SD decrease in the test score was reached and the age at which 1.0 SD in the test score was reached. The Word Learning Test Recognition subtask and the gait domains "tandem" and "base of support" did not reach a score of 0.5 SD lower at a certain age than the score at age 45 years and are therefore not shown. This sequence of decline was estimated based on the total study population. Note that not all participants had all cognitive and motor tests completed.

Discussion

In this population-based study, we showed that both cognitive and motor function generally decline linearly between the ages of 45 and 65 years, followed by a steeper decline after the age of 65–70 years. Test scores for cognitive and motor function declined similarly, with high variation in the rate of decline across age for individual tests. Importantly, whereas a higher level of education was associated with higher cognitive function, there was no association between level of education and function on the majority of the motor tests.

Various studies have reported changes in cognitive function with aging, but evidence on the temporal relation between change in cognitive compared to motor function is limited. Most evidence comes from memory clinics[78], or from studies that solely rely on gait speed to assess motor function.[78,82,101–105] These studies have closely linked global cognitive function to gait speed. As yet, no studies have investigated differences in performance on specific cognitive tests nor studied other facets of motor function, such as fine motor skill. These knowledge gaps remain unaddressed since prior studies found that decline of cognitive and motor function may vary, or that one may predate the other.[79,106–108] Most of these studies were conducted in older participants (aged 70 years and older), with a limited sample size (varying between 488 and 2276 participants), or with relatively short follow-up (ranging from 5 to 7 years). The current study is able to extend these findings by leveraging a detailed set of cognitive and motor tests among a broader age range (ages 45–90 years) in a larger, population-based sample ($N \geq 8297$) with up to 6 repeated assessments during a maximum follow-up of 19.4 years.

We did not find distinct patterns of an overall decline in cognitive function preceding motor function or vice versa, yet we observed large variability in test-specific decline. For instance, tendency to shuffle ("phases" gait domain) and fine motor function generally started to show initial signs of decline up to 25 years earlier than widely used cognitive (screening) tests, such as the Word Learning Tests Delayed recall and Recognition.[78,101–103,109] These findings may be explained by accelerating changes in brain structure during aging, with loss of white matter preceding loss of gray matter.[57,83] We indeed observed the earliest changes in cognitive and motor domains that depend on white matter integrity, including information processing speed, executive function, and the gait domain "phases".[83,110–112] In contrast, cognitive and motor domains related to alterations in gray matter volume (ie, memory and the gait domain "base of support") showed a later decline in function than those related to white matter integrity. [83,111–113]

Variability in test-specific decline may also be explained by diseases and common

comorbidities in these older adults, such as cardiovascular diseases, depression, respiratory diseases, cancer, or impairments in sensory organs.[114–117] These may differentially influence cognitive compared to motor function in some individuals. As an example, presence of peripheral artery disease or arthrosis limits walking speed, but does not directly influences executive functioning as assessed by the Stroop Task.[118] The contribution of these potentially modifiable determinants to sequence of test-specific decline and the shape of these trajectories was beyond the scope of the present study, and warrants further investigation using more advanced statistical models.

As expected, we found that participants with a higher educational level had higher baseline scores (scores at age 45 years) for cognitive tests than participants with a lower educational level. Regarding the rate of change in cognitive function, we found that participants with a "primary" educational level declined faster on most tests than higher educated participants. The declines over time were largely similar among "lower," "intermediate," and "higher" educated participants. This implies that higher educated individuals are generally older when they reach the same cognitive test performance than lower educated individuals. As an example, comparing performance between lower and higher educated participants on the Word Learning Test Delayed subtask score, reveals that at age 45 years, the lowest educated individuals remembered on average 8 of the 15 originally presented words after 10 minutes. The highest educated individuals, however, attained this same score when they were on average 73 years. Yet, no association was found between educational level and motor function for the majority of the motor tests. These findings support emerging evidence that cognitive reserve, operationalized by for example educational attainment, could have long-lasting compensatory effects on cognitive but not on motor function, with the potential to postpone cognitive decline and thereby the clinical diagnosis of dementia.[119–121]

Limitations and Strengths

This study has several limitations. First, given that participants underwent most cognitive tests at the research center, we cannot exclude that selection bias may have influenced our results, with those who are considered less healthy being less likely to participate. Therefore, the presented test scores on cognitive and motor function may be an overestimation of the true performance in the general population, especially for those at older ages.[122] Second, repetitive administering of cognitive tests can lead to learning effects, which could have led to overestimating performance with increasing age. However, these effects are expected to be limited, since the median test interval was 5.1 years for cognitive assessments and 5.4 years for motor assessments. Third, in the early nineties, the completed level of education

was determined by several factors including sex and social economic status. As such, educational attainment in this study may not be a proper proxy for cognitive reserve in women. Lastly, we estimated trajectories of cognitive and motor function on a population level, yet deviations from this pattern on an individual level may signal an under recognized group at high risk for neurodegenerative diseases and stroke. Strengths of this study include the large sample size and the repeated and simultaneous assessments of cognitive and motor function in a single, community-dwelling population.

Conclusions

In this study, we present trajectories of decline of both cognitive and motor functioning among individuals aged 45–90 years in the general population. Such data are essential to understand the natural course of cognitive and motor function during aging. Cognitive and motor function decline similarly during aging, characterized by a linear decline between the ages of 45 and 65, and a steeper decline thereafter. Higher educational attainment was related to higher cognitive function at baseline and to a slower rate of subsequent decline, but it did not affect motor function. In the sequence of decline across individual tests, up to a 25-year age difference between the fastest and slowest declining test scores was observed.

2.3

Cortical gyrification in relation to age and cognition in older adults

S Lamballais, EJ Vinke, MW Vernooij, MA Ikram, RL Muetzel

NeuroImage. 2020; 212:116637

Gyrification of the cerebral cortex changes with aging and relates to development of cognitive function during early life and midlife. Little is known about how gyrification relates to age and cognitive function later in life. We investigated this in 4397 individuals (mean age: 63.5 years, range: 45.7 to 97.9) from the Rotterdam Study, a population-based cohort. Global and local gyrification were assessed from T1-weighted images. A measure for global cognition, the g-factor, was calculated from five cognitive tests. Older age was associated with lower gyrification (mean difference per year = −0.0021; 95% confidence interval = −0.0025; −0.0017). Non-linear terms did not improve the models. Age related to lower gyrification in the parietal, frontal, temporal and occipital regions, and higher gyrification in the medial prefrontal cortex. Higher levels of the g-factor were associated with higher global gyrification (mean difference per g-factor unit = 0.0044; 95% confidence interval = 0.0015; 0.0073). Age and the g-factor did not interact in relation to gyrification ($p > 0.05$). The g-factor bilaterally associated with gyrification in three distinct clusters. The first cluster encompassed the superior temporal gyrus, the insular cortex and the postcentral gyrus, the second cluster the lingual gyrus and the precuneus, and the third cluster the orbitofrontal cortex. These clusters largely remained statistically significant after correction for cortical surface area. Overall, the results support the notion that gyrification varies with aging and cognition during and after midlife, and suggest that gyrification is a potential marker for age-related brain and cognitive decline beyond midlife.

Introduction

Gyrification is one of the most fundamental and distinguishing properties of the human cerebral cortex. The folding patterns of the cortex are highly heritable[123], evolutionarily conserved, and similar amongst closely related animal species[124]. Abnormalities in gyrification, such as polymicrogyria and pachygyria, lead to altered brain function, which can manifest as impairments in speech and cognition. Similarly, both global and regional abnormalities in gyrification have been found in patients with autism[125,126], schizophrenia[127,128], and bipolar disorder[128]. A deeper understanding of gyrification may therefore lead to better insight into a broad range of diseases.

Gyrification changes with age and in turn affects cognitive function.[128–131] The global degree of gyrification is often expressed as the Gyrification Index (GI). The GI peaks during childhood, rapidly declines during adolescence and the decline slows down as adulthood progresses[128–130]. Regional patterns of gyrification can be quantified

with the Local Gyrification Index (LGI).[132] The regions surrounding the angular gyrus, i.e. the parietal cortex, seem most prone to age-related decline in the LGI.[131] The association between the LGI and cognition has been studied in both pediatric and adult cohorts, and it showed the strongest effect in the frontal and parietal regions as well as the temporoparietal junction.[130] These findings consolidate the relevance of gyrification in the normal development of the brain.

Several knowledge gaps still remain. Limited work exists on cortical gyrification during middle adulthood, i.e. 40–65 years of age, and late adulthood, i.e. beyond 65 years of age. Other aspects of the cerebral cortex – such as cortical surface area – change significantly during middle and late adulthood.[83] Furthermore, atrophy of the cerebral cortex seems to accelerates towards the end of life[133], and the rates of atrophy differ between brain regions[83]. How gyrification changes during late life and how the changes are distributed across the brain remains to be elucidated. Similarly, cognitive function declines in aging, which in turn may affect if and how cognition and gyrification relate. Finally, most previous studies were performed in clinical samples or clinic-based settings, limiting the external validity of the findings. The use of population-based studies would allow for better generalization of the results.

The aim of the present study was to elucidate the associations of age and cognition with gyrification during middle and late adulthood. The study was performed using data from the Rotterdam Study cohort, a prospective population-based cohort study of individuals aged 40 years and higher. We hypothesized that age and the GI showed a non-linear association across middle and late adulthood, where the rate of loss of gyrification accelerates with age. Furthermore, based on previous volumetric work, we expected to find that the shape of the association between age and the LGI differed across the brain, with regions near the angular gyrus showing the fastest decline towards the end of life. Finally, we hypothesized that cognition positively associated with the GI, and with the LGI in frontal and temporal regions.

Methods

Study population

The Rotterdam Study is a prospective cohort study based in the Ommoord district of Rotterdam, the Netherlands, that has been ongoing since 1989.[36] The second recruitment wave started in 2000, and the third wave in 2006. All participants are re-invited for an interview and in-person examinations every 4–6 years. The study has included 14,926 participants 45 years of age and older. Neuroimaging was

introduced in 2005.[37] The current study population included individuals who were eligible to participate in a research center visit between 2006 and 2015 with cognitive testing and neuroimaging (n = 6647). Of these, 38 had no cognitive test battery data, 980 had incomplete data, 417 did not participate in the MRI study, in 462 the image surface tesselation in FreeSurfer failed, and 145 were excluded due to poor quality of the T1-weighted images. We further excluded participants with prevalent stroke (n = 126) or prevalent dementia (n = 82). The final sample consisted of 4397 participants. A flow chart of the study population is shown in Supplementary Figure 1. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015). All participants provided written informed consent.

Assessment of cognitive function

All participants underwent a cognitive test battery.[73] The battery consisted of five tests, each assessing different cognitive domains. The first test was the 15-word learning test (15WLT), to assess verbal learning and verbal memory.[134] The 15WLT consisted of three trials where 15 words were presented visually, and after each trial participants had to name all words they could remember (i.e. immediate recall). At least 10 min after the third trial, participants were again asked to name all words that they could still remember (i.e. delayed recall). We used the number of words in the delayed condition as the measurement outcome. The second test was the Stroop task[135], a task that assesses selective attention and automaticity. Participants had to read aloud the names of colors (red, green, blue, yellow) as fast and flawless as possible. The words were printed on paper in mismatching colors (e.g. “blue” printed in the color red) to interfere with the naming process. The time to read all words was adjusted for the number of errors by calculating the time per word and adding one-and-a-half that time for each error. Thus, the Stroop task is inversely coded compared to the other tests, where a higher score relates worse performance. The total was then log transformed and used as the outcome measure. The third test was the letter-digit substitution test (LDST)[90], in which participants have to write down the corresponding digits next to letters according to a dictionary table. This assesses processing speed as well as executive function. Fourth, a word fluency test (WFT) was administered to assess efficiency of searching long term memory[85]. Participants had to name as many animal species in a span of 1 min, with the total number of unique species as the outcome. Finally, we administered the Purdue pegboard test (PPB)[136], where participants had to place small metal pins into holes across three trials: left hand only, right hand only, and both hands. The sum of the number of pins over all trials was used as a measure for fine motor dexterity and psychomotor ability.

To summarize all tests into a single score for global cognition, known as the g-factor, we used principal component analysis and isolated the first component.[137] The g-factor explained 50.6% of the variance amongst the cognitive tests which is in agreement with previous literature.[138]

Image acquisition

Neuroimaging was performed on a 1.5T magnetic resonance imaging (MRI) scanner with an eight-channel head coil (GE Signa Excite, General Electric Healthcare, Milwaukee, USA). The imaging sequences have been described extensively elsewhere[37]. Axial T1-weighted images were collected using a 3D Spoiled Gradient Recalled sequence (TR = 13.8 ms, TE = 2.8 ms, TI = 400 ms, flip angle = 20°, bandwidth = 12.5 kHz, voxel size = 0.8 mm isotropic). The images were subsequently stored in an extensible neuroimaging archive toolkit (XNAT) database.[139]

Image processing

Images were processed using the FreeSurfer analysis suite (version 6.0).[140] The standard reconstruction was conducted, where non-brain tissue was removed, voxel intensities were corrected for B1 field inhomogeneities, voxels were segmented into white matter, gray matter and cerebrospinal fluid, and surface-based models of gray and white matter were generated. The GI was calculated as the ratio between the outer contour of the cortex and the pial surface of the whole cerebrum. The LGI was estimated at each vertex along the cortical ribbon[132,141], and each vertex was automatically assigned an anatomical label according to a predefined atlas[142]. All measures were co-registered to a standard stereotaxis space and smoothed with a full-width half-max Gaussian kernel, 5 mm for the LGI given inherent smoothness and 10 mm for all other measures.

A multistep procedure was used to identify datasets of insufficient quality for analysis. First, we used an automated tool to obtain a quality metric for each T1-weighted scan that assesses artifacts related to motion[143]. Next, we visually inspected FreeSurfer reconstructions from 200 randomly selected scans. The visual ratings consisted of inspecting segmented brain images in the coronal, sagittal and axial directions, as well as 3D reconstructions of the pial surface. The segmentation was rated as a fail if FreeSurfer did not succeed to consistently trace the white and pial surfaces. Next, we established that the automated quality metric value predicted strongly whether a test passed or failed. We subsequently set a threshold above which all scans were of sufficient quality, and all scans below the threshold were excluded.

Measurement of covariates

Hypertension was defined as a resting blood pressure exceeding 140/90 mmHg or the use of blood pressure lowering medication. Blood pressure was measured twice with a sphygmomanometer after 5 min of rest, and the average of the two measurements was used. Use of blood pressure lowering medication was derived from information collected by a physician at the research center. Alcohol use was assessed during home interviews with questions based on beer, wine, liquor and other alcoholic beverages such as sherry and port. Based on these data, an established method was used to calculate alcohol in grams per day.[144] BMI was calculating using the height and weight obtained during the research center visit. Smoking status was obtained during home interviews and individuals were classified as never smokers, past smokers or current smokers. Education level was assessed during the home visit interview and classified into four categories according to the United Nations Educational, Scientific and Cultural Organization classification: primary (no or primary education), low (unfinished secondary and lower vocational), intermediate (secondary or intermediate vocational) or high education (higher vocational or university).

Statistical analyses

All statistical analyses were performed in R 3.4.3.[145] To assess the relation of age and cognition with the GI we used linear regression models. Surface-based LGI analyses were performed to study the spatial distributions of these associations along the cortex. This was done with vertex-wise analyses using the R package QDECR (https://github.com/slamballais/QDECR). Resulting p-value maps were corrected for multiple comparisons at the vertex level using Gaussian Monte Carlo Simulations[146]. Surface-based analyses on cortical thickness and similar measures may show non-Gaussian patterns of spatial correlations, which would increase the false positive rate higher than 0.0531. We therefore set the cluster forming threshold to $p = 0.001$, as this has shown high correspondence with actual permutation testing across all surface measures.[147] We further applied Bonferroni correction to account for analyzing both hemispheres separately (i.e. $p < 0.025$ cluster-wise).

Age-related atrophy of the brain accelerates with age, which may also affect cortical gyrification. We therefore studied three types of associations between age and cortical gyrification: (1) a linear age term, (2) orthogonal linear and quadratic age terms, (3) a B-spline for age with two or three degrees of freedom. The spline knot for the two-fold spline was set at the median age, and the knots of the three-fold spline at the first and second tertiles. The shape of the relationship between age and gyrification was assessed in two steps by evaluating model fit. The linear and non-

linear model fits for the GI were compared by calculating the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for each model. Next, we created a linear model for age and the LGI, and additionally a non-linear model depending on the AIC and BIC for the GI models.

Specific domains of cognition map to different functional regions of the cerebral cortex.[148] Therefore, in addition to the g-factor we also studied whether the scores from the individual cognitive tests associated with GI and LGI. Furthermore, to inspect whether the association between cognition and gyrification changes with age we created a separate model with an interaction term for age and the g-factor.

Models were adjusted for covariates to account for potential confounding. The age analyses were corrected for sex and for study cohort, i.e. the first, second or third cohort of the Rotterdam Study. The cognition analyses were adjusted in three separate models, which allows for the impact of each set of new confounders on model estimates to be described. Model 1 was adjusted for sex, cohort, age at cognitive testing and age difference between cognitive testing and the MRI scan. The way that age entered the model as a covariate – linear, quadratic or with splines – was dependent on the results from the analyses on age and the GI. Model 2 was additionally adjusted for hypertension, alcohol intake, smoking status and BMI. Lastly, Model 3 was additionally adjusted for education level. The p-values for the associations between the potential confounders and the global gyrification index are shown in Supplementary Table 1. To assess whether image quality could affect the results we ran sensitivity analyses with the image quality metric for each scan as a covariate.

Gyrification is calculated as the ratio of the pial surface and the outer surface of the brain.[141] However, the cortical surface area itself has also been shown to relate to cognitive function.[149] Any association between the LGI and cognition may therefore be driven by cortical surface area, and potentially by cortical thickness as well. To further assess this, we performed a sensitivity analysis per cluster. In each model we defined cognition as the outcome, and both the mean LGI and the mean cortical surface area or the mean cortical thickness of each cluster as the determinants. The models were further corrected for all covariates as used in Model 3. We then assessed whether the association between cognition and LGI remained statistically significant, taking into account cortical surface area or thickness.

All covariates had less than 1% missing data except for alcohol use (4.8%). In order to maximize power, missing covariate data were imputed thirty times using multiple imputation by chained equations.[100] Imputed models were subsequently pooled per vertex according to Rubin's rules.[150] We also performed a non-response analysis to

examine whether the individuals who were not included into the final sample were in any way different than those who were included (Table 1). This was done through logistic regression, where inclusion was entered as the outcome and all other variables were included as predictors.

Of note, in all models, we defined age or cognition as the determinants (predictors) and gyrification as the outcome, as limitations in vertex-wise analyses generally only allow for the vertex measure to be modelled as the outcome. Thus, while cognition is generally considered a consequence of brain structure, due to limitations in the vertex-wise software it was defined as a determinant of gyrification in the models. As a sensitivity analysis, we created models for each statistically significant LGI

Table 1. Baseline characteristics of the study population.The excluded sample (n = 2250) were all participants that were eligible for cognitive testing and the neuroimaging study, but did not end up in the final sample (see Supplementary Figure 1).

Characteristics	**Included sample N = 4397**	**Excluded sample N = 2250**	**p-value[b]**
Age at MRI (years)	63.5 ± 10.1	69.5 ± 1.13	<.001
G-factor	0.00 ± 1.00		
Cohort (%)			<.001
RS-I	15.3	33.9	
RS-II	25.3	26.7	
RS-III	59.4	39.5	
Sex, female (%)	55.3	57.8	.278
Time between cognition and MRI (years)	0.3 ± 0.4	0.2 ± 0.3	.847
Hypertensive (%)[a]	61.4	76.1	.053
Alcohol per day (grams)[a]	9.2 ± 10.1	8.5 ± 9.6	<.001
Body mass index (kg/m^2)[a]	27.4 ± 4.1	28.1 ± 4.8	.003
Smoking status (%)[a]			.406
Never	30.9	29.9	
Past	49.1	49.4	
Current	20.0	20.7	
Education level (%)[a]			.002
Primary	7.8	13.1	
Low	37.7	41.0	
Intermediate	30.3	28.4	
High	24.2	17.5	
Mean GI	2.55 ± 0.08	2.52 ± 0.09	<.001

[a]Missingness of data for all variables was below 1% except for alcohol consumption (4.8%).

[b]Differences between inclusion and exclusion were tested through multiple logistic regression.

cluster where the cluster-wise mean LGI was defined as the determinant and the g-factor as the outcome.

All reported results focus on the beta coefficients and the 95% confidence intervals (CIs) rather than p-values. Confidence intervals give insight into the range of values within which the true parameter will likely be, whereas p-values do not[151]. Any reported result that is stated as statistically significant will have a p-value below the threshold of 0.05.

Results

Baseline characteristics of the study population (n = 4397) are displayed in Table 1. The mean age of the participants was 63.5 years (SD: 10.1, range: 45.7 to 97.9) and 55.3% were female. We analyzed whether any differences were present between individuals included in the analysis and those who were eligible for MRI but did not end up in the final sample (Supplementary Figure 1). Excluded participants tended to be older (mean = 6.2 years), were more often from the first cohort of the Rotterdam Study (33.9% versus 15.3%), were less likely to drink alcohol (mean = −0.7 g per day), had a higher BMI (mean = 0.7 kg/m2), were more likely to have only primary education (13.1% versus 7.8%) and had a lower GI (mean = −0.03).

Age and global gyrification

A scatterplot of age and the GI is shown in Figure 1A, and the results of the different models are shown in Table 2. In the linear model one year increase in age associated with a −0.0021 (95% CI: -0.0025; −0.0017) lower GI. For the 2nd polynomial model both the linear term ($p < 0.001$) and the quadratic term ($p = 0.027$) were also statistically significant. All spline coefficients were statistically significant for the natural cubic splines with both two and three degrees of freedom. The AIC and BIC of all models were highly similar, suggesting that the linear fit sufficiently describes the association of age during mid and late adulthood with the GI.

Cognition and global gyrification

A scatterplot of the g-factor and the GI is shown in Figure 1B, and regression coefficients for the g-factor and all separate cognitive tests are shown in Table 3 for all three adjustment models. Higher levels of the g-factor were associated with a higher GI, with similar results across Model 1 ($\beta = 0.0045$, 95% CI = 0.0018; 0.0073) to Model 3 ($\beta = 0.0044$, 95% CI = 0.0015; 0.0073). We examined the individual cognitive tests to see which cognitive tests drove most of the association. Three cognitive tests yielded statistically significant results, namely the LDST

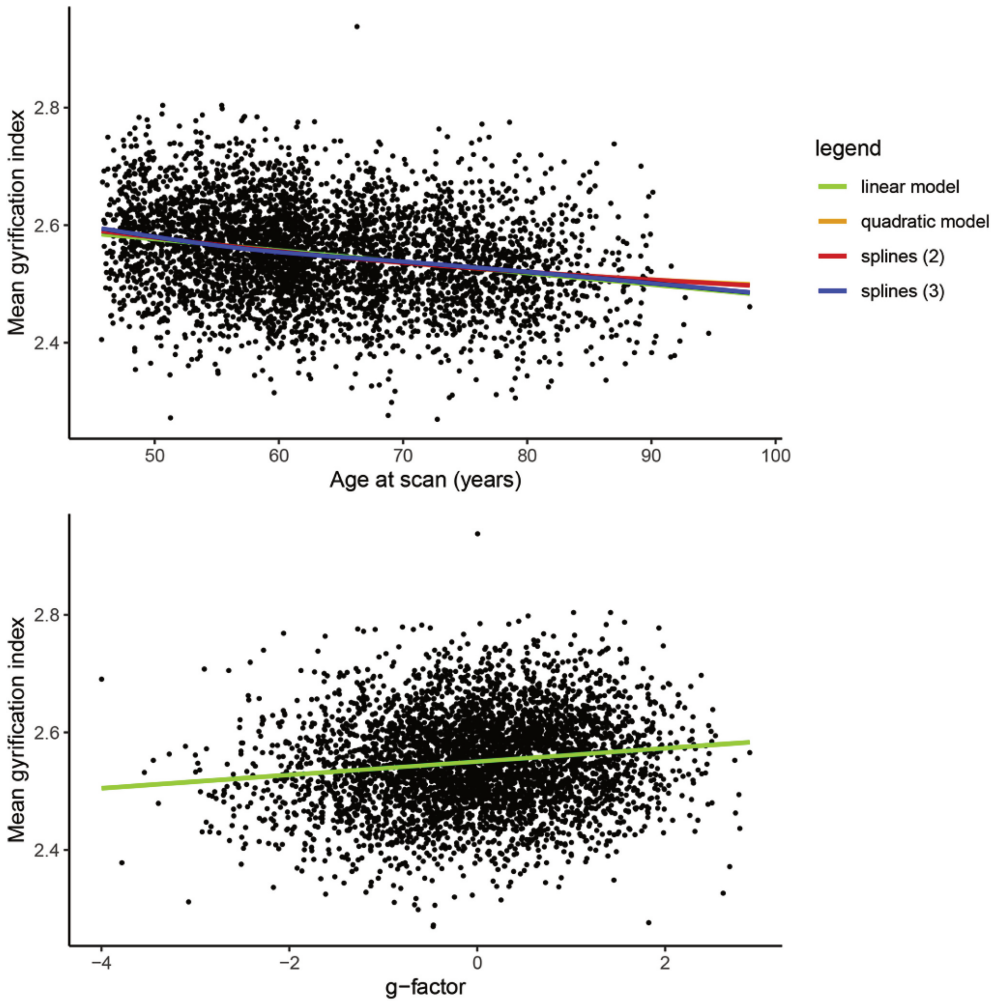


Figure 1. Scatterplot of age (A) and cognition (B) with GI. For age, four models were plotted (linear, quadratic, and the two and three spline models). The lines are not fully visible due to the extensive overlap. The plot for cognition only shows the linear model.

(β = 0.0005, 95% CI = 0.0001; 0.0009), the WFT (β = 0.0009, 95% CI = 0.0005; 0.0013) and the Stroop task (β = −0.0081, 95% CI = −0.0164; −0.0002). Of these, the Stroop task had the strongest association with gyrification. Of note is that the association with the Stroop task was negative due to the lower scores on the Stroop task reflecting higher cognitive performance. Finally, the interaction term between age and cognition did not reach statistical significance in any of the models (all $p_{unadjusted} > 0.05$), thus the magnitude of the association between cognition and the GI was stable during and after midlife.

Table 2. The associations between age and GI.

Model[a]	Type	β	95% CI	AIC	BIC
Linear	Linear	−0.0021	−0.0025; −0.0017	−10252.60	−10214.27
Quadratic	1st polynomial	−0.0053	−0.0081; −0.0024	−10255.47	−10210.75
	2nd polynomial	−0.000025	−0.000003; −0.000046		
Spline (2)	1st spline	−0.1235	−0.2376; −0.1020	−10256.33	−10211.61
	2nd spline	−0.0703	−0.0947; −0.0458		
Spline (3)	1st spline	−0.0552	−0.0384; −0.0718	−10203.46	−10254.57
	2nd spline	−0.1179	−0.0915; −0.1442		
	3rd spline	−0.0826	−0.0550; −0.1102		

[a]The models were adjusted for study cohort and sex.

Abbreviations: GI: Gyrification Index; CI: Confidence interval; AIC: Akaike information criterion; BIC: Bayesian information criterion

Table 3. The associations between cognition and GI.

Domain	GI (Model 1[a])		GI (Model 2[b])		GI (Model 3[c])	
	β	95% CI	β	95% CI	β	95% CI
g-factor	0.0045	0.0018; 0.0073	0.0047	0.0019; 0.0075	0.0044	0.0015; 0.0073
15WLT	−0.0001	−0.0009; 0.0007	−0.0001	−0.0010; 0.0007	−0.0003	−0.0011; 0.0006
Stroop task[d]	−0.0090	−0.0171; −0.0010	−0.0090	−0.0172; −0.0010	−0.0081	−0.0164; −0.0002
LDST	0.0005	0.0001; 0.0009	0.0005	0.0002; 0.0009	0.0005	0.0001; 0.0009
WFT	0.0009	0.0005; 0.0013	0.0009	0.0005; 0.0013	0.0009	0.0005; 0.0013
PPB	0.0001	−0.0004; 0.0007	0.0002	−0.0004; 0.0007	0.0002	−0.0004; 0.0007

[a]Adjusted for age at MRI scan (years), study cohort, sex and age difference between cognitive testing and MRI scan (years).

[b]Additionally adjusted for hypertension (yes/no), alcohol intake (grams per day), BMI and smoking status (never/past/current).

[c]Additionally adjusted for education level (primary/low/intermediate/high).

[d]The Stroop task is inversely coded compared to the other tests, where a higher score relates to worse performance.

Abbreviations: GI: Gyrification Index; CI: Confidence interval; WLT: 15 Word learning test; WFT: Word fluency test; PPB: Purdue pegboard test.

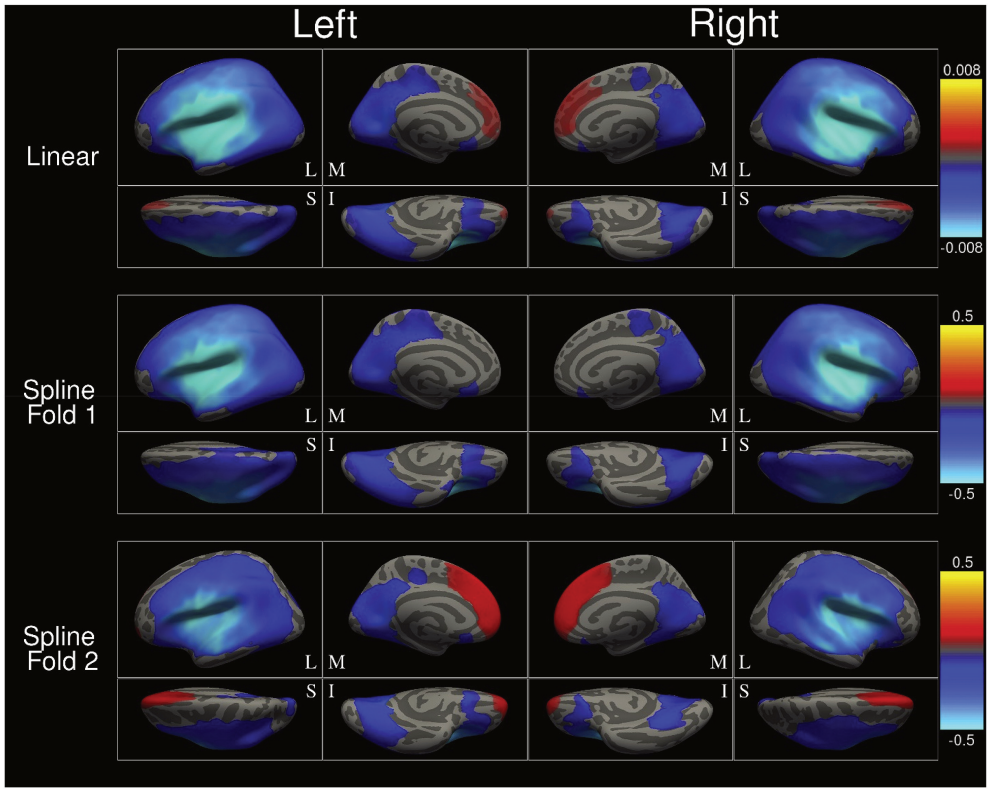


Figure 2. Vertex-wise associations between age – the linear model and the two-fold spline model – and the local gyrification index (LGI). The color scale represents the regression coefficients. The models were adjusted for study cohort and sex. L = Lateral; M = Medial; S = Superior; I = Inferior.

Age and local gyrification

In order to determine the precise spatial extent of associations between age and gyrification, we performed surface-based vertex-wise analyses. Due to the similar fits between the models of age and the GI we opted to further investigate the linear model and the two-fold spline model with the LGI. Figure 2 displays the vertex-wise associations between age and the LGI. In the linear model the LGI decreased with age in the parietal, temporal, occipital and frontal regions. The effect sizes were generally larger than those found when examining the association between age and the GI. A second cluster arose in the frontal pole and medial prefrontal cortex, where the LGI increased with age. The significant clusters were similar across hemispheres in both size and strength. The two-degree spline model differed from the linear model. The first spline fold, i.e. ages between 45.7 and 61.6 years, associated negatively with the LGI in the parietal, frontal, temporal and occipital regions. Unlike the linear model, no cluster was present near the frontal pole or the medial prefrontal pole. In the second spline fold, i.e. ages between 61.6 and 97.9 years, the

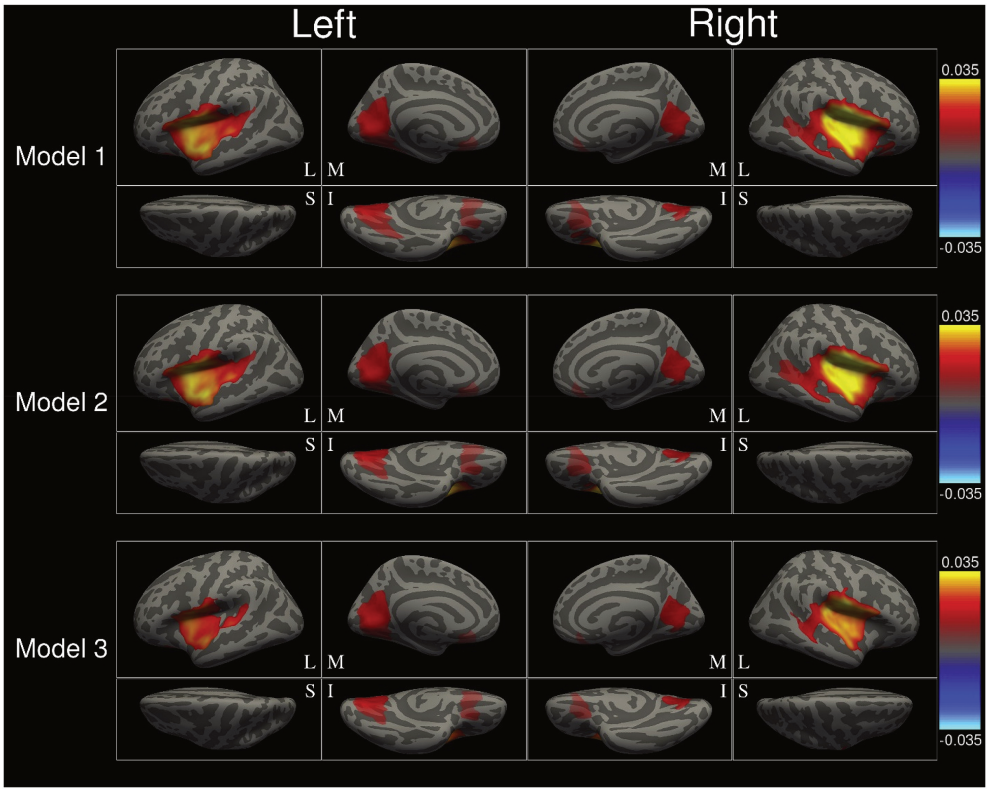


Figure 3. Vertex-wise associations between the g-factor and the LGI. The color scale represents the regression coefficients. Model 1 was adjusted for age (linear term), study cohort, sex and the time difference between the cognitive test battery and the MRI visit. Model 2 was additionally adjusted for hypertension status, alcohol intake, BMI and smoking. Model 3 was additionally adjusted for education level. L = Lateral; M = Medial; S = Superior; I = Inferior.

negative associations were more restricted to the temporal and parietal regions, and the lateral part of the frontal cortex. In addition, a positive cluster was present in the medial prefrontal cortex and the frontal pole, stronger than in the linear model. The age-gyrification association shows a clear deviation in its shape in the medial prefrontal gyrus compared to other regions (Supplementary Figure 2). The findings were robust upon further correction for the image quality metric (Supplementary Figure 3).

Cognition and local gyrification

Figure 3 displays the vertex-wise associations of the g-factor with the LGI for the three adjustment models. Associations between g-factor and the LGI were mostly present in three clusters: (1) the superior temporal gyrus, the insular cortex and the

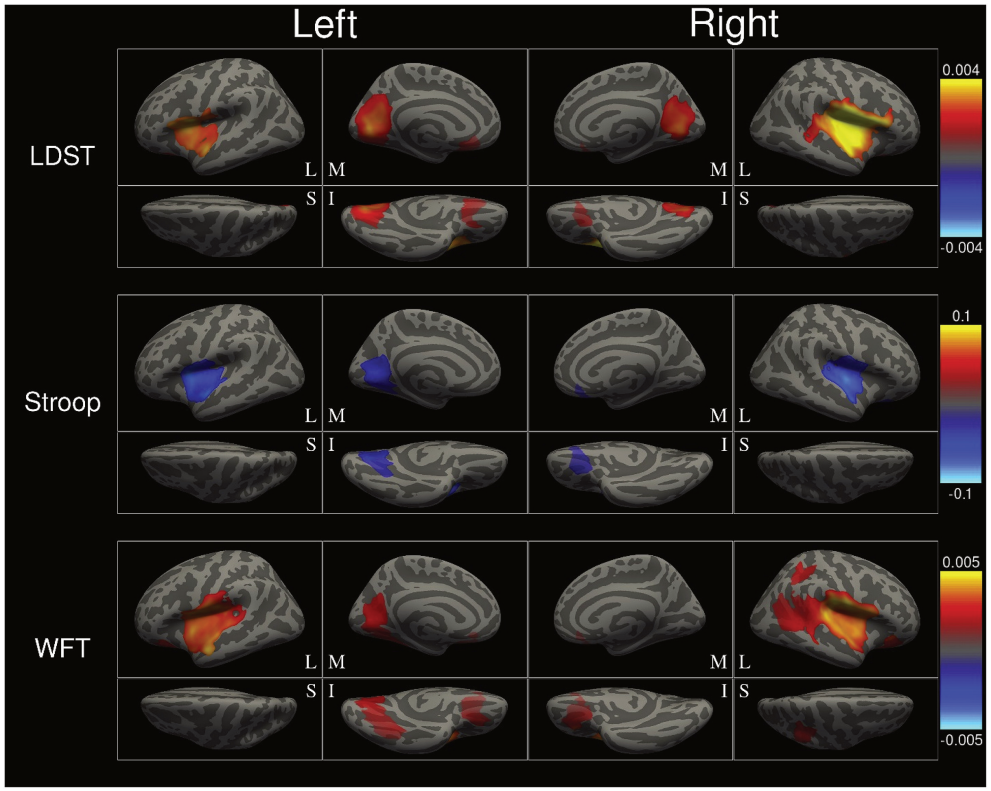


Figure 4. Associations between the individual cognitive tests and the LGI. The color scale represents the regression coefficients. The images show the results for Adjustment Model 3. The Stroop task is inversely coded compared to the other tests, where a higher score relates worse performance. No statistically significant clusters were identified for the 15 word learning test or the Purdue pegboard test, thus these are not displayed. LDST = Letter digit substitution test; WFT = Word fluency test. L = Lateral; M = Medial; S = Superior; I = Inferior.

postcentral gyrus, (2) the lingual gyrus, the precuneus and the pericalcarine cortex and (3) the orbitofrontal gyrus. These clusters roughly presented bilaterally. Similar patterns were found for the associations between the individual cognitive tests and the LGI (Figure 4). The LGI of the cuneate gyrus, insular cortex and superior temporal gyrus were all associated with the Stroop task, the LDST and the WFT, although more so on the right than the left hemisphere. Additionally, the WFT also associated with the LGI in several additional regions, namely the supramarginal gyrus on both hemispheres and the lateral orbitofrontal cortex, the angular gyrus and the superior parietal gyrus on the right hemisphere. Further correction for the image quality metric did not affect these findings (Supplementary Figure 4). Finally, as cognition was originally specified as the determinant, we constructed cluster-wise models with the cluster-wise mean gyrification as the determinant and cognition as

the outcome. For the identified LGI clusters the associations with cognition remained unattenuated (all clusters $p < 0.00001$).

Table 4. The associations of the mean LGI and mean surface area with g-factor per previously identified LGI cluster. Each model contained both LGI and surface area. Both the GI and surface area were standardized in order to be able to compare the magnitude of their effects. All models were adjusted for age at MRI scan (years), study cohort, sex, age difference between cognitive testing and MRI scan (years), hypertension (yes/no), alcohol intake (grams per day), BMI (kg/m^2), smoking status (never/past/current), and education level (primary/low/intermediate/high).

Hemisphere	#	Location	Standardized mean LGI			standardized mean surface area		
			β	95% CI	p	β	95% CI	p
Left	1	Temporal	0.025	[-0.006; 0.055]	.115	0.056	[0.024; 0.088]	.001
Left	2	Cuneus	0.047	[0.016; 0.077]	.002	0.025	[-0.005; 0.056]	.106
Left	3	Orbitofrontal	0.025	[-0.001; 0.050]	.056	0.067	[0.038; 0.096]	<.001
Right	1	Temporal	0.045	[0.014; 0.076]	.004	0.041	[0.009; 0.073]	.012
Right	2	Cuneus	0.028	[-0.005; 0.073]	.094	0.029	[-0.004; 0.063]	.085
Right	3	Orbitofrontal	0.030	[0.005; 0.054]	.019	0.061	[0.034; 0.089]	<.001

Table 5. The associations of the mean LGI and mean cortical thickness with g-factor per previously identified LGI cluster. Each model contained both LGI and cortical thickness. Both the LGI and cortical thickness were standardized in order to be able to compare the magnitude of their effects. All models were adjusted for age at MRI scan (years), study cohort, sex, age difference between cognitive testing and MRI scan (years), hypertension (yes/no), alcohol intake (grams per day), BMI (kg/m2), smoking status (never/past/current), and education level (primary/low/intermediate/high).

Hemisphere	#	Location	Standardized mean LGI			standardized mean thickness		
			β	95% CI	p	β	95% CI	p
Left	1	Temporal	0.063	[0.038; 0.088]	<.001	0.062	[0.039; 0.085]	<.001
Left	2	Cuneus	0.069	[0.045; 0.094]	<.001	0.045	[0.022; 0.068]	<.001
Left	3	Orbitofrontal	0.044	[0.020; 0.068]	<.001	−0.025	[-0.049; −0.000]	.047
Right	1	Temporal	0.075	[0.050; 0.100]	<.001	0.061	[0.037; 0.085]	<.001
Right	2	Cuneus	0.058	[0.034; 0.083]	<.001	0.050	[0.027; 0.073]	<.001
Right	3	Orbitofrontal	0.049	[0.026; 0.073]	<.001	−0.003	[-0.027; 0.021]	.790

Both vertex-wise cortical surface area and thickness associate with age (Supplementary Figure 5) and the g-factor (Supplementary Figure 6). In order to assess whether the associations between cognition and the LGI were driven by cortical surface area or thickness, we created a new model for each significant LGI cluster with the g-factor as the outcome and both the mean LGI and the mean cortical thickness or surface area as determinants. Associations remaining after concurrent adjustment for cortical surface area or thickness suggest an independence between LGI the other measures. The results are shown in Table 4 for surface area and in Table 5 for cortical thickness. After adjustment for surface area, the LGI remained associated with the g-factor in the left hemisphere in the cluster near the cuneus ($p_{unadjusted}$ = .002) but not the clusters in the orbitofrontal cortex ($p_{unadjusted}$ = .056) or the temporal cortex ($p_{unadjusted}$ = .115). In the right hemisphere the association between the LGI and cognition was unaffected by surface area in both the orbitofrontal ($p_{unadjusted}$ = .019) and the temporal clusters ($p_{unadjusted}$ = .004), but not in the cluster in the cuneus ($p_{unadjusted}$ = .094). In all clusters, the LGI was unaffected by additional corrections for cortical thickness.

Discussion

We show in a large population-based setting that global gyrification of the cerebral cortex decreases during middle and late adulthood. This decline in gyrification is mainly driven by regions close to the Sylvian fissure. A specific cluster within the medial prefrontal cortex showed more gyrification with increasing age, particularly during late adulthood. Furthermore, we also found that global cognition positively associates with gyrification, in particular in the temporal regions, the lingual gyrus and the cuneus.

The findings for age and gyrification are in line with previously reported findings in smaller or younger age samples. A study from 2017 attempted to map the life course trajectory of the GI in a cross-sectional sample of 881 participants.[128] The authors found that the GI trajectory can be described as a negative logarithmic function, with the decline in gyrification slowing with age. However, their sample included only about 30 participants above the age of 60. Another study reported on the association of age and gyrification in 322 healthy adults of whom 116 were aged 60 or older.[131] They found that the LGI had non-linear associations with age in certain brain regions, especially the orbitofrontal and dorsomedial prefrontal cortex. In particular, LGI in these regions seemed to increase towards the end of life. Our study builds upon these findings, with a much larger number of participants beyond the age of 70 years, allowing us to more precisely study how gyrification changes during late adulthood. We found that the association is essentially linear, which matches a

negative logarithmic life course pattern.[128] We also established non-linear patterns in the regional surface-based patterns, albeit the increase in LGI was only seen in the medial prefrontal cortex and only towards the end of life. These findings further consolidate the global and local dependence of gyrification on age.

Gyrification associates with cognition during and after midlife, and this association does not change with age. Furthermore, in half of the significant clusters we found that cortical surface area is likely driving the associations. This is not surprising as loss of surface area leads to lower folding thus lower gyrification within that region. Still, three out of six LGI clusters remained associated with cognition after adjusting for surface area, suggesting that gyrification harbors independent information. Furthermore, the pattern of LGI clusters within our study was similar amongst the individual cognitive tests, suggesting that the LGI captures a more general aspect of cognitive function. Gyrification may therefore play a unique role in cognitive function, which could prove useful in the study of normal and abnormal cognitive aging. For example, other cortical characteristics such as thickness and surface area have distinct contributions to cognitive decline as seen in Alzheimer's disease[152–154], yet any such contributions from gyrification remain to be elucidated.

The temporal lobe has previously been linked to cognitive processes such as language[155] and memory[156] as well as psychiatric disorders like adulthood autism spectrum disorders[157] and schizophrenia[158–160]. Interestingly, these disorders have also been linked to abnormal gyrification.[125–128] Genetic mechanisms underlying cognitive processes and neuropsychiatric disorders may also affect cortical morphology in the temporal region, and in particular gyrification. Previous studies have found links between genes underlying cognition function and temporal lobe structure[161,162], although the results are inconsistent.[163] Thus further work is needed to elucidate the presence of a genetic pleiotropic link between gyrification and function of the temporal lobe.

Several mechanisms could explain how gyrification changes with age. One explanation is that during brain development the cortical surface buckles due to differential rates of growth of cortical layers[164], and the opposite may occur during adulthood. The rate of atrophy is higher in gray than white matter during early and mid adulthood.[165,166] Gray matter atrophy is mostly through the reduction of surface area of the cortex, which leads to more shallow sulci and consequently a lower GI. The rate of atrophy of white matter starts to exceed the rate for gray matter during late adulthood[83], which could in turn lead to an increase in gyrification with age. Interestingly, a previous study found that after the age of 60 years the cingulate cortex thickens and that the rate of thinning of the medial prefrontal cortex declines[167], which could explain the increased gyrification of the medial prefrontal

cortex in our study. Indeed, we see a similar thinning of the cingulate cortex. However, our findings also suggest that gyrification overall keeps decreasing during older adulthood, thus other mechanisms than the different gray and white matter atrophy rates are also likely involved.

Another plausible explanation is the "axon tension" theory[168], which states that axonal tension pulls the gyral walls inwards, thus folding the cortex. The axonal tension may depend on the health status of the axon, and damage to axons could lead to reduced tension and consequently decreased gyrification. White matter microstructure decreases with age[169,170] and white matter lesions accumulate from mid adulthood onwards[83], and could explain the decrease in gyrification. However, further experimental work has discredited axonal tension as a cause of cortical folding. For example, if axonal tension causes gyrification then cutting the gyrus transaxially should unfold the gyrus, and experiments have shown that this is not the case.[171] Thus, further work is needed to elucidate the causes of cortical (un)folding during adulthood.

Gyrification may also associate with age due to more technical aspects of the data collection itself. Head motion may affect the relation between age and GI.[172] The reasoning for this is that older participants tend to move more with their head while in the MRI. A previous study confirmed this and also found that head motion related to LGI, although the association was not very strong.[173] We attempted to minimize the impact of head motion on the analyses by conservatively excluding all raw images with suboptimal quality and further by performing sensitivity analyses with the image quality metric as a covariate.

The study has several limitations. First of all, we relied on a cross-sectional study design to examine age effects on gyrification. Cross-sectional estimation of age-related changes may yield inaccurate estimates compared to longitudinal designs.[54] Second, changes in gyrification likely cause changes in cognition, but the models were specified with cognition as the determinant and gyrification as the outcome due to limitations in the vertex-wise analysis modeling. Rerunning the models per cluster with proper specification did show that the LGI clusters indeed associated with cognition, suggesting that the models hold under proper specification. Third, the cognitive test battery that was used does not cover all aspects of cognitive function. Due to the emphasis on verbal tests we were not able to fully assess the scope of cognition and gyrification. Fourth, in the case of cognition there may be reverse causality, as higher intelligence tends to lead to a healthier lifestyle and thus better brain health. We corrected for a number of variables related to lifestyle and their effect on the association was minimal. Despite this there could still be residual confounding by other variables that we did not account for. Fifth, while we excluded

those with prevalent stroke and dementia, there could be other medical conditions and confounders that could bias the results. For example, traumatic brain injury and substance abuse disorders are known to accelerate brain and cognitive aging[174–176], and could subsequently affect the association of age and cognition with gyrification. Our study also had several strengths. First, this is the largest sample size to date in a study of gyrification and age or cognition, leading to sufficient statistical power to find associations, and unravel new regional differences. Second, the individuals were sampled from a wide age-range, thus enabling making accurate inferences about gyrification even in the later phases of late adulthood. Third, the sample was drawn from a population-based cohort, thus the findings can be generalized beyond a clinical setting.

In conclusion, gyrification globally decreases linearly with age across the entirety of adulthood, and gyrification in the medial prefrontal cortex increases towards the end of life. Furthermore, gyrification increases with higher levels of cognitive performance in some clusters irrespective of surface area. These findings consolidate the importance of gyrification in normal brain function. Whether gyrification is a viable marker for abnormal brain aging and cognitive decline towards the end of life remains to be elucidated.

2.4

Identifying subtypes in brain ageing through data-driven disease progression models

EJ Vinke, AL Young, NP Oxtoby, DC Alexander, V Venkatraghavan, EE Bron, MA Ikram, MW Vernooij

Submitted

2.5

Brain aging: more of the same?!

EJ Vinke, MA Ikram, MW Vernooij

Aging (Albany NY). 2019;11(3):849-850

Editorial

With the number of patients suffering from brain diseases set to soar in coming decades[84] it has become of great importance to not only understand normal functioning of the brain but also unravel when, how, and where in the brain deviations take place from the path of normal aging towards pathological degeneration and ultimately clinical disease. A first essential step is to map normal aging trajectories of the structure and function of the brain. This research question is as simple in its statement as it is complex in its operationalization.

Magnetic resonance (MR) imaging has been the single most important contributor to the in vivo investigation of brain structure and function. Over the last three decades, an ever-increasing number of imaging datasets have become available capturing every phase of the human lifespan. With these datasets, trajectories of brain imaging markers in an aging population can be estimated, based on a single brain scan per subject. With a single measurement, taken at a single moment in time, merely a snapshot of the dynamic process of aging is taken. Considering the high inter-subject variation, investigating what brain aging is with only cross-sectional data is challenging.

While these cross-sectional studies have laid an important and solid foundation, to investigate brain aging in more depth and to better discern between patterns of normal versus abnormal aging, longitudinal data is essential to provide better insight into the timing and sequence of changes in aging.[54] Compared to cross-sectional studies, with longitudinal data one is able to investigate when and how the deviations from normal aging occur, rather than the average absolute differences between young and older subjects. Distinguishing different trajectories based on longitudinal data can be a starting point to study why certain persons show a different aging pattern than others, which factors drive these differences, and what functional outcomes these relate to.

A frequently used approach to investigate different patterns of changing imaging markers over time, is to simply subtract measurements from two different time points to identify subjects that increase, decrease or remain stable over time. Once the subjects belonging to each of these categories are identified, comparing the population characteristics between these groups could point towards potential factors influencing the trajectories. Although this approach could be an important first exploration, we believe that this use of longitudinal data is not living up to its

full potential. First, classification of subjects into these three categories based on only two longitudinal measurements could be very sensitive to noise and therefore lead to misclassification. Second, the simplification of the course of the trajectories makes it less sensitive to detect more subtle slope differences. Third, though in this approach the slope of the subject-specific trajectories is taken into account, the age on which these slope differences occur is not.

To give an example of how one can use longitudinal data to further explore different patterns of imaging markers, assume the hypothetical situation that we have identified using the method described above a potentially important factor that could influence the aging trajectory: smoking. Figure 1A shows fictional subject-specific trajectories of smokers and non-smokers for a specific imaging marker. We can then use the subject-specific trajectories to study whether and how the differences in the trajectories are explained by smoking. Instead of assuming that everyone follows the same aging trajectory, only allowing for a different starting point (Figure 1B), approaching smoking as an effect modifier of the effect of age on the marker is a way to capture more subtle changes in the shape of the trajectory (Figure 1C). Even though everyone follows their own trajectory, the effect of smoking can be determined by estimating to which extent each subject-specific trajectory is explained by smoking. Combining all that information gives an overall approximation of the effect of smoking, that would have been missed or underestimated when simply assuming that everyone has the same shape and that smoking only influences the intercept.

Next to the ability to investigate different trajectories of a single imaging marker, longitudinal data on not just one, but several markers within the same subject are essential to assess how changes in these markers coincide and interact, and to assess the temporality of these events. In our recent work we investigated the trajectories and sequence of changing structural brain imaging markers in a large aging population, using longitudinal brain imaging data.[83] The resulting sequence of changing markers could be interpreted as an average sequence of the broad spectrum of normal aging in this population. We believe that within aging research, the field can take example of research performed in the setting of diseases with a very heterogeneous clinical presentation, such as multiple sclerosis. In these diseases, timing of events such as the onset of certain symptoms or presence of disease markers are already being used to investigate and identify subtypes of disease and to predict progression of disease. Considering that we may never be able to draw a clear line between normal aging and abnormal aging, we believe that with a special focus on the timing and sequence of events in brain aging, we may also be able to identify different patterns of aging in a similar way. This could greatly advance research into brain health in old age.

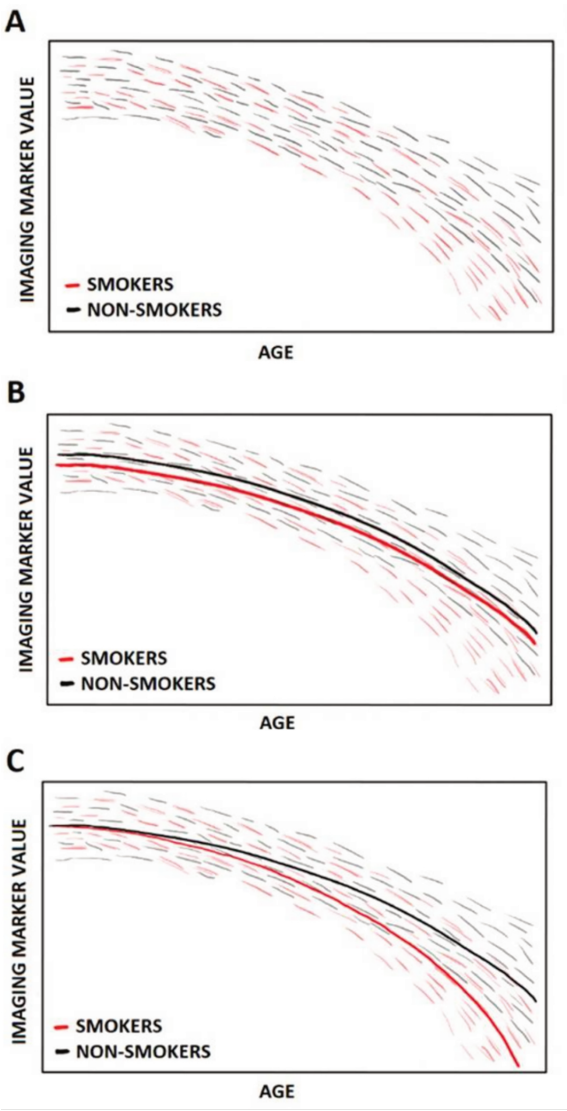


Figure 1: Fictional subject-specific trajectories of an imaging marker in aging, from smokers and non-smokers (A), with population-based trajectories assuming that every subject follows the same trajectory only allowing for a different starting point (B) and population-based trajectories with smoking as an effect modifier (C).

To summarize, more imaging data on the same subjects gives us the opportunity to focus on timing and sequences of changes, which can help us to identify different patterns within the broad spectrum of normal aging. This would bring us one step closer towards understanding the sources of variability, and their implications, within 'normal' aging.

3

Determinants of brain aging

3.1

Intracranial arteriosclerosis is related to cerebral small vessel disease and accelerated brain atrophy: a prospective cohort study

EJ Vinke*, P Yilmaz*, JE van der Toorn, R Fakhry, K Frenzen, F Dubost, S Licher, M de Bruijne, M Kavousi, MA Ikram, MW Vernooij, D Bos

Neurobiology of Aging. 2021;105:16-24

Intracranial arteriosclerosis has been increasingly recognized as a risk factor for cognitive impairment and even dementia. A possible mechanism linking intracranial arteriosclerosis to cognitive impairment and dementia involves structural brain changes including cerebral small vessel disease (CSVD). To assess whether intracranial carotid artery calcification (ICAC) and vertebrobasilar artery calcification (VBAC), as proxies for intracranial arteriosclerosis, are related to CSVD. Within the population-based Rotterdam Study, between 2003 and 2006 a computed tomography (CT)-based measurement of ICAC and VBAC and at least one magnetic resonance imaging (MRI) measurement of structural brain changes were performed from 2005 onwards in 1,489 participants. To estimate the burden of calcification independent of age, we computed age-adjusted percentile curves for ICAC and VBAC separately, based on the calcification volumes. Using the longitudinal MRI data, we assessed whether a larger calcification burden accelerates structural brain changes using appropriate statistical models for repeated outcome measures. A larger burden of ICAC and VBAC was associated with an increase of CSVD markers accelerating over time. A larger burden of ICAC and VBAC was not significantly ($p > 0.05$) associated with accelerated brain atrophy. Arteriosclerosis is related to accelerating structural brain changes over time.

Introduction

Increasing evidence suggests an important role for arteriosclerosis in the etiology of cognitive decline, and even dementia.[208,209] The potential mechanism underlying the link of arteriosclerosis with cognitive decline and dementia may be the presence of subclinical, structural brain changes, including cerebral small vessel disease (CSVD). Indeed, previous studies have shown that markers of systemic arteriosclerosis are associated with white matter hyperintensities (WMH) and cerebral microbleeds (CMB), and smaller brain tissue volumes.[210–215] Despite these important insights into the role of arteriosclerosis in the etiology of structural brain changes, important knowledge gaps remain.

First, most studies focusing on the link between arteriosclerosis and structural brain changes were cross-sectional and thus unable to capture the aging process in the relation between arteriosclerosis and brain structure. Longitudinal data combining image-based measurements of arteriosclerosis (CT) with those assessing brain structure (MRI) are needed. Linkage of such data in a population-based setting with

a broad age-span could answer the question whether arteriosclerosis is associated with accelerated brain structure changes, or whether this is merely an overall difference in brain structure for which it is unclear when this difference originated.

Second, during the last decade it has become increasingly evident that the burden of arteriosclerosis, although a systemic disease, may differ considerably across arteries within the same person.[216] This has also shown to translate into a differential predictive value for various clinical manifestations of arteriosclerosis depending on the proximity of the artery to the organ under study.[208,217] Hence, vascular disease in the intracranial arteries in particular may play an important role in the etiology of structural brain changes. Yet, until now, most studies that investigated the effect of arteriosclerosis on brain structure focused on the coronary or carotid arteries only.[213–215,218–221] Moreover, if studies investigated intracranial arteriosclerosis at all, these generally focused on the anterior cerebral circulation.[219,222] Yet, the posterior, vertebrobasilar arterial system accommodates critical blood supply to the posterior structures of the brain, but little is known regarding the influence of arteriosclerosis in this vascular system on brain structure in aging.

To address these two major knowledge gaps, we investigated whether intracranial carotid artery calcification (ICAC) and vertebrobasilar artery calcification (VBAC), as proxies for intracranial arteriosclerosis, are associated with accelerated structural brain changes over time.

Methods and Materials

Study population

This study is embedded within the Rotterdam Study, an ongoing population-based study designed to investigate causes and determinants of age-related diseases since 1990. Participants aged ≥45 years are interviewed at home and examined at the research center at baseline and during follow-up visits every 3−4 years. The design of the Rotterdam Study has been described previously.[194]

Between 2003 and 2006, participants were invited to undergo multi-detector computed tomography (MDCT) of the intracranial internal carotid arteries and vertebrobasilar arteries, as part of a CT-imaging study aimed to visualize calcification in major arteries. In total, 2,524 participants were scanned (response rate, 78%). Due to imaging artefacts, 60 examinations were not gradable for the presence of calcification in above mentioned arteries, leaving a total of 2,464 participants available for analyses. Since August 2005, brain MRI was implemented in the core Rotterdam Study protocol, and repeated brain imaging was performed

during consecutive research visits.[194] out of the 2,464 participants with complete CT examination, 1,681 participants underwent ≥1 MRI scan(s). A flowchart of the number of in- and excluded participants with a CT, and the number of MRI scans for each individual is shown in Supplemental Figure 1. All available MRI scans from participants with a complete CT examination were included (n = 3,322) of which brain MRI with incomplete acquisition, processing or CSVD ratings were excluded (n = 250). Additionally, scans from participants who had a history of dementia or symptomatic stroke (n =153) and scans with MRI-defined cortical infarcts (n = 69) were excluded. In total, 2,850 MRI scans from 1,489 participants were available for analysis.

Assessment of intracranial arteriosclerosis

A 16-slice (n = 724) or 64-slice (n = 1690) MDCT scanner (Somatom Sensation 16 or 64, Siemens, Forchheim, Germany) was used to perform non-contrast CT scanning. Detailed information regarding imaging parameters is described elsewhere.[216] On a scan, ranging from the aortic root to the Circle of Willis, calcification in the intracranial carotid arteries was quantified from the horizontal segment of the petrous internal carotid artery to the top of the internal carotid artery. (Bos et al., 2012a) Calcification in the intracranial vertebral arteries was assessed from the level at which the vertebral arteries enter the skull to the level of merging into the basilar artery. Calcification in the basilar artery was assessed from the merge of the vertebral arteries to the top of the basilar artery.[223]

ICAC and VBAC were quantified by a semiautomatic scoring method that allows for manual segmentation of calcification on CT-images[223,224]. After manual delineation of the calcification, calcification volumes (mm3) were computed by multiplying the number of pixels within the delineated area above 130 Hounsfield units by the pixel-size and slice increment.[218] The total volume of ICAC was calculated by summing up the calcification volumes of the left and right intracranial carotid arteries. VBAC was calculated as the sum of the calcification volumes of the vertebral arteries and the basilar artery.

Assessment of structural brain changes

Brain MRI scanning was performed in all participants during the entire study period on the same single 1.5-tesla MRI scanner (GE Signa Excite; GE Healthcare, Milwaukee, USA). The scan protocol and sequence details have been described elsewhere. (Ikram et al., 2020) In short, the protocol comprised of T1-weighted, PD, Fluid-attenuated inversion recovery (FLAIR) and T2* gradient-recalled echo images for assessment of focal CSVD markers and quantification of volumetric markers.

Focal markers

Trained research physicians rated the presence, number and location of cortical infarcts and lacunes, CMB and PVS.[194] Infarcts showing involvement of cortical gray matter were classified as cortical infarcts. Lacunes were defined as subcortical lesions between ≥3mm and <15mm with signal intensity similar to CSF on all sequences, and a hyperintense rim on the FLAIR sequence when located supratentorially.[44] CMB were rated as focal areas <10mm of very low signal intensity on T2*-weighted imaging.[194] PVS were determined according to a standardized protocol and defined as linear, ovoid or round-shaped hyperintensities of ≥1mm and <3mm on proton density-weighted images.[45] The numbers of enlarged PVS in the midbrain, hippocampi, basal ganglia and centrum semiovale were quantified using an automated PVS algorithm, developed by Dubost et al.[225] Further description of the PVS algorithm can be found in the Supplemental Materials. Fluid-attenuated inversion recovery (FLAIR) scans were used for the automated segmentation of WMH volume, which is described in further detail for the brain atrophy measures.

Volumetric markers

For brain atrophy, T1-weighted (voxel size 0.49 × 0.49 × 1.6mm3), proton density weighted (voxel size 0.6 × 0.98 × 1.6mm3), and FLAIR (voxel size 0.78 × 1.12 × 2.5mm3) scans were used for automated segmentation of supratentorial gray matter, white matter, cerebrospinal fluid (CSF) and WMH. This automated segmentation was based on a k-nearest neighbor classification algorithm.[38,39] All scans were transformed to the high-resolution data set (256 × 256 × 128) using tri-linear interpolation. Automated processing tools from the Brain Imaging Center, Montreal Neurological Institute and McGill University (www.bic.mni.mcgill.ca) were used to co-register MRI data (based on mutual information) and subsequently normalize the intensities for each feature image volume using N3[40]. All segmentations were visually inspected and manually corrected if needed. Total cerebral volume was the sum of gray matter, normal appearing white matter and WMH volumes. Supratentorial intracranial volume was estimated by summing gray and white matter (including the sum of normal appearing white matter and WMH volumes) and CSF volumes.[38] Furthermore, T1-weighted scans were processed using FreeSurfer[140] (version 5.1) to obtain white matter and cortex volumes of the cerebellum.

Statistical analysis

Characteristics of the study population at time of MDCT scan are presented in means ± standard deviations or medians with interquartile ranges for (skewed) continuous

variables and in counts and percentages for categorical variables. We determined the number of repeated MRI scans available for analyses. Since the time between the intracranial calcification assessment and the first MRI was variable (a difference ranging from 0 to 12 years), the increase in calcification at the time of the MRI scans in this longitudinal study design was taken into account by creating age-adjusted percentiles of calcification volumes. These percentile curves describe the burden of calcification in comparison to the 'average' amount of calcification in an aging population. Using this calcification percentile as the measure of calcification, we assume that at the times that the follow-up MRI scans are performed, the increase in calcification in that participant is in accordance with the expected increase of calcification of the percentile curve. In other words, although the calcification is expected to increase with age, a participants' percentile remains stable over age. These percentiles were created using the following strategy. First, to assure standard normal distributed (natural log-transformed) calcification values at each age, we used natural log-transformed values of the calcification volume and participants with absence of calcification were not used for the estimation of the percentile curves. Second, ICAC percentile curves were fitted on 2,021 participants and VBAC percentile curves were fitted on 515 participants with presence of calcification (Online Figure 1) using the Lambda-Mu-Sigma (LMS) method.[226] In the Supplemental Materials further explanation of this method can be found. Third, using the percentile curve models which were fitted on participants with presence of calcification, subject specific percentile values of all 2,464 participants with a complete calcification assessment were determined for ICAC and VBAC. In this step, to deal with calcification values of zero, 1mm3 was added to the non-transformed values [Ln(calcification volume + 1.0mm3)]. Accordingly, the ICAC and VBAC percentile values were used as the main determinants in this study, and higher percentile values are interpreted as a larger burden of calcification.

We analyzed the relation between the percentile values, that is age-adjusted calcification burden, and brain structure changes. We used generalized estimating equations (GEE) for repeatedly measured categorical outcomes, and linear mixed models for repeatedly measured continuous outcomes. Both GEE and linear mixed models are statistical models that can deal with data from participants with only a single MRI, as well as participants that have multiple MRI scans. Accordingly, GEE models were used to assess the probability of having at least one microbleed or lacune. To account for possible nonlinear trajectories, exploratory analyses were performed to assess whether splines of age (with increasing degrees of freedom) improved the model compared with a linear age term. As a result of these analyses, splines of age with 1 knot were used in GEE analyses. An interaction of the calcification percentile and the spline-coefficients of age was included, to allow

slope differences in the relation of age and outcome explained by the burden of calcification. Other covariables in the model were intracranial volume, sex, the interaction sex and the spline-coefficients of age, to take potential slope differences between men and women into account. In addition, stratified GEE analyses were performed for the location of CMB (lobar, deep, infratentorial) and lacunes (supratentorial, infratentorial).

Analysis of WMH and the total number of enlarged PVS were assessed using linear mixed models with random intercepts and slopes. The total number of enlarged PVS was calculated by summing up the number of PVS in the midbrain, hippocampi, basal ganglia and centrum semiovale. WMH was natural log-transformed to account for the skewed distribution. In each linear mixed model, age of the participant at each MRI measurement was used as the time variable. Similar to the construction of the GEE to account for possible nonlinear trajectories, exploratory analyses were performed to assess whether splines of age (with increasing degrees of freedom) improved the model compared with the linear age term. As a result, splines of age with 2 knots were used in the linear mixed models. Similar to the GEE models, in the linear mixed models an interaction of the calcification percentile, the spline-coefficients of age, and the abovementioned covariables were included.

To assess the relation of ICAC and VBAC burden, as expressed in percentile values, with brain atrophy, we also used linear mixed models with random intercepts and slopes. The brain atrophy measures were global volumetric imaging markers including volumes of white matter, normal appearing white matter, gray matter; total cerebral volume; cerebellar white matter and cerebellar cortex volume. The linear mixed models were constructed similarly as described above.

Two sensitivity analysis were performed: (1) to assess the role of cardiovascular risk factors and extracranial carotid artery calcification on the relation between intracranial arteriosclerosis and structural brain changes in aging, we corrected for hypertension, type 2 diabetes mellitus, current smoking, obesity, hypercholesterolemia and extracranial carotid artery calcification volume (mm3) in all participants of whom information on these factors was available at the time of the first MRI scan (787 participants with in total 1544 MRI scans) and (2) to assess whether our statistical approach handles the time between CT and MRI sufficiently, we performed our analysis in only the participants who had their first MRI within 1 year after the CT (641 participants with in total 1567 MRI scans).

Statistical analyses were performed with R statistical software (R-project, Institute for Statistics and Mathematics, R Core Team (2013), version 3.4.1) using the "geeglm" and "lme" functions from the R-packages "geepack" and "nlme"[46,47].

Results

In total 1,489 participants had a CT as well as at least one brain MRI available [2,850 MRI scans in total were included in this study (Supplemental Figure 1)]. Out of 1,489 participants, 650 participants had a single brain MRI scan, 367 participants had two MRI scans, 422 participants had three MRI scans and 50 participants had four MRI scans available for analysis. Characteristics of the participants included in this study are presented in Table 1. The mean age at CT in this population was 67.9 years (range 59.0–91.9 years) and 52.0% were women. The average time difference between CT and first MRI was 5.4 years (range 0.06–12.2 years). In Supplemental Table 1, the brain structure characteristics of the study population are shown.

The ICAC and VBAC percentile curves are shown in Supplemental Figure 2. The ICAC percentile curves showed a larger calcification volume with increasing age, whereas for the VBAC percentile this increase in volume over age was less apparent.

Trajectories of focal markers and white matter hyperintensities

Table 1. Characteristics of the study population

Characteristic	Mean (SD), N = 1489
Age at CT, years	67.9 (5.7)
Sex, women	774 (52.0)
ICAC volume[a], mm^3	31 (98.8)
ICAC absent	313 (21.0)
ICAC present (> 0 mm^3)	1176 (79.0)
ICAC percentile	37.7 (32.5)
VBAC volume[b], mm^3	0 - 211.9
VBAC absent	1237 (83.1)
VBAC present (> 0 mm^3)	252 (16.9)
VBAC percentile	24.3 (16.4)
Availability of MRI scans with acceptable segmentation	
Number of participants with a single MRI scan	650 (43.7)
Number of participants with two MRI scans	367 (24.6)
Number of participants with three MRI scans	422 (28.3)
Number of participants with four MRI scans	50 (3.4)
Number of available MRI scans with acceptable segmentation	2850
Scan interval between MRI scans,years	3.3 (1.1)

Continuous variables are presented as means (standard deviations) and categorical variables as number (percentages).

Key: ICAC, intracranial carotid artery calcification; N: number of participants; VBAC: vertebrobasilar artery calcification.

[a]ICAC volume is presented as median (interquartile range).

[b]VBAC volume is presented as minimum and maximum value.

The estimated trajectories of focal markers in aging for the different calcification

Table 2. Overview of the quantitative measures of the trajectories of focal markers and white matter hyperintensity volume shown in Figure 1, with the addition of values for the percentile curves 25 and 75. The probability difference at age 60 and 80 represents the difference in probability between the trajectory of percentile 0 at that specific age, compared to the percentile curve. A negative difference means a lower volume compared to percentile 0. *p*-values below 0.05 were marked bold.

	Calcification	Percentile	Difference in probability at age 60	Difference in probability at age 80	*p*-value percentile term	*p*-value percentile*age interaction term
White matter hyperintensity volume*	ICAC	25	0.13	0.56	0.006	0.339
		50	0.27	1.17		
		75	0.41	1.83		
		100	0.57	2.54		
	VBAC	25	0.21	0.97	**0.021**	0.136
		50	0.44	2.06		
		75	0.69	3.29		
		100	0.96	4.69		
Microbleeds	ICAC	25	0.02	0.04	0.092	0.078
		50	0.05	0.08		
		75	0.08	0.12		
		100	0.11	0.17		
	VBAC	25	-0.02	0.06	0.435	**0.044**
		50	-0.04	0.12		
		75	-0.05	0.19		
		100	-0.06	0.26		
Lacunes	ICAC	25	0.01	0.01	**$8.8 * 10^{-5}$**	**0.005**
		50	0.02	0.03		
		75	0.04	0.05		
		100	0.10	0.06		
	VBAC	25	0.00	0.01	0.411	0.337
		50	0.01	0.03		
		75	0.02	0.04		
		100	0.03	0.06		
Enlarged perivascular spaces**	ICAC	25	0.3	-0.08	0.303	**0.032**
		50	0.59	-0.16		
		75	0.89	-0.24		
		100	1.19	-0.31		
	VBAC	25	0.86	0.70	0.126	0.579
		50	1.73	1.39		
		75	2.59	2.09		
		100	3.46	2.79		

* Difference at age 60 and difference at age 80 expressed in ml (back-transformed);

**Difference at age 60 and difference at age 80 expressed in number of enlarged perivascular spaces.

percentiles are depicted in Figure 1. In Table 2 the trajectories of microbleeds and lacunes were quantified by showing the probability difference at age 60 and 80, representing the difference in probability between the trajectory of percentile 0 at that specific age, compared to the percentile curve. For white matter hyperintensity volume, the difference at age 60 and difference at age 80 was expressed in ml (back-transformed). For enlarged perivascular spaces, the difference at age 60 and difference at age 80 was expressed in number of enlarged perivascular spaces. A larger burden of ICAC and VBAC, that is higher percentile values, resulted in either intercept differences, slope differences or both. Overall, trajectories of CSVD markers showed more WMH and a higher probability of microbleeds and lacunes with increasing burden of ICAC or VBAC. Increasing ICAC or VBAC burden significantly increased WMH volume, however the slope was not significantly affected by the burden of calcification. For microbleeds, the effect of a larger ICAC burden seemed to be rather constant, whereas a larger VBAC burden seemed to mainly influence the slope of the microbleeds probability over age.

Trajectories of location-specific microbleeds showed the same patterns for lobar, deep and infratentorial microbleeds with larger ICAC and VBAC burden (Supplemental Figure 3 and Supplemental Figure 4), where especially the probability of deep microbleeds is significantly affected by the main effect of ICAC and the interaction of VBAC and age ($p = 0.041$ and $p = 0.031$, Supplemental Table 2) .

For lacunes, ICAC significantly increased the probability of lacunes and significantly modified the slope, whereas larger burden of VBAC did not significantly affect the probability of lacunes. Trajectories of location-specific lacunes showed similar patterns with significant effects of ICAC for both the main effect and the interaction of ICAC and age on supratentorial lacunes. Only the main effect of ICAC was statistically significant for the infratentorial lacunes ($p = 0.021$). The main effect of VBAC was only significant in supratentorial lacunes ($p = 0.023$).

For the number of enlarged PVS, a larger burden of both ICAC and VBAC did not significantly affect the number of enlarged PVS, nonetheless ICAC did significantly modify the slope. Trajectories of location-specific enlarged PVS showed different patterns with the number of hippocampal PVS being the most affected by both ICAC and VBAC, and a significant increased number of PVS in the midbrain with significant modified slopes of ICAC percentiles.

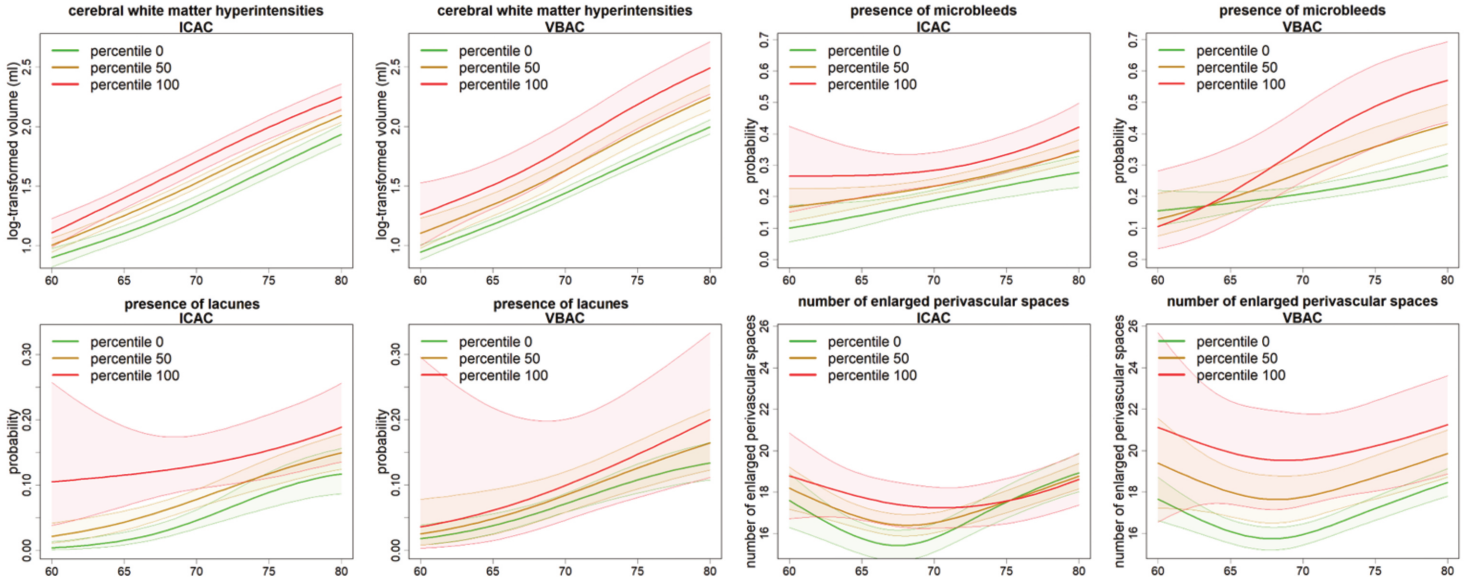


Figure 1. Trajectories of focal markers and white matter hyperintensity volume in aging for different ICAC and VBAC percentiles**.** The x-axis represents the age at time of the MRI scan, the y-axis represents the natural log-transformed volume in ml (estimated for a subject with an average intracranial volume) for the cerebral white matter hyperintensities, the probability of having one or more microbleeds and lacunes and the number of enlarged perivascular spaces.

Table 3. Overview of the quantitative measures of the global structural brain volume trajectories shown in Figure 2, with the addition of values for the percentile curves 25 and 75. The volume difference at age 60 and 80 represents the volume difference between the trajectory of percentile 0 at that specific age, compared to the percentile curve, expressed in milliliter (ml). A negative difference means a lower volume compared to percentile 0.

	Calcification	Percentile	Volume difference at age 60	Volume difference at age 80	p-value percentile term	p-value percentile*age interaction term
Total cerebral volume	ICAC	25	-1.72	-2.53	0.087	0.091
		50	-3.45	-5.05		
		75	-5.17	-7.58		
		100	-6.89	-10.11		
	VBAC	25	-2.73	-3.37	0.15	0.246
		50	-5.47	-6.74		
		75	-8.2	-10.11		
		100	-10.94	-13.48		
Cerebral gray matter volume	ICAC	25	-0.27	0.29	0.794	0.312
		50	-0.54	0.57		
		75	-0.81	0.86		
		100	-1.08	1.14		
	VBAC	25	-5.22	-1.03	0.057	0.167
		50	-10.45	-2.07		
		75	-15.67	-3.1		
		100	-20.9	-4.13		
Cerebral white matter volume	ICAC	25	-1.84	-2.8	0.223	0.563
		50	-3.69	-5.61		
		75	-5.53	-8.41		
		100	-7.38	-11.22		
	VBAC	25	0.52	-1.87	0.878	0.466
		50	1.04	-3.75		
		75	1.56	-5.62		
		100	2.08	-7.5		
Cerebral normal appearing white matter	ICAC	25	-2.23	-3.54	0.135	0.626
		50	-4.45	-7.08		
		75	-6.68	-10.61		
		100	-8.91	-14.15		
	VBAC	25	0.64	-2.63	0.872	0.374
		50	1.28	-5.26		
		75	1.92	-7.89		
		100	2.57	-10.52		
Cerebellar white matter volume	ICAC	25	-0.06	-0.02	0.162	0.155
		50	-0.13	-0.04		
		75	-0.19	-0.06		
		100	-0.25	-0.08		
	VBAC	25	0.1	0.01	0.321	0.248
		50	0.2	0.02		
		75	0.3	0.04		
		100	0.4	0.05		
Cerebellar cortex	ICAC	25	-0.1	-0.13	0.497	0.382
		50	-0.2	-0.26		
		75	-0.31	-0.38		
		100	-0.41	-0.51		
	VBAC	25	-0.47	-0.08	0.152	0.145
		50	-0.93	-0.15		
		75	-1.4	-0.23		
		100	-1.87	-0.31		

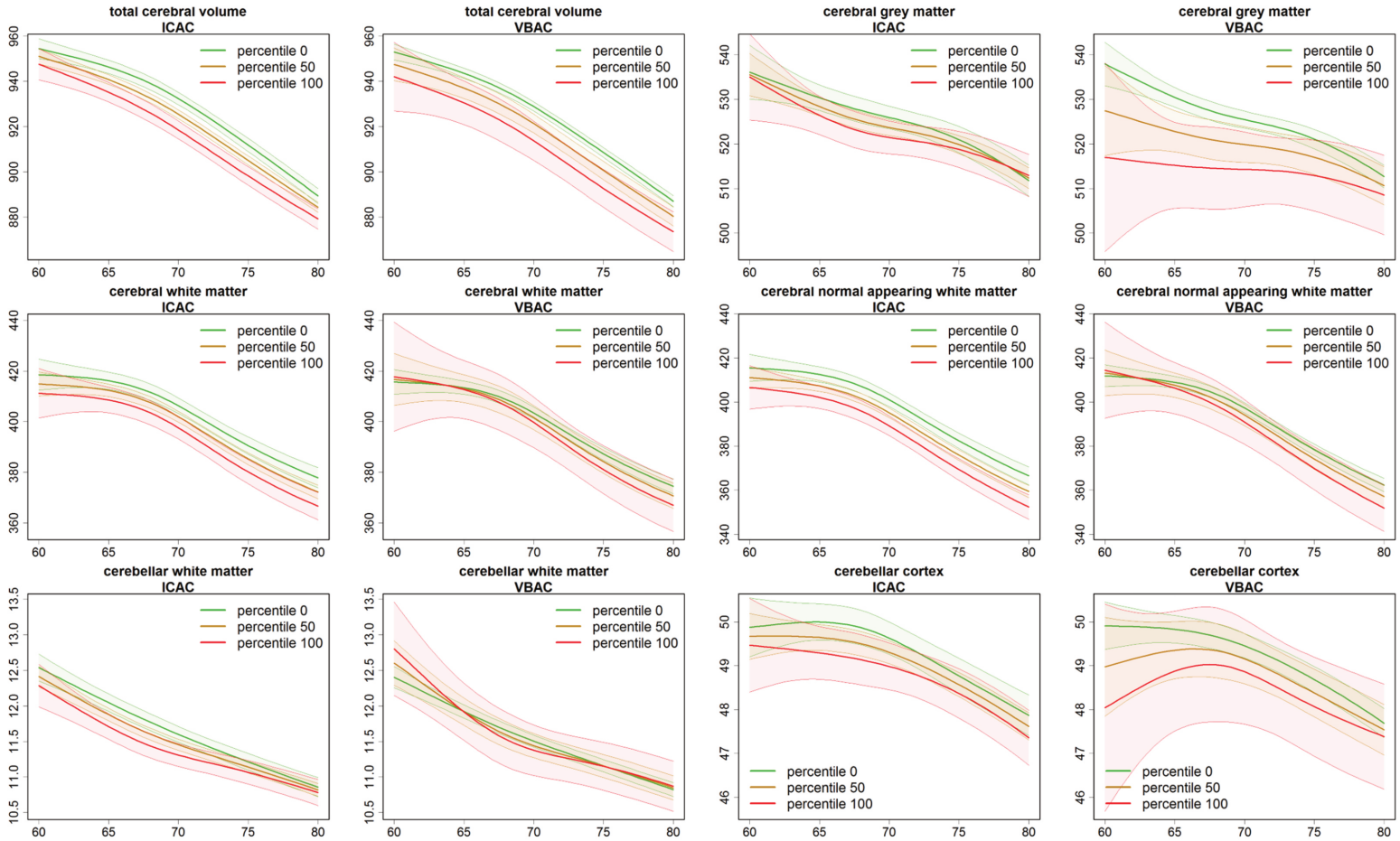


Figure 2. Trajectories of the global structural brain volumes in aging for different ICAC and VBAC percentiles**.** The x-axis represents the age at time of the MRI scan, the y-axis represents volume in ml, estimated for a subject with an average intracranial volume.

Trajectories of global volumetric markers

Trajectories of the global volumetric imaging markers in aging, for the different calcification percentiles are shown in Figure 2. Although the trajectories showed a pattern of smaller global volumes with larger burden of ICAC and VBAC, the main effect of the percentiles on the volumetric outcomes and the interaction between age and the percentiles were not statistically significant ($p > 0.05$). The effect of the calcification burden on the volumetric outcomes and the corresponding statistical significance level of the slope and interaction with age are presented in Table 3. To further quantify these effects, we showed the volume difference at age 60 and age 80 between the trajectory of percentile 0 at that specific age, compared to the percentile curve.

Sensitivity analyses

The associations between arteriosclerosis and structural brain changes in aging showed similar effects and trends when corrected for cardiovascular risk factors and extracranial carotid artery calcification, as without the correction for these factors. Supplemental Figure 5 and 6 show the trajectories of CSVD markers, including brain atrophy in aging for different arteriosclerosis percentiles when corrected for these factors. In Supplemental Table 3, an overview of the presence of cardiovascular risk factors within the study population is shown. Furthermore the results within participants with less than 1 year between the CT and first MRI were similar to the results with all participants included.

Discussion

Within this population-based study, we found that a larger burden of ICAC and VBAC, as proxy of intracranial arteriosclerosis, was associated with accelerating small vessel disease and overall smaller brain volumes over time, which includes accelerated brain atrophy in aging.

Of all brain measurements, a larger burden of ICAC was most strongly associated with WMH, deep microbleeds, supratentorial and infratentorial lacunes, hippocampal and midbrain PVS. For larger VBAC burden, we found significant relations with WMH, supratentorial lacunes, hippocampal PVS. Cross-sectionally, we previously reported that ICAC is related to larger WMH volume, and presence of lacunes, though no relation with microbleeds was found.[218] This could partially be explained by limited statistical power and the cross-sectional study design. Other studies reported associations between arteriosclerosis in the internal carotid artery and WMH and lacunes which are consistent with the findings in this study.[227] Other

studies also observed higher WMH burden, in population-based and patient cohorts that used different imaging techniques like ultrasound and MRI to visualize carotid atherosclerosis.[210–213,220]

To our knowledge, the current study is the first that showed the role of intracranial arteriosclerosis in the anterior- and posterior circulation on brain changes while capturing the concomitant aging process using longitudinal data. Muller et al. concluded that in patients with arteriosclerotic disease, only severe and bilateral carotid stenosis were related to progression of brain atrophy with 4 years of follow-up.[214] We have shown a pattern of a higher burden of arteriosclerosis being related to accelerated brain atrophy over time, however these effects were not statistically significant. Cross-sectional studies also found that markers of carotid arteriosclerosis were associated with smaller brain volumes.[219,220,228] In the Rotterdam study, we previously described this matter and have found significant results for smaller total brain and white matter volumes,[219] albeit we could not reproduce these findings in the present study. A possible explanation for the less pronounced finding in our study is that in the cross-sectional study, the effect of carotid arteriosclerosis on brain atrophy may be driven by age, simply because age was only corrected for linearly, whereas in this study we conclude that arteriosclerosis has a non-linear relation with age. Furthermore, with the longitudinal data we were able to account for the complex non-linear relations between age and arteriosclerosis as well as age and brain structure.

Given potential variation in genetic susceptibility to arteriosclerosis, and differences in anatomical structure between the vertebrobasilar and intracranial carotid arteries in terms of curvatures and branches, consequences of arteriosclerosis are likely to be artery-specific or at least – besides the influence of systemic arteriosclerosis - contain an artery-specific component[223,229]. Therein VBAC is commonly neglected in studies as an investigation site for arterial lesion.[230] Identifying pathophysiological mechanisms of intracranial arteriosclerosis at large could aid development of location-specific therapeutic and preventive implications[231]. Beyond etiology, the predictive value of intracranial calcification needs to be evaluated for its potential as a noninvasive marker to reflect the severity of intracranial vascular disease. Similar to the coronary artery calcium score predicting acute coronary disease,[232] such insight could benefit the identification of high risk individuals for neurological diseases.

The trajectories of brain structure changes in aging and how these are related to the burden of intracranial arteriosclerosis provide insight in the effect of arteriosclerosis over time. However, the current study does not address whether the relation between intracranial arteriosclerosis and the structural brain changes is causal. Although we

performed a sensitivity analysis that showed similar results when corrected for cardiovascular risk factors and extracranial carotid artery calcification at the time of the first MRI, we did not account for time-dependent confounding of these factors. Therefore, the associations found in this study could be partly driven by other factors, including changes in cardiovascular risk factors over time and progression of arteriosclerosis at large instead of solely the burden of intracranial arteriosclerosis at the time of the CT examination. However, identifying the potential effect of arteriosclerosis on brain structure changes is an important step forward that can motivate future studies to explore the significance of these associations.

Strengths

Strengths of this study are the large study population including extensive and longitudinal data on brain markers. Repeated measurements of brain MRI allowed us to take into account the effect of arteriosclerosis on brain changes, longitudinally. To thoroughly account for the effect of aging, which plays a pivotal role in development of arteriosclerosis, we applied age-adjusted percentile curves. Moreover, we performed sensitivity analyses to account for nonlinear trajectories with splines of age and interaction terms.

Limitations

Yet potential limitations also need to be addressed. First, to quantify intracranial arteriosclerosis, only calcified plaques were assessed since non-calcified plaques, also known as vulnerable soft plaques are impossible to visualize with CT. Nevertheless, arterial calcification has been recognized an adequate marker of the total plaque burden[233]. Second, although the average follow-up time was 5.4 years, which is a relatively long follow-up time within the field, a longer follow-up time would better allow to capture the aging effect on brain health, especially in terms of evaluating brain volumes. Third, as a subgroup of participants had multiple scans available for analysis a healthy survivor effect may have occurred, which could reduce the range of calcification in the vessels and thus attenuate true associations. Fourth, only longitudinal measurements for brain MRI were available but we had no follow-up CT measurements. With the use of the calcification percentiles we take into account the burden of calcification with respect to the normal aging process. The percentile curves describe the burden of calcification in comparison to the 'average' amount of calcification in an aging population. Using this calcification percentile of a study participant at a certain age as our measure of arteriosclerosis, we assume that at the times that the follow-up MRI scans are performed, the increase in arteriosclerosis in that participant is in accordance with the expected increase of the percentile curve. In other words, although the burden of arteriosclerosis is expected

to increase with age, the participants' percentile remains stable over age. Whether this assumption holds needs to be assessed in future research. Especially for VBAC it is uncertain whether this assumption holds, since the prevalence of VBAC in our population is rather low. Therefore the effect of the volume of VBAC may be less apparent while we might fail to recognize the value of VBAC presence, and incidence during follow-up of which the latter we have not been able to measure. Lastly, although a major strength of this study is that we addressed both the anterior cerebral circulation and the posterior anterior cerebral circulation, we cannot rule out the possibility that our findings may partly reflect the influence of systemic arteriosclerosis.. However, cross-sectional studies within the same study populations have shown that arteriosclerosis within the intracranial arteries in particular, as compared to other major vessel beds, was most strongly related to smaller white matter volume, and larger WMH volumes[218,219].

Conclusions

A larger burden intracranial arteriosclerosis is associated with accelerating small vessel disease over time. Identifying pathophysiological mechanisms of intracranial arteriosclerosis could help to lower the vascular burden in the brain and eventually prevent stroke, cognitive impairment and dementia.

3.2

Cardiovascular health in relation to trajectories of imaging markers in brain ageing

EJ Vinke, HA Vrooman, WJ Niessen, T Voortman, MA Ikram, MW Vernooij

Submitted

3.3

Hearing loss and cognitive decline in the general population: a prospective cohort study

PH Croll*, EJ Vinke*, NM Armstrong, S Licher, MW Vernooij, RJ Baatenburg de Jong, A Goedegebure, MA Ikram

Journal of Neurology. 2021;268(3):860-871

Previous studies identifying hearing loss as a promising modifiable risk factor for cognitive decline mostly adjusted for baseline age solely. As such a faster cognitive decline at a higher age, which is expected considering the non-linear relationship between cognition and age, may have been overlooked. Therefore it remains uncertain whether effects of hearing loss on cognitive decline extend beyond age-related declines of cognitive function. 3,590 non-demented participants were eligible for analysis at baseline, and a maximum of 837 participants were eligible for the longitudinal analysis. Hearing loss was defined at baseline. Cognitive function was measured at baseline and at follow-up (4.4 years [SD: 0.2]). Multivariable linear regression analysis was used for the cross-sectional analysis. Linear mixed models were used to assess the longitudinal association between hearing loss and cognitive decline over time while adjusting for confounders and the interaction of age and follow-up time. Hearing loss was associated with lower cognitive function at baseline. Moreover, hearing loss was associated with accelerated cognitive decline over time on a memory test. After additionally adjusting for the interaction between age and follow-up time, we found that hearing loss did not accelerate cognitive decline anymore. Hearing loss was associated with lower cognitive function at baseline and accelerated cognitive decline on a memory test. The association between hearing loss and accelerated cognitive decline was non-significant after additional adjustment for non-linear age effects. More evidence is needed to ensure the role of hearing loss as a modifiable risk factor for cognitive decline.

Introduction

Recently, hearing loss has been put forward as a promising modifiable risk factor for cognitive decline and dementia.[121,244–247] Both the prevalence of hearing loss and dementia will increase substantially due to the aging of the worldwide population.[121,248,249] With the increasing numbers of both conditions, it is of great importance to determine if hearing loss is independently associated with cognitive decline in dementia-free participants. As such, more can be said on whether hearing rehabilitative treatments may potentially alter or delay the progression of cognitive decline.

Several longitudinal studies reported associations between hearing loss and poorer cognitive function[250], and with an increased risk of dementia.[244,246,248,251–254] Despite these promising results, several methodological issues should be considered. First,

both hearing loss and cognitive impairment are heavily dependent on age, reflected in a steep increase of the prevalence of both with increasing age.[74,249] Therefore, it is of importance to adjust for both linear and non-linear age effects in the association between hearing loss and cognition. To our knowledge, only one other study incorporated age non-linearly in their models.[250] Yet, it is plausible that older people may decline faster over time on cognitive abilities compared to their younger counterparts. Keeping this age-related decline into account can be accomplished by adding an interaction between age and follow-up time into statistical models. Second, some studies rely on a limited battery of neuropsychological tests for cognitive assessment.[245,253,255–265] This potentially increases the likelihood of misclassification of cognitive impairment[266], especially in those with higher levels of hearing impairment. Lower scores on cognitive tests may partially be falsely attributed to cognitive impairment, as individuals might not be able to hear verbal test instructions properly.[267,268] Third, hearing loss does not necessarily accurately reflect an inability to follow speech in noisy environments.[269] To our knowledge, only one other study incorporated a measure of speech understanding in their analyses.[265]

Against this background, we aimed to elucidate whether hearing loss accelerates cognitive decline over time in dementia-free participants whom are at risk of cognitive decline and cognitive impairment from a large population-based study. We measured hearing loss, including speech understanding, and repeatedly assessed cognitive functioning with comprehensive cognitive testing. We examined whether trajectories of cognitive decline differed across degrees of hearing impairment while adjusting for potentially strong effects of age.

Methods

Study setting and population

This study is embedded in the Rotterdam Study, a prospective, population-based cohort study. The Rotterdam Study was initiated in 1989 and investigates determinants and consequences of aging. Details of the study have been described previously.[194] The entire study population consists of 14,926 individuals aged ≥ 45 years from the Ommoord area, a suburb of Rotterdam, the Netherlands, who undergo extensive examinations at the research center at study entry and subsequent visits every 3–4 years. In 2011, hearing assessment was introduced into the study protocol. For the present study, we sampled two study populations, described in detail below.

Hearing loss and cognitive function: cross-sectional study population

In total, 3739 participants underwent baseline hearing assessment (2011–2014). We excluded participants with probable conductive hearing loss (air–bone gap ≥ 15 dB; N = 83), participants with a history of dementia or those who were insufficiently screened for dementia at baseline (N = 51), and participants who developed dementia during follow-up (N = 15), leaving 3590 participants with baseline hearing assessment. From those 3590 participants, data were available on different cognitive tests, namely the MMSE (N = 3584), the Stroop test (N = 3500), the Word Fluency test (WFT) (N = 3536), the Letter Digit Substitution test (LDST) (N = 3507), the Word Learning test (WLT) (N = 3239), and the Purdue Pegboard test (PPT) (N = 3264). There were 3498 participants with both data on hearing thresholds and speech understanding in noise.

Hearing loss and cognitive decline: longitudinal study population

Data on the different cognitive tests from participants who were re-invited for follow-up measurements and with available cognitive data at baseline, were available at follow-up (2015–2016) for the longitudinal analysis. At follow-up, 837 participants had data available for the MMSE, 764 participants for the Stroop test, 519 participants for the WFT, 780 participants for the LDST, 755 participants for the WLT, and 714 participants for the PPT. The mean time interval between cognitive baseline assessment and re-examination was 4.4 years (SD: 0.2). See supplementary methods for an explanation regarding the attrition rate.

Hearing

Hearing thresholds measured with pure-tone audiometry

To determine hearing loss expressed by hearing thresholds in decibel (dB), pure-tone audiometry (PTA) was performed in a soundproof booth.[194] A computer-based audiometry system (Decos Technology Group, version 210.2.6, AudioNigma interface) and TDH-39 headphones were used. dB hearing levels were measured according to the ISO-standard 8253-1 (International Organization for Standardization, 2010). Air conduction (frequencies 0.25–8 kilohertz [kHz]) and bone conduction (0.5 and 4 kHz) were tested for both ears while masking according to the method of Hood.[270] The best hearing ear was determined by taking the average hearing thresholds over all frequencies and identified by the lowest hearing threshold of one of both ears. Of the best hearing ear, we determined the average speech frequencies threshold (average of 0.5, 1, 2, and 4 kHz) levels.

Finally, we determined degrees of hearing loss: normal hearing (0–20 dB), mild hearing loss (20–35 dB), moderate hearing loss (35–50 dB), and severe hearing loss (≥ 50 dB).[194,271]

Speech understanding in noise measured with the digits-in-noise test

To measure speech understanding in noise, we derived a signal-to-noise ratio (SNR; in dB) from the digits-in-noise (DIN) test, a 3-min test of speech understanding in noise.[272] Both speech and noise signal were presented in the participant's better hearing ear. Pre-recorded male-spoken speech-signal consisted of 24 digit triplets. Initially, the triplet was presented at 0 dB SNR. In case of an incorrect response, the next triplet was presented more intensely. After the first correct response, the speech level was decreased and a new stimulus was presented. For a correct response, the intensity was decreased again, while an incorrect response lead to an increase of the intensity. This was repeated until 24 triplets were repeated. SNR was the average of the last 20 triplets. We defined hearing categories based on optimal SNR cut points defined by clinically relevant degree of hearing loss using Youden's Index (Supplementary Figure 1).[273]

Cognition

Cognitive function was assessed in detail with an extensive neuropsychological test battery comprising the MMSE, the Stroop test (adjusted interference score; inverted as higher scores indicate worse performance), the WFT (amount of animals named within 60 s), the LDST (number of correct digits within 60 s), the 15-WLT (total number of words remembered at least 10 min after immediate recall), and the PPT (sum score of three trials). The MMSE was administered during a home visit, the other tests were administered at the research center. All tests instructions were presented verbally. The MMSE is a validated screening tool for cognitive decline and cognitive impairment.[195] The Stroop test is a validated test measuring executive functioning, more specifically it measures the ability to inhibit cognitive interference.[274] To accurately and reliably measure verbal fluency, the WFT was used[275]. With the validated LDST, we measured executive functioning including processing speed and attention.[90] The 15-WLT is a validated test measuring memory functioning.[88] Results of the WLT are not negatively influenced by hearing status, as the 15 different words are visually presented to the participants. The PPT is a validated measure of unilateral and bilateral fine manual dexterity.[95]

Covariates

During home interviews, educational level was assessed and categorized as primary education, lower education, intermediate vocational education and higher education. Smoking habits were assessed during the same interview and subsequently classified into never, former and current smoking.[194] Alcohol consumption was assessed through self-report with the food-frequency questionnaire[236], and we subsequently calculated daily alcohol consumption in grams.[236] Systolic and diastolic blood pressures were measured twice on the right arm with a random-zero sphygmomanometer; the mean of these readings was used for the analyses. Use of antihypertensive medication was assessed by interview.[194] Participants were screened for dementia at baseline and follow-up examinations using a protocol described in detail elsewhere.[197]

Statistics

We investigated whether baseline characteristics differed between participants with just a baseline assessment and participants with both a baseline and a follow-up assessment using T tests, $\chi 2$ tests, and Mann–Whitney U Tests when appropriate. Subsequently, we present three sequential analyses to examine the association between hearing loss and cognition. First, we assessed the cross-sectional association between hearing loss (all frequencies, speech frequencies, degrees of hearing loss and SNR) and cognitive functioning at baseline using multivariable linear regression models. We adjusted for age, age^2, sex, education, alcohol consumption, smoking behavior, systolic- and diastolic blood pressure, and use of blood pressure lowering medication. All SNR analyses were additionally adjusted by PTA hearing levels for all frequencies.

Second, we used linear mixed models with random intercepts and slopes to elucidate the longitudinal association between degrees of hearing loss (mild, moderate or severe compared to normal hearing defined by either PTA or SNR) and cognitive trajectories per test. In each model, we entered follow-up time in years after baseline measurement to use as time variable. For adjustment, we used the same models as described above. In a second model, a two-way interaction between age and follow-up time was added to account for possible slope differences for cognition over time, depending on the baseline age. All SNR analyses were additionally adjusted by PTA hearing thresholds. Next to the linear effects of hearing loss on cognition, an interaction of hearing loss and follow-up time was incorporated in all models, to allow slope differences in the relationship between cognitive functioning and time explained by degree of hearing loss. The linear hearing loss term (intercept difference) and the interaction term between hearing loss and follow-up time (slope

difference) are the main terms of interest in this longitudinal analysis. For SNR analysis, random slopes were not included as the models failed to converge.

Third, we performed similar linear mixed models to study the longitudinal association between hearing levels (all frequencies, speech frequencies, and SNR) and cognitive trajectories per test.

In sensitivity analyses, we explored whether longitudinal associations between hearing levels and cognitive trajectories differed between men and women and between mid-life (51–70 years) compared to late life (70–99 years). Originally, the MMSE was designed as a cognitive screening tool and is therefore limited in its capability to truly measure global cognitive functioning.[195] In an additional sensitivity analyses, we created a global cognition score, a g factor, by z transforming and averaging performance across each of the tests (except for the MMSE). Results were non-significant and effect estimates were smaller than those for the MMSE, indicating that the g factor in this sample cannot be considered as a more sensitive marker of global cognition than the MMSE. To facilitate interpretability and comparability (previous studies often used the MMSE), we chose to show the results for the MMSE and omit results regarding the g-factor from the final manuscript.

IBM SPSS Statistics version 25 (International Business Machines Corporation, Armonk, New York) and RStudio; integrated development environment for R, version 3.5.1. (RStudio, Boston, Massachusetts) were used for statistical analyses. Analyses with linear mixed models were done using the "lme" function of the R-package "nlme".[46] A p value < 0.05 was considered statistically significant.

Results

Table 1 shows the baseline characteristics of the study population. Mean age was 65.2 years (SD: 7.3). 56.2% of our population were female. Participants had a mean all frequency hearing threshold of 22.8 dB (SD: 11.1). 44.6% of the population had normal hearing threshold levels. For speech understanding in noise, mean SNR was − 4.06 dB (SD: 4.2). Participants with a follow-up assessment compared to participants with only a baseline assessment were significantly older, had a lower alcohol intake and were unhealthier (Supplementary Table 1).

Cross-sectional results

Table 2 shows the cross-sectional association between hearing loss and cognitive function. Elevated hearing thresholds and diminished speech in noise understanding were associated with lower scores on all cognitive tests, and appeared to be most pronounced for participants with severe hearing loss as compared to normal hearing on the Stroop test, WFT, LDST and the PPT (Table 2).

Longitudinal results

In the first model, mild hearing loss showed statistically significant intercept differences, compared to normal hearing thresholds on the WFT, LDST, and the PPT (Table 3). In line with this, mild and moderate degrees of hearing loss, showed intercept differences for all cognitive tests, though not statistically significant (Table 3; model 1). Longitudinally, moderate hearing loss as compared to normal hearing thresholds modified the slope of memory functioning as measured with the 15-WLT significantly over time. For the other cognitive tests no significant slope differences were identified (Table 4; model 1). No significant slope differences were found for any hearing loss, as compared to normal speech understanding in noise (Table 4, model 1). The significant slope difference of the 15-WLT became statistically non-significant, and slope differences of other cognitive tests became small or close to zero (Table 3; model 2; Figure 1) after additional adjustment for the interaction between age and follow-up time. Comparable results were found for degrees of hearing loss as measured with the DIN test (Table 4; model 2).

Moreover, assessing hearing levels continuously showed that the additional change in cognitive functioning attributable to hearing loss were small and non-significant for both hearing thresholds and speech understanding in noise (Supplementary Table 2). Results did not differ between males and females or between midlife and late-life (Supplementary Tables 3 and 4).

Table 1. Baseline characteristics

Baseline characteristics	***N* = 3590**
Age [years (SD)]	65.2 (7.3)
Age (range)	51.5–98.6
Female [*N* (%)]	2016 (56.2)
Educational level [*N* (%)]	
Primary	264 (7.4)
Lower	1330 (37.0)
Intermediate vocational	1049 (29.2)
Higher	925 (25.8)
Alcohol consumption[a], gram (IQR)	7.9 (1.4–19.1)
Smoking [*N* (%)]	
Never	1134 (33.5)
Past	1828 (50.9)
Current	611 (17.0)
Systolic blood pressure, mmHg (SD)	139.5 (21.0)
Diastolic blood pressure, mmHg (SD)	83.1 (11.2)
Use of blood pressure lowering medication [*N* (%)]	1449 (40.4)
Hearing thresholds measured with pure-tone audiometry	
All frequency hearing loss [dB (SD)]	20.8 (9.7)
Speech frequency hearing loss [dB (SD)]	20.0 (10.7)
Degree of hearing loss [*N* (%)]	
Normal (0–20 dB)	1601 (44.6)
Mild (20–35 dB)	1456 (40.6)
Moderate (35–50 dB)	425 (11.8)
Severe (50 dB)	79 (2.2)
Speech understanding in noise measured with the digits-in-noise test	*N* = 3498
Signal-to-noise ratio [dB (SD)]	− 4.06 (4.2)
Degree of hearing loss [*N* (%)]	
Normal (0–20 dB)	1662 (46.3)
Mild (20–35 dB)	837 (23.3)
Moderate/severe (35–50 dB)	1,091 (30.4)
Cognitive abilities	
Mini-Mental State Examination score[a] (IQR)	29.0 (27.0–29.0)
Stroop Test interference score[a] (IQR)	44.5 (37.9–54.1)
Word Fluency Test score[a] (IQR)	23.0 (19.0–27.0)
Letter Digit Substitution Test score[a] (IQR)	30.0 (26.0–35.0)
Word Learning Test delayed recall score[a] (IQR)	8.0 (6.0–10.0)
Purdue Pegboard Test sum score[a] (IQR)	36.0 (33.0–39.0)

Values are mean (standard deviation [SD]) for continuous variables or a median (interquartile range [IQR]) for non-normally distributed continuous variables and percentages for categorical variables. The amount of hearing loss is expressed in dB, i.e. a higher dB value reflects more hearing loss. Abbreviations: dB, decibel; mmHg, millimetres of mercury.

Table 2. Effect estimates of hearing loss and cognitive function based on the cross-sectional analysis.

Hearing loss	Mini− mental state examination score	Stroop test interference score	Word fluency test score	Letter digit substitution test score	Word learning test delayed recall	Purdue pegboard test sum score
	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)
Hearing loss measured with pure-tone audiometry						
Hearing thresholds per 10 dB increase						
All frequencies	−0.04 (−0.14, 0.06)	−0.63 (−1.31, 0.04)	**−0.42 (−0.65, −0.20)**	**−0.38 (−0.62, −0.14)**	**−0.11 (−0.23, −0.00)**	**−0.33 (− 0.52, − 0.14)**
Speech frequencies	0.01 (−0.09, 0.11)	−0.49 (−1.15, 0.18)	**−0.37 (−0.59, −0.15)**	**− 0.27 (− 0.51, − 0.03)**	− 0.10 (− 0.21, 0.01)	**−0.28 (−0.47, − 0.09)**
Degree of hearing loss						
Normal (0-20dB)	Reference	Reference	Reference	Reference	Reference	Reference
Mild (20-35 dB)	− 0.07 (− 0.25, 0.12)	−0.75 (−2.18, 0.67)	**−1.02 (−1.50, −0.55)**	−0.42 (−0.93, 0.09)	− 0.20 (−0.44, 0.03)	**−0.52 (−0.92, −0.13)**
Moderate (35-50 dB)	− 0.10 (−0.44, 0.23)	− 1.84 (−4.05, 0.37)	**− 0.77 (−1.50, −0.03)**	−0.66 (− 1.45, 0.12)	−0.31 (−0.68, 0.05)	**−0.83 (−1.45, −0.21)**
Severe (≥ 50 dB)	**− 0.98 (−1.94, −0.02)**	0.02 (−4.58, 4.61)	**− 1.88 (−3.40, −0.37)**	**−1.91 (−3.54, −0.28)**	−0.59 (−1.35, 0.16)	**− 1.38 (−2.67, −0.09)**
Hearing loss measured with the digits-in-noise test						
Speech understanding in noise per 1 dB increase						
Speech reception threshold	**−0.07 (−0.10, −0.04)**	**−0.59 (−0.96, −0.23)**	−0.03 (−0.12, 0.05)	**−0.19 (−0.30, −0.09)**	**−0.07 (−0.12, −0.02)**	−0.03 (−0.06, 0.00)
Degree of hearing loss[a]						
Normal (≤ -5.55 dB)	Reference	Reference	Reference	Reference	Reference	Reference
Mild (-5.55 to -3.8 dB)	−0.14 (−0.37, 0.09)	−2,08 (−4.75, 0.59)	**−0.80 (−1.45, −0.14)**	**−0.85 (−1.59, −0.11)**	**−0.41 (−0.77, −0.05)**	0.01 (−0.19, 0.22)
Moderate/severe (> - 3.8 dB)	**−0.36 (−0.64, −0.08)**	**−6.19 (−9.38, −2.99)**	**−0.92 (−1.71, −0.14)**	**−1.56 (−2.45, −0.66)**	**−0.53 (−0.96, −0.09)**	**−0.29 (−0.54, −0.04)**

Difference: represents the difference in cognitive score per 10 dB increase in hearing acuity or the difference in cognitive score per 1 dB increase in speech understanding in noise or the difference in degree of hearing loss (both hearing acuity (PTA) and speech understanding (DIN)) as compared to normal hearing. All frequencies: 0.25, 0.50, 1, 2, 4, and 8 kHz. Speech frequencies: 0.5, 1, 2, and 4 kHz. The amount of hearing loss is expressed in dB, i.e. a higher dB value reflects more hearing loss

CI confidence interval, dB decibel

[a]Defined by digits-in-noise score cut-offs. Adjusted for age, age^2, sex, education, alcohol consumption, smoking, diastolic and systolic blood pressure, and use of blood pressure lowering medication. Analyses using speech understanding were further adjusted for hearing acuity

Statistically significant effect estimates ($p < 0.05$) are indicated in bold

Table 3. Effect estimates of the degree of hearing loss and cognitive function based on the longitudinal analysis (intercept difference)

Degree of hearing loss		Mini-mental state examination score	Stroop test interference score	Word fluency test score	Letter digit substitution test score	Word learning test delayed recall	Purdue pegboard test sum score
		Difference (95% CI)	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)	Difference (95% CI)
Degrees of hearing loss as measured with pure-tone audiometry							
Normal (0–20 dB)	Model 1	Reference	Reference	Reference	Reference	Reference	Reference
Mild (20–35 dB)	Model 1	−0.08 (−0.38, 0.23)	0.11 (−3.45, 3.67)	**−1.11 (−2.19, −0.04)**	− 0.64 (− 1.79, 0.51)	− 0.18 (−0.71, 0.36)	**−1.01 (−1.84, −0.17)**
Moderate (35–50 dB)	Model 1	−0.14 (−0.50, 0.21)	− 1.76 (− 5.91, 2.39)	− 1.09 (− 2.34, 0.16)	− 1.07 (− 2.41, 0.27)	− 0.37 (− 0.99, 0.26)	−0.80 (−1.77, 0.17)
Severe (≥50 dB)	Model 1	−0.33 (−0.98, 0.32)	− 1.70 (− 9.39, 6.00)	− 2.00 (− 4.33, 0.33)	− 2.42 (− 4.92, 0.07)	− 0.26 (− 1.41, 0.86)	−0.89 (−2.79, 1.00)
Normal (0–20 dB)	Model 2	Reference	Reference	Reference	Reference	Reference	Reference
Mild (20–35 dB)	Model 2	−0.09 (−0.39, 0.21)	0.12 (−3.43, 3.68)	**−1.17 (−2.24, −0.09)**	− 0.69 (−1.84, 0.46)	− 0.22 (− 0.75, 0.32)	**−1.03 (−1.87, −0.20)**
Moderate (35–50 dB)	Model 2	−0.17 (−0.53, 0.18)	− 1.73 (− 5.88, 2.42)	− 1.23 (− 2.48, 0.03)	− 1.19 (− 2.54, 0.15)	− 0.47 (− 1.10, 0.15)	−0.87 (− 1.84, 0.11)
Severe (≥50 dB)	Model 2	−0.39 (−1.04, 0.27)	− 1.72 (− 9.41, 5.96)	− 2.24 (− 4.57, 0.10)	**−2.65 (−5.16, −0.14)**	−0.46 (−1.61, 0.69)	−1.04 (−2.95, 0.86)
Degrees of hearing loss as measured with digits-in-noise test							
Normal (≤ − 5.55 dB)	Model 1	Reference	Reference	Reference	Reference	Reference	Reference
Mild (− 5.55 to − 3.80 dB)	Model 1	−0.04 (−0.32, 0.25)	− 1.27 (−4.53, 1.98)	**−0.88 (−1.69, −0.07)**	**−0.92 (−1.83, −0.01)**	**−0.59 (−1.03, −0.16)**	0.05 (−0.20, 0.30)
Moderate/severe (>−3.80 dB)	Model 1	−0.27 (−0.60, 0.07)	**−5.30 (−9.18, −1.42)**	**−1.07 (−2.04, −0.11)**	**−1.50 (−2.60, −0.40)**	**−0.73 (−1.25, −0.21)**	**−0.31 (−0.60, −0.02)**
Normal (≤ − 5.55 dB)	Model 2	Reference	Reference	Reference	Reference	Reference	Reference
Mild (− 5.55 to − 3.80 dB)	Model 2	−0.05 (−0.34, 0.23)	−1.63 (−4.87, 1.62)	**−0.92 (−1.73, −0.11)**	**−0.95 (−1.86, −0.03)**	**−0.62 (−1.05, −0.18)**	0.04 (− 0.21, 0.29)
Moderate/severe (>−3.80 dB)	Model 2	−0.03 (−0.64, 0.04)	**−5.96 (−9.83,−2.08)**	**−1.15 (−2.12, −0.19)**	**−1.55 (−2.65, −0.45)**	**− 0.77 (− 1.29, − 0.25)**	**− 0.33 (− 0.62, − 0.03)**

Difference: represents the intercept difference in cognitive score per degree hearing loss (both hearing threshold as measured with pure-tone audiometry and speech understanding in noise as measured with the digits-in-noise test) as compared to normal hearing

CI confidence interval, dB decibel. Model 1 adjusted for age, sex, education, alcohol consumption, smoking, diastolic and systolic blood pressure, and use of blood pressure lowering medication. Model 2 additionally adjusted for the interaction between age and follow-up time. Analyses using speech understanding were further adjusted for hearing thresholds as measured with pure-tone audiometry

Statistically significant effect estimates ($p < 0.05$) are indicated in bold

Table 4. The additional change in cognitive score per year attributed to different degrees of hearing loss based on the longitudinal analysis (slope differences)

Degree of hearing loss		Mini-mental state examination score Difference (95% CI)	Stroop test interference score Difference (95% CI)	Word fluency test score Difference (95% CI)	Letter digit substitution test score Difference (95% CI)	Word learning test delayed recall Difference (95% CI)	Purdue pegboard test sum score Difference (95% CI)
Degrees of hearing loss as measured with pure-tone audiometry							
Normal (0–20 dB)	Model 1	Reference	Reference	Reference	Reference	Reference	Reference
Mild (20–35 dB)	Model 1	−0.01 (−0.10, 0.07)	−0.13 (−1.29, 0.43)	−0.00 (−0.24, 0.23)	0.01 (−0.18, 0.21)	− 0.09 (−0.20, 0.03)	0.02 (−0.16, 0.19)
Moderate (35–50 dB)	Model 1	−0.04 (−0.14, 0.06)	−0.47 (−1.47, 0.52)	−0.05 (−0.35, 0.24)	0.00 (−0.22, 0.23)	**−0.17 (−0.30, −0.03)**	− 0.01 (−0.22, 0.19)
Severe (≥50 dB)	Model 1	−0.05 (−0.23, 0.13)	−1.39 (−3.20, 0.43)	−0.19 (−0.96, 0.58)	0.30 (−0.14, 0.73)	− 0.03 (−0.27, 0.22)	− 0.15 (−0.54, 0.24)
Normal (0–20 dB)	Model 2	Reference	Reference	Reference	Reference	Reference	Reference
Mild (20–35 dB)	Model 2	− 0.00 (−0.09, 0.08)	− 0.23 (−1.09, 0.63)	0.04 (−0.20, 0.27)	0.04 (−0.16, 0.24)	− 0.06 (−0.18, 0.06)	0.03 (−0.14, 0.21)
Moderate (35–50 dB)	Model 2	− 0.02 (−0.12, 0.08)	− 0.03 (−1.05, 0.98)	0.03 (−0.28, 0.33)	0.06 (−0.17, 0.30)	− 0.11 (−0.25, 0.03)	0.03 (−0.18, 0.24)
Severe (≥50 dB)	Model 2	− 0.01 (−0.19, 0.18)	− 0.47 (−2.35, 1.41)	−0.02 (−0.80, 0.75)	0.42 (−0.03, 0.88)	0.09 (−0.17, 0.35)	−0.05 (−0.46, 0.36)
Degrees of hearing loss as measured with the digits-in-noise test							
Normal (≤ − 5.55 dB)	Model 1	Reference	Reference	Reference	Reference	Reference	Reference
Mild (− 5.55 to − 3.80 dB)	Model 1	−0.03 (−0.12, 0.05)	−0.15 (−1.01, 0.71)	0.06 (−0.14, 0.25)	0.04 (−0.15, 0.24)	0.07 (−0.04, 0.18)	−0.04 (−0.11, 0.03)
Moderate/severe (> − 3.80 dB)	Model 1	−0.07 (−0.15, 0.00)	−0.45 (−1.20, 0.30)	0.04 (−0.14, 0.21)	− 0.04 (−0.21, 0.13)	0.04 (−0.06, 0.13)	−0.02 (−0.09, 0.04)
Normal (≤− 5.55 dB)	Model 2	Reference	Reference	Reference	Reference	Reference	Reference
Mild (− 5.55 to − 3.80 dB)	Model 2	−0.03 (−0.11, 0.06)	0.06 (−0.79, 0.91)	0.08 (−0.12, 0.27)	0.06 (−0.14, 0.25)	0.08 (−0.03, 0.19)	− 0.03 (−0.11, 0.04)
Moderate/severe (> − 3.80 dB)	Model 2	−0.04 (−0.12, 0.03)	0.10 (−0.66, 0.87)	0.10 (−0.08, 0.28)	0.00 (−0.17, 0.18)	0.07 (−0.03, 0.17)	− 0.01 (−0.08, 0.05)

Difference: represents the additional change in cognitive score per year increase in follow-up time per degree hearing loss (both hearing threshold as measured with pure-tone audiometry and speech understanding in noise as measured with the digits-in-noise test) as compared to normal hearing

CI confidence interval. dB: decibel. Model 1 adjusted for age, sex, education, alcohol consumption, smoking, diastolic and systolic blood pressure, and use of blood pressure lowering medication. Model 2 additionally adjusted for the interaction between age and follow-up time. Analyses using speech understanding were further adjusted for hearing thresholds as measured with pure-tone audiometry

Statistically significant effect estimates ($p < 0.05$) are indicated in bold

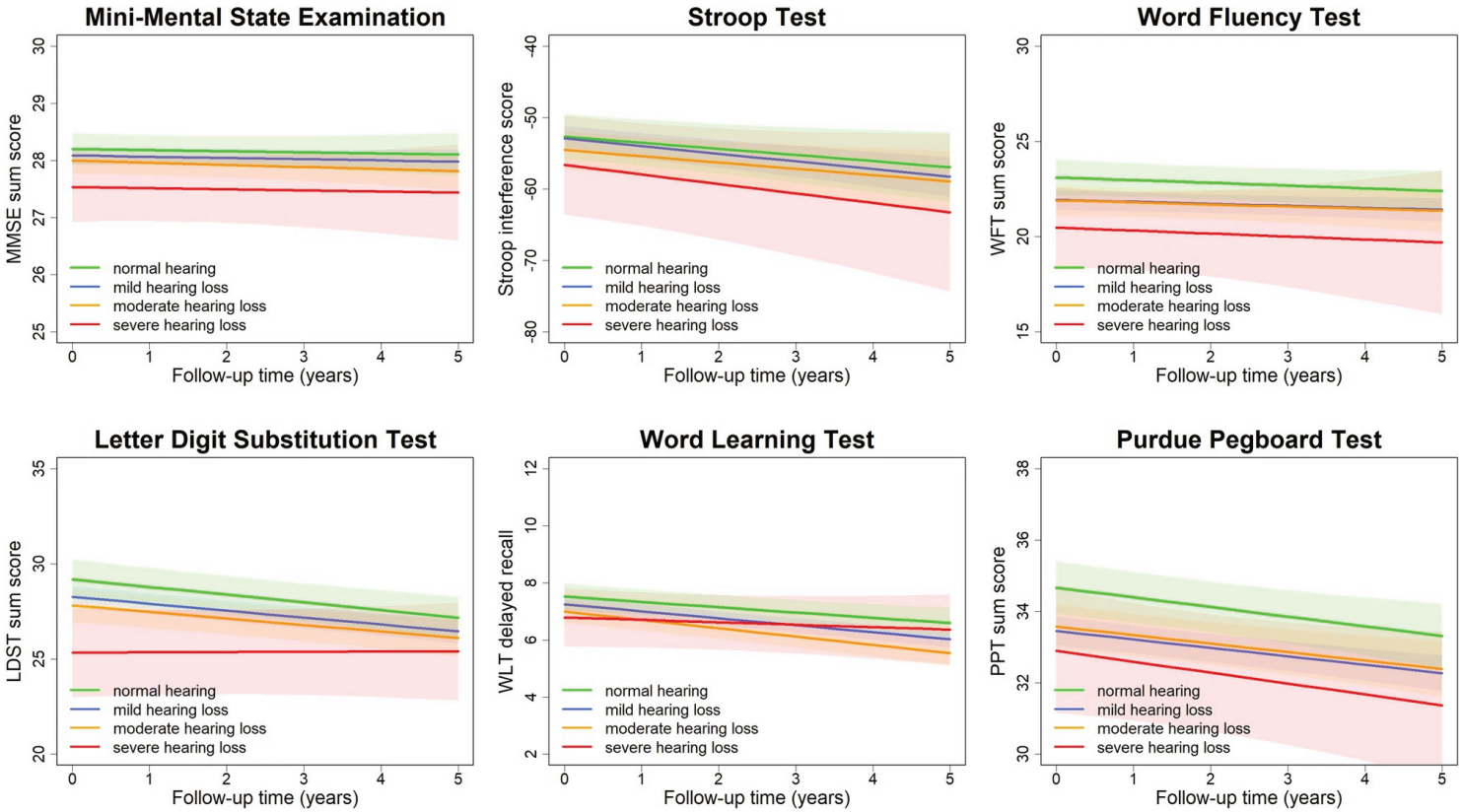


Figure 1. Estimated cognitive function trajectories over time for different degrees of hearing loss as measured with pure-tone audiometry, with corresponding 95% confidence intervals: adjusted for age and sex.

Discussion

In this large population-based study in non-demented older adults at risk for cognitive decline and cognitive impairment, we found that hearing loss was associated with poorer global cognitive functioning, executive functioning, verbal fluency, attention, memory, and manual dexterity. After adjustment for the possible non-linear effects of age on cognitive change during follow-up, we did not find that hearing loss for either hearing thresholds or speech understanding in noise accelerates cognitive decline over time.

Strengths of this study are its prospective and longitudinal population-based design, the large sample size and the standardized assessment of hearing thresholds with pure-tone audiometry and a speech-in-noise test as well as cognitive functioning with comprehensive cognitive testing. However, the following limitations of this study must be considered. First, although we extensively adjusted for age and other important confounders, residual confounding might still be present. For example, frailty and psychosocial well-being may confound our results as those are known to be highly related to age-related hearing loss.[276,277] Second, dementia incidence of participants with a baseline hearing assessment is small (N = 15), precluding the possibility to analyze whether hearing loss is associated with an increased risk of dementia in this sample.

Our cross-sectional results were comparable with other studies, reflected in lower scores on cognitive tests with higher levels of hearing loss.[245,255,278] It is of great interest that based on our results hearing loss (both peripheral and central) seems to affect executive functioning, verbal fluency, memory, manual dexterity and to some extent global cognitive functioning. Previous studies have argued that hearing loss leads to an increased cognitive load, shifting cognitive capacities towards sensory impairments rather than cognition.[254] Therefore, processes such as attention, memory, executive functioning, inhibitory control and verbal fluency may be compromised as a result of hearing loss. Moreover, cognitive decline, and especially diminished executive functioning, has repeatedly been linked to an increased risk of general frailty.[279] Even though we cannot infer on causality in this cross-sectional analysis, it does shed important light on a general risk of frailty in elderly with hearing impairment and comorbid compromised cognitive functioning. This underlines the great importance of investigating whether timely rehabilitative hearing treatment may alter or delay cognitive decline and possibly lowering the risk of full-blown dementia and/or general frailty.[277,280]

In our longitudinal analysis we found an accelerated decline in memory function (as measured with the 15-WLT) with moderate hearing loss, which is comparable to the

results and effect estimates of other population-based studies.[250,253] Unexpectedly, we did not find such an association for participants with severe levels of hearing loss, which could be explained by its relatively low prevalence (2.2%). Importantly, with further adjustment for confounding by age, the association between hearing loss and memory function became weaker and statistically non-significant. The prevalence of both hearing- and cognitive impairment increases substantially with age.[121,249] Moreover, it is also important to consider, especially in longitudinal studies with a wider age range that older individuals may decline faster on cognitive test performance between baseline and follow-up measurement than their younger counterparts.[267] Therefore, we added the interaction between baseline age and follow-up time into our statistical models, which seemed to explain most of the effects of hearing loss on memory function as the slope difference becomes statistically non-significant in the second model. To our knowledge, only one other study incorporated non-linear effects of age in their statistical model.[250] Therefore, verification in future studies is needed to explore whether effects of hearing loss on cognitive decline extend beyond 'normal' age-related decline of cognitive function.

Besides elevated hearing thresholds, speech understanding in noise could contribute towards accelerated cognitive decline. The ability to understand speech in noise is a complex process in which elements of peripheral processing interact with more centrally located elements of auditory processing.[272] As such, it may be hypothesized that a potential association with cognitive functioning may even be stronger when specifically speech understanding is reduced. Interestingly, we found the same (non-significant) results between speech understanding in noise and cognitive decline. This may be explained by the fact that there is a high correlation between hearing loss based on audiometry and speech in noise results in our population.[272]

It is also worthwhile considering whether found associations in our and previous studies might be driven by confounding and/or bias. The absence of an effect of hearing loss on cognitive decline in the current study is not explained by selection bias, as the sample with both a baseline- and a follow-up measurement were significantly older than the participants with just a baseline measurement. Moreover, it has been proposed that upstream common causes, i.e., inflammation, vascular pathology, and other systemic neurodegenerative processes, may lead to both hearing loss and cognitive decline through central nervous system-wide functional decline, rather than that those two are directly related to one another.[267] As such, greater sensitivity in one domain could identify impairments in that domain prior to the other, leading to the appearance of a false direct association.[267,268]

We should also acknowledge that our follow-up time (mean = 4.4 years) may have been too short to capture a possible small, but significant effect of hearing loss on

cognition. Epidemiological evidence has shown that elevated blood pressure in mid-life, an established modifiable risk factor of dementia, increases the risk of cognitive impairment 20–30 years later.[281–284] In contrast, another study with a follow-up of 8 years did not find an association between hypertension and cognitive functioning.[285] The differences in these results suggest that the follow-up time would need to be longer to show a potential association of hearing loss with cognitive decline. Nevertheless, despite the relatively short follow-up time, we do find an effect of mild hearing loss on memory functioning in the first model which is both statistically significant as well as clinically relevant.[244,250] Would our follow-up time truly been too short to capture an effect of hearing loss on cognitive function, we would have expected non-significant results.

In conclusion, hearing loss was significantly associated with compromised cognitive function and with accelerated decline on the 15-WLT measuring memory function. Notably, the latter association seemed to be driven by non-linear effects of age. Future, population-based studies are needed to further confirm the role of hearing loss as a potential modifiable risk factor for cognitive decline, whilst paying attention to effects of age on cognition. Even though more research is needed to strengthen evidence between hearing loss and accelerated neurodegeneration, our results do underline the great importance to acknowledge the effects of hearing loss (whether it being direct or indirect) on compromised cognitive function and associated general frailty within the elderly.

3

4

Clinical implications of brain aging

4.1

Normative brain volumetry derived from different reference populations: impact on single-subject diagnostic assessment in dementia

EJ Vinke, W Huizinga, M Bergtholdt, HH Adams, RME Steketee, JM Papma, FJ de Jong, WJ Niessen, MA Ikram, F Wenzel, MW Vernooij

Neurobiology of Aging. 2019;84:9-16

Brain imaging data are increasingly made publicly accessible, and volumetric imaging measures derived from population-based cohorts may serve as normative data for individual patient diagnostic assessment. Yet, these normative cohorts are usually not a perfect reflection of a patient's base population, nor are imaging parameters such as field strength or scanner type similar. In this proof of principle study, we assessed differences between reference curves of subcortical structure volumes of normal controls derived from two population-based studies and a case-control study. We assessed the impact of any differences on individual assessment of brain structure volumes. Percentile curves were fitted on the three healthy cohorts. Next, percentile values for these subcortical structures for individual patients from these three cohorts, 91 mild cognitive impairment and 95 Alzheimer's disease cases and patients from the Alzheimer Center, were calculated, based on the distributions of each of the three cohorts. Overall, we found that the subcortical volume normative data from these cohorts are highly interchangeable, suggesting more flexibility in clinical implementation.

Introduction

Methods to assist (early) diagnosis of neurological diseases and neuropsychiatric disorders in a clinical setting are of great importance. Noninvasive brain imaging, for example, with magnetic resonance imaging (MRI), is an increasingly applied diagnostic tool to detect brain pathology. To detect pathology on brain imaging, an understanding of what is normal is important, especially in diseases with a strong age-related component. A background of “normal aging” should therefore be taken into account, something that is difficult to estimate on a visual assessment alone. Many studies have focused on creating normative values of a broad spectrum of imaging markers of the human brain. By combining small to relatively large imaging data samples of healthy controls from different studies to one large imaging data set, normative values for different brain structure regions in aging were estimated and presented for clinical use.[286–290] With more and more brain imaging data from large population cohorts being publicly accessible, simply choosing a single population cohort to use as reference data would be feasible and in many (clinical) settings the most pragmatic option. However, the ideal reference population is the base population from which that individual patient arises, but data from such a population are rarely available in the clinical setting. Although the added value for diagnostic purposes of the use of normative values on top of visual assessment alone in a clinical setting is increasingly recognized[291–294], it is not known to what extent

variations in reference populations may affect the individual patient comparison to reference data. Furthermore, the choice of reference population is accompanied by differences in scanner types, field strength, and acquisition parameters between normative cohorts, which could introduce variation in results obtained from automated brain segmentation methods. Regarding the latter, several studies examined the robustness of automated segmentation methods across field strengths and scanner types, which have shown that reproducible segmentations can be obtained with residual volumetric variability of a few percent.[295–299] Yet, even with a perfectly robust segmentation method, the question remains whether population differences in structural brain volumes may impact individual patient comparison and whether this would lead to different clinical management. Are reference populations derived from case-control studies, "healthy controls" for example, similar to reference populations derived from population-based cohorts? Or does a reference population need to be similar to the base population from which an individual patient arises? Studies using normative reference data for diagnosis of neurological diseases, such as Alzheimer's disease (AD), commonly focus on volumetric changes in cortical gray matters areas.[287,300] More recently, interest in the role of volume and shape of subcortical brain structures is growing as relevant (early) brain imaging markers.[301–303] A novel approach for subcortical brain segmentation in T1-weighted MRI brain scans was recently presented, based on a shape-constrained deformable surface model.[304] Experiments on data both 3T and 1.5T for different scanners indicate good agreement with respect to independent ground truth segmentations of the subcortical structures using this model-based brain segmentation (MBS) approach, regardless of the field strength or vendor. In this proof-of-principle study, we assessed differences in normative reference curves for subcortical structure volumes (including hippocampal volume) segmented with the MBS method, between reference populations derived from two population-based studies and normal controls from a large case-control study. Furthermore, we assessed the impact of using these different cohorts on individuals with a higher risk of developing AD (APOE ε4 allele carriers and subjects with mild cognitive impairment (MCI)) and patients with AD.

Material

Reference populations

In this study, cross-sectional samples of three reference populations were used to estimate and compare the subcortical volume percentile curves. The reference populations included the Rotterdam Study, the United Kingdom Biobank (UKBB), and normal controls from the Alzheimer's Disease Neuroimaging Initiative (ADNI).

These studies were approved by a research or medical ethical committee, and informed consent was obtained from all subjects. From each study, 3D T1-weighted imaging data were used for subcortical structure segmentation. In Supplemental Figure 1 the age-distribution of the healthy participants of the Rotterdam Study, ADNI and UKBB are shown.

Rotterdam Study

We included 895 T1-weighted scans (median age = 66.4, interquartile range (IQR) = 22.7, 504 women) from the population-based Rotterdam Study, a prospective longitudinal study among community-dwelling subjects aged 45 years and over.[36] Scans were randomly selected from the study such that the age at time of the scan was uniformly distributed within a range of 45-95 years. All brain scans were acquired on a single 1.5-Tesla MRI system (GE Healthcare, US)[36]. In total, 225 of the 895 participants were APOE ε4 carriers (25.1%).

Alzheimer's Disease Neuroimaging Initiative

We included 430 (median age = 74.1, IQR = 7.5, 217 women) baseline 3D T1-weighted MRI scans from healthy controls from ADNI. ADNI is a longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD (adni.loni.usc.edu, for up-to-date information, see www.adni-info.org). With the ADNI data set being the smallest cross-sectional data set of the three cohorts, no further selection based on age was performed, resulting in an age range between 55 and 90 years. Participants were scanned on a 1.5- (n = 231, 53.7%) or 3-Tesla (n = 199, 46.3%) MRI system from GE Medical (n = 162), Philips (n = 71), or Siemens (n = 197). In total, 114 of the 430 participants were APOE ε4 carriers (26.5%).

United Kingdom Biobank

We included 876 (median age = 55.0, IQR = 15.0, 428 women) 3D T1-weighted scans from UKBB, all scanned with a 3-Tesla MRI system (Siemens Healthcare, UK). Scans were randomly selected from the study such that the age at time of the scan was uniformly distributed within a range of 40-70 years. UKBB is a prospective resource gathering extensive questionnaires, physical and cognitive measures, and biological samples in a cohort of 500,000 participants.[305] In total, 238 of the 876 participants were APOE ε4 carriers (27.2%).

Patient data for subject-specific comparison

We assessed 3D T1-weighted scans from participants with MCI and AD from the

Rotterdam Study and ADNI database and a sample of the APOE ε4 allele carriers from the healthy participants from the three reference populations, to evaluate whether subject-specific percentile estimations of different participant groups (APOE ε4 allele carriership, MCI or AD) depend on the chosen reference population. Furthermore, as an independent patient data set, we included MCI and AD cases from the Alzheimer Center Erasmus MC.

AD and MCI cases from the Rotterdam Study and ADNI

From the Rotterdam Study, 3D T1-weighted scans were selected from study participants with MCI (n = 41, age = 72 ± 6.4, 22 women) and prevalent AD (n = 45, age = 81.9 ± 4.6, 25 women) at time of the scan. From the ADNI data set, we selected the baseline 3D T1-weighted scan from patients with MCI (n = 50, age = 75.6 ± 7.0, 19 women) and patients with AD (n = 50, age = 75.1 ± 7.7, 28 women).

AD and MCI cases from the Alzheimer Center Erasmus MC

Scans from patients with MCI and AD from the Alzheimer Center Erasmus MC, Rotterdam, The Netherlands, were used as an independent set. Use of clinical data from the Alzheimer Center for research purposes was approved by the local medical ethical committee. Informed consent was obtained from all patients. We used 19 3D T1-weighted scans from patients with MCI (8 women, age = 69.4 ± 5.6) and 43 3D T1-weighted scans from patients with AD (15 women, age=66.8 ± 9.6) who visited the Alzheimer Center Erasmus MC between 2011 and 2016. All patient data were acquired on a single 1.5T MRI system (GE Healthcare, US).

Participant groups

In the rest of the article, the term "participant groups" will be used to describe the different subgroups on which the analyses are performed. The participant groups consist of the following:

- **Healthy:** healthy participants from the three reference populations (N_{total} = 2201, Rotterdam Study: 895, ADNI: 430, UKBB: 876).
- **APOE ε4 carriers:** Healthy participants from the three reference populations who carry one or two APOE ε4 allele(s) (N_{total} = 158, Rotterdam Study: 47, ADNI: 61, UKBB: 50).
- **MCI:** participants from the Rotterdam Study and ADNI data set with MCI (N_{total} = 91, Rotterdam Study: 41, ADNI: 50).
- **AD:** participants from the Rotterdam Study and ADNI data set with AD (N_{total} = 95, Rotterdam Study: 45, ADNI: 50).
- **MCI AC**: patients with MCI who visited the Alzheimer Center (N_{total} = 19).
- **AD AC:** patients with AD who visited the Alzheimer Center (N_{total} = 43).

Methods

Segmentation of subcortical structures on 3D T1-weighted data

This work is based on an MBS as described by Wenzel et al.[304], utilizing a shape-constrained deformable surface model for segmentation of subcortical brain structures from T1-weighted MRI. Adaptation of subcortical brain surfaces is performed stepwise, starting with global rigid and affine adaptation and followed by multi-affine and fully deformable adaptation. In each step, a weighted sum of internal and external energy is minimized. Here, internal energy relates to deviations from a shape/point distribution model of a training data set. The external energy component is based on the triangle-specific spatial distance to a target point along its normal. Target points are estimated with boundary detector functions that have been trained via a simulated search on the same training data set. For the used version of MBS, the training data set included 96 manually delineated 3T scans, equally distributed between patients with AD and healthy controls between ages 50 and 90 year as well as three device manufacturers (Philips, Siemens, and GE). The segmentation software is optionally available as part of the IntelliSpace Discovery workstation for data analytics in medical imaging.

Percentile curve fitting

For fitting of percentile curves for each subcortical volume in each of the three normative cohorts, we used the lambda-mu-sigma (LMS) method.[226] The LMS method can deal with skewed distributions and results in smooth percentile curves. The assumption of the LMS method is that the data are standard normally distributed after applying the Yeo-Johnson transformation, which is an extension of the Box-Cox transformation.[226] This method estimates the λ-parameter of the Yeo-Johnson transformation[306] (L), the median (M), and coefficient of variation (S) for the appropriate subcortical structure volume at each age. With the parameters L, M, and S, percentiles can be computed at each age to obtain a smooth curve. The smoothness of the fitted curves is influenced by the degrees of freedom δ, a user-defined parameter. In our experiments, we set the smoothness parameter δ to a value of 2 and we utilized the R-package VGAM[307] for the percentile curve fitting. The volume of a brain region may also be influenced by other covariates than age, for example, sex and head size. Including a covariate in the LMS model results in an age-dependent correction for the confounder. We therefore included sex in the LMS model as an confounder, which allows different percentile curves for men and women. To ensure an head size correction independent of age, head size was regressed out before fitting the LMS models. The precision of the estimated percentile curves depends on the number of data points in the appropriate age range. If the data are nonuniformly

distributed over age, it could be that the curve estimation is not precise in the part where there are very few data points. To assess the precision of the fitted curves, we used a bootstrapping procedure, by random sampling subjects with replacement and re-estimating the percentile curves. A distribution of possible curves was collected, from which confidence intervals were estimated.[308]

Percentile curves were fit on the Rotterdam Study, UKBB, and ADNI reference populations separately for the subcortical volumes of the hippocampus, amygdala, putamen, thalamus, caudate and nucleus accumbens, and globus pallidus. With the MBS, the volume of the caudate and the nucleus accumbens are combined into one volume. Furthermore, for the analysis, the subcortical volumes were the sum of the left and right volume. The MBS method does not segment the extraventricular cerebrospinal fluid (CSF); therefore, the exact intracranial volume (i.e., the sum of brain tissue and all CSF) was not available. To correct for head size, the "estimated intracranial volume" was constructed as the sum of total brain volume and the intraventricular CSF volume. An explorative comparison of the estimated intracranial volume and the intracranial volume segmented previously in the Rotterdam Study for other purposes with FreeSurfer 5.1 showed a good correlation (0.93); therefore, the estimated intracranial volume was used to correct for head size.

Subject-specific comparison

To assess the influence of using a specific reference population on subject-specific percentile values, scans from all three cohorts served as a joint test set to reduce a cohort-specific bias caused by the different age range covered by each cohort. We estimated the percentile value for every subcortical structure, for all participant groups based on each of the three reference cohorts. This results in three percentile values per subcortical volume for each participant. To assess differences in these percentile values, the distributions of the percentile values based on the three reference populations within the different participant groups are compared using a Welch's two-sample t-test. In addition, the shift function, as described by Rousselet and Wilcox et al.[309], was used to describe the differences between the percentile distributions based on the three different populations, to account for non-normally distributed percentile distributions within the participant groups.

Results

Table 1 shows the characteristics of the different participant groups. Characteristics of the participant groups per cohort are shown in Supplementary Table 1.

Table 1. Characteristics of the participant groups.

Characteristic	Healthy	APOE ε4 carriers	MCI	AD	MCI AC	AD AC
Age (y)[a]	63.6 (20.5)	68.8 (18.7)	74.5 (10.7)	79.2 (10.4)	70.1 (6.9)	66.0 (11.9)
Sex, women	1149 (0.52)	74 (0.47)	41 (0.45)	53 (0.56)	8 (0.42)	15 (0.35)
Hippocampus volume (mL)	6.3 (0.8)	6.3 (0.8)	5.7 (0.9)	4.7 (0.8)	6.0 (0.7)	5.3 (0.9)
Amygdala volume (mL)	1.9 (0.3)	1.9 (0.3)	1.7 (0.3)	1.4 (0.3)	1.9 (0.3)	1.6 (0.3)
Putamen volume (mL)	8.2 (1.0)	8.3 (1.0)	7.7 (1.0)	7.2 (0.8)	7.9 (1.0)	7.3 (0.8)
Thalamus volume (mL)	13.0 (1.5)	13.1 (1.4)	12.2 (1.5)	11.2 (1.0)	12.2 (1.3)	12.2 (1.7)
Caudate and accumbens[b] volume (mL)	7.5 (0.9)	7.6 (0.9)	7.2 (0.9)	6.6 (0.8)	7.2 (1)	6.8 (1.4)
Globus pallidus volume (mL)	2.8 (0.4)	2.8 (0.3)	2.7 (0.4)	2.5 (0.3)	2.8 (0.3)	2.7 (0.3)
Estimated intracranial volume (mL)	1229.4 (127.2)	1245.0 (126.6)	1229.4 (144.8)	1137.7 (123.6)	1239.4 (137.9)	1212.2 (140.9)

Continuous variables are presented as means (standard deviations), and categorical variables as numbers (percentages).

Key: APOE ε4 carriers, healthy participants from the three reference populations who carry one or two APOE ε4 allele(s) (N_{total} = 158); AD AC, patients with AD who visited the Alzheimer Center (N_{total} = 43); AD, participants from the Rotterdam Study and ADNI data set with AD (N_{total} = 95); Healthy, healthy participants from the three reference populations (N_{total} = 2201); MCI, participants from the Rotterdam Study and ADNI data set with MCI (N_{total} = 91); MCI AC, patients with MCI who visited the Alzheimer Center (N_{total} = 19).

[a]Age is presented as the median and interquartile range because of the non-normal distribution of age.

[b]Combined caudate and nucleus accumbens volume (mL).

Normative percentile curves

In Figure 1, the normative percentile curves based on the Rotterdam Study, ADNI, and UKBB data sets are shown for the subcortical structure volumes: (A) hippocampus, (B) amygdala, (C) thalamus, (D) putamen, (E) caudate and nucleus accumbens, and (F) globus pallidus. Considering the percentile curves and the corresponding confidence intervals around each curve, the percentile curves of hippocampus volume and caudate and nucleus accumbens of the three normative cohorts largely overlap, with a slightly higher volume in ADNI compared with Rotterdam Study and UKBB. For the amygdala, the percentile curves show small differences with higher volumes for UKBB, followed by ADNI and the lowest volumes for Rotterdam Study. For the putamen, thalamus, and globus pallidus, the ADNI and UKBB curves largely overlap, but the Rotterdam Study percentile curves show a lower volume. Furthermore, for almost all subcortical structures, the Rotterdam Study percentile curves show a larger decrease in volume over age than ADNI; however, the steepness of the curves between Rotterdam Study and UKBB seems comparable.

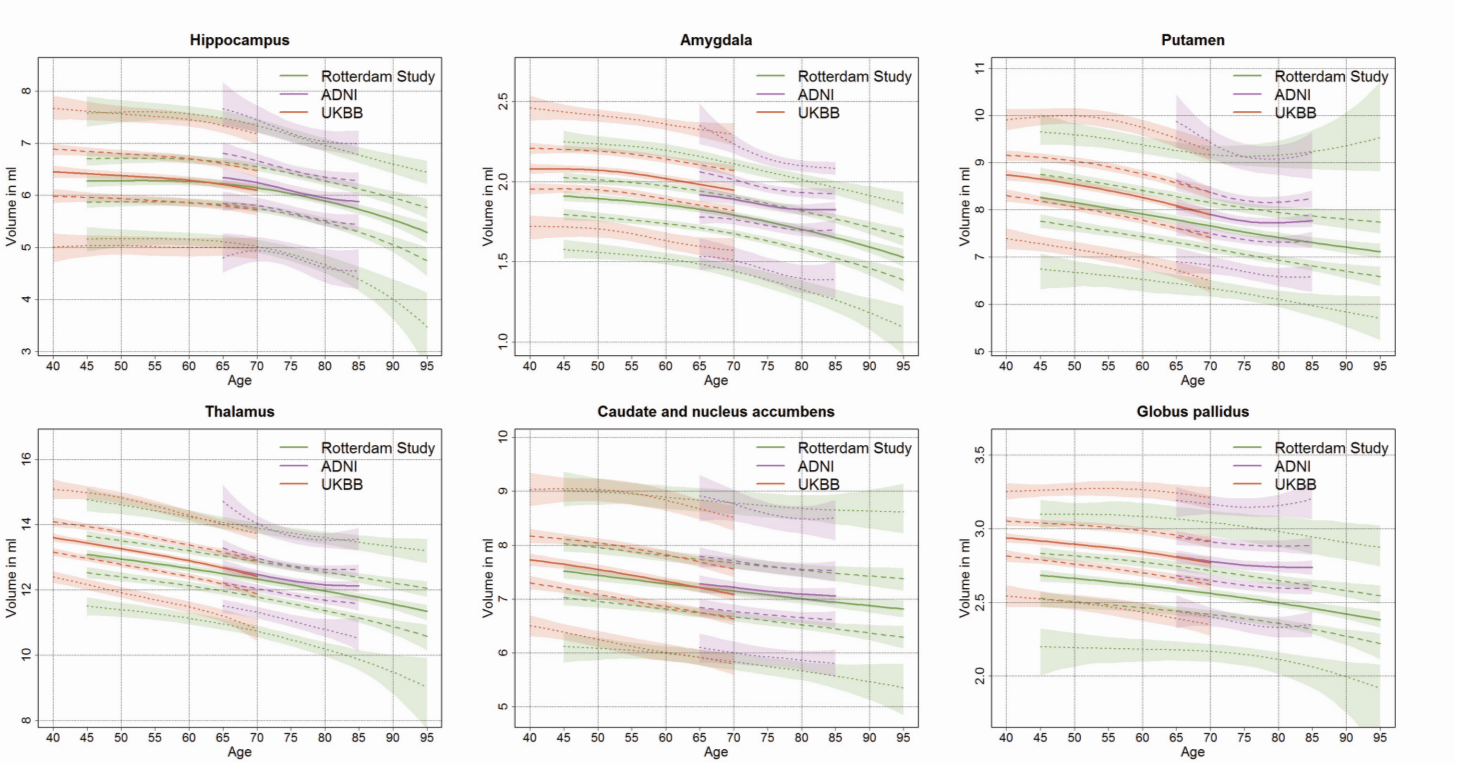


Figure 1. Percentile curves of the subcortical structure volumes based on Rotterdam Study, UK Biobank (UKBB), and Alzheimer's Disease Neuroimaging Initiative (ADNI) data. The lower and upper dotted lines represent the 2.5% and 97.5% percentile curves, the lower and upper dashed lines represent the 25% and 75% percentile lines, and finally, the solid line represents the 50% percentile line, respectively. Around each percentile line the confidence interval is shown.

Table 2. Comparison of percentile values of the participant groups based on each of the three cohorts as reference curves.

Subcortical structure	Participant group	RS_{ref}	$ADNI_{ref}$	$UKBB_{ref}$	Difference (p-value)	N_{total} (RSS;ADNI;UKBB)
Hippocampus	Healthy	0.51 (0.29)	0.51 (0.30)	0.51 (0.30)	1	2201 (895; 430; 876)
	APOE ε4 carriers	0.51 (0.31)	0.51 (0.32)	0.51 (0.32)	1	158 (47; 61; 50)
	MCI	0.34 (0.30)	0.33 (0.30)	0.34 (0.32)	1	91 (41; 50; 0)
	AD	0.14 (0.21)	0.14 (0.21)	0.13 (0.21)	1	95 (45; 50; 0)
Amygdala	Healthy	0.62 (0.30)	0.56 (0.31)	0.38 (0.29)	<0.001[a]	2201 (895; 430; 876)
	APOE ε4 carriers	0.60 (0.32)	0.55 (0.33)	0.38 (0.29)	<0.001[b]	158 (47; 61; 50)
	MCI	0.40 (0.31)	0.34 (0.30)	0.20 (0.24)	<0.001[b]	91 (41; 50; 0)
	AD	0.22 (0.27)	0.18 (0.24)	0.09 (0.16)	<0.001[c]	95 (45; 50; 0)
Putamen	Healthy	0.57 (0.28)	0.55 (0.30)	0.49 (0.30)	<0.001[b]	2201 (895; 430; 876)
	APOE ε4 carriers	0.60 (0.26)	0.58 (0.29)	0.52 (0.29)	0.39	158 (47; 61; 50)
	MCI	0.51 (0.25)	0.47 (0.28)	0.46 (0.27)	1	91 (41; 50; 0)
	AD	0.48 (0.29)	0.43 (0.31)	0.50 (0.30)	1	95 (45; 50; 0)
Thalamus	Healthy	0.56 (0.28)	0.61 (0.31)	0.52 (0.29)	<0.001[a]	2201 (895; 430; 876)
	APOE ε4 carriers	0.59 (0.27)	0.64 (0.29)	0.56 (0.28)	0.17	158 (47; 61; 50)
	MCI	0.41 (0.29)	0.43 (0.32)	0.46 (0.29)	1	91 (41; 50; 0)
	AD	0.42 (0.27)	0.39 (0.29)	0.51 (0.26)	0.028[d]	95 (45; 50; 0)
Caudate and nucleus accumbens	Healthy	0.51 (0.28)	0.53 (0.30)	0.53 (0.29)	1	2201 (895; 430; 876)
	APOE ε4 carriers	0.55 (0.26)	0.57 (0.28)	0.57 (0.28)	1	158 (47; 61; 50)
	MCI	0.45 (0.28)	0.45 (0.30)	0.50 (0.30)	1	91 (41; 50; 0)
	AD	0.38 (0.27)	0.38 (0.29)	0.47 (0.30)	0.54	95 (45; 50; 0)
Globus Pallidus	Healthy	0.66 (0.28)	0.42 (0.32)	0.40 (0.30)	<0.001[e]	2201 (895; 430; 876)
	APOE ε4 carriers	0.7 (0.26)	0.47 (0.32)	0.46 (0.31)	<0.001[e]	158 (47; 61; 50)
	MCI	0.67 (0.25)	0.40 (0.30)	0.41 (0.28)	<0.001[e]	91 (41; 50; 0)
	AD	0.61 (0.30)	0.36 (0.31)	0.39 (0.30)	<0.001[e]	95 (45; 50; 0)

Mean and standard deviation of the percentiles of the different participant groups (healthy participants, APOE ε4 carriers, participants with MCI, and participants with AD) based on the reference curves of each of the three normative cohorts ($RS_{ref,}$ $ADNI_{ref}$, and $UKBB_{ref}$).

Difference: smallest *p*-value of the paired *t*-tests; N_{total}: sample size of the participant groups.

Key: RS, Rotterdam Study; UKBB, United Kingdom Biobank; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE ε4 carriers, healthy participants from the three reference populations who carry one or two APOE ε4 allele(s); AD, participants from the Rotterdam Study and ADNI data set with AD; AD AC, patients with AD who visited the Alzheimer Center; Healthy, healthy participants from the three reference populations; MCI, participants from the Rotterdam Study and ADNI data set with MCI; MCI AC, patients with MCI who visited the Alzheimer Center.

[a]percentile values based on the three normative cohorts are all significantly different from each other.

[b]percentile values based on UKBB data are significantly different from those based on the other cohorts.

[c]percentile values based on Rotterdam Study data are significantly different from those based on the UKBB data.

[d]percentile values based on ADNI data are significantly different from those based on the UKBB data.

[e]percentile values based on Rotterdam Study data are significantly different from those based on the other cohorts.

Subject-specific comparison

In Table 2, the average percentile values and standard deviations are shown for the different participant groups when based on each of the three normative cohorts. In general, differences shown in the percentile curves in Figure 1 result in significant differences in the percentile distributions. For hippocampus and caudate volume, there are no significant differences in using the percentile curves from the Rotterdam Study, ADNI, or UKBB for any of the different participant groups. For the volumes of the putamen, there were significant differences in the percentiles within the healthy participants. Yet, for the APOE ε4 carriers, patients with MCI and AD, these differences were not significant. For the volumes of the thalamus, both the percentiles of the healthy participants and the patients with AD were statistically significant.

For the amygdala volume, the percentile values based on the three cohorts were all significantly different. However, for the APOE ε4 carriers, MCI and AD cases, only the percentile values based on UKBB were significantly lower than the other reference cohorts. For the globus pallidus, there was a significantly higher percentile value based on the Rotterdam Study data versus the other two cohorts, which is a reflection of the significantly lower percentile curves for the Rotterdam Study, as shown in Figure 1.

In Supplemental Figure 2, the results from the shift-function analyses are shown, for the four different participant groups. The results show overall a straight line for all participant groups, indicating a fixed percentile difference when comparing the percentiles based on two different populations, which is independent of the percentile value itself. The exceptions are the comparison of amygdala percentiles based on UKBB compared with those based on the Rotterdam Study and ADNI. Here, a higher percentile value is related to a larger percentile difference. The same holds for the globus pallidus percentile based on the Rotterdam Study compared with ADNI and UKBB.

Finally, in Table 3, the mean and standard deviation of the estimated percentiles for the participants with AD and MCI from the Alzheimer Center are shown. Within this sample, there was only a significant difference for the globus pallidus volume in the patients with AD. Other percentile estimations in these groups did not differ depending on the reference curves applied.

Table 3. Comparison of percentile values of the participants with MCI and AD from the Alzheimer Center, based on the reference curves from each of the three normative cohorts (RS_{ref}, $ADNI_{ref}$, and $UKBB_{ref}$)

Subcortical structure	Participants	Average percentile (SD)			Difference	N
		RS_{ref}	$ADNI_{ref}$	$UKBB_{ref}$		
Hippocampus	MCI	0.36 (0.32)	0.36 (0.33)	0.38 (0.33)	1	19
	AD	0.16 (0.23)	0.16 (0.22)	0.16 (0.22)	1	43
Amygdala	MCI	0.58 (0.36)	0.53 (0.36)	0.39 (0.33)	1	19
	AD	0.27 (0.3)	0.22 (0.28)	0.14 (0.24)	0.5	43
Putamen	MCI	0.49 (0.3)	0.46 (0.32)	0.42 (0.31)	1	19
	AD	0.3 (0.23)	0.26 (0.24)	0.23 (0.23)	1	43
Thalamus	MCI	0.33 (0.23)	0.34 (0.26)	0.31 (0.22)	1	19
	AD	0.34 (0.29)	0.35 (0.32)	0.31 (0.3)	1	43
Caudate and nucleus accumbens	MCI	0.41 (0.3)	0.41 (0.31)	0.43 (0.3)	1	19
	AD	0.33 (0.33)	0.34 (0.34)	0.35 (0.34)	1	43
Globus pallidus	MCI	0.67 (0.24)	0.4 (0.32)	0.4 (0.3)	0.08	19
	AD	0.66 (0.31)	0.44 (0.33)	0.43 (0.32)	0.023[a]	43

Difference: smallest p-value of the paired t-tests; N_{total}: sample size of the participant groups; N: sample size of the Alzheimer Center set.

Key: AD, Alzheimer's disease; MCI, mild cognitive impairment; RS, Rotterdam Study; UKBB, United Kingdom Biobank; ADNI, Alzheimer's Disease Neuroimaging Initiative.

[a]percentile values based on Rotterdam Study data are significantly different from those based on other cohorts.

Discussion

In this study, we calculated normative reference curves for subcortical structure volumes from reference populations that were either derived from population-based studies or from normal controls of a case-control study. We used a segmentation method for which previous experiments on data from both 3T and 1.5T for different scanners indicate good agreement with respect to independent ground truth segmentations, regardless of the field strength or vendor. We found that for most subcortical structures, the percentile curves of the subcortical structures largely overlap. This indicates only small differences between the subcortical volumes of these reference populations, regardless of differences in vendors, field strength, acquisition, and population differences. When estimating the percentile values for various participant groups that may be evaluated in a clinical setting (APOE ε4 carriers, and patients with MCI and AD), the choice of reference population did not influence the percentile distribution significantly, except for the smallest subcortical structures: amygdala and globus pallidus.

In particular, the hippocampus percentile curve was very robust across the participant groups. This indicates that individual diagnostic assessment in a clinical setting, based on subcortical volume information, may not be biased by the use of a specific reference population.

Strengths and limitations

A major strength of this study is the use of a single segmentation tool on MRI scans from various different large reference populations, giving a comprehensive overview of subcortical volumes in aging in these populations. Another strength of the study is the availability of scans from patient groups (MCI and AD) from the Rotterdam Study and ADNI, as well as a patient population independent from the reference populations, that is, the Alzheimer Center data. There are a number of limitations associated with this study. First, a limitation concerning the volume segmentation method used in this study is the lack of segmentation of extraventricular cerebrospinal fluid because of which the intracranial volume estimated in this study gives an underestimation of the true intracranial volume (or head size). This underestimation may lead to an underestimation of the atrophy effect in aging, when the changes in ventricular cerebrospinal fluid are not representative of the extraventricular cerebrospinal fluid changes in aging. Yet, a sensitivity analysis within the Rotterdam Study population in whom both estimated intracranial volume and exact intracranial volume were available showed these effects to be negligible. Second, the LMS method used in this study for the estimation of the percentile curves results in smooth percentile curves in aging, which can deal with skewed distributions. Within this study, other methods to estimate percentile curves would also have been suitable, assuming that the subcortical volumes over age are normally distributed. Within the context of this study, we believe that the impact of the choice of the percentile curve estimation method on the differences between populations is minimal, as long as the same percentile curve fitting method is the same for the different reference populations. Third, a limitation concerning the generalizability of these percentile curves is the fact that the vast majority of the healthy study participants and the participant groups are Caucasian. Therefore, differences in percentile curves which could result from differences in ethnicity of study populations are not assessed in this study. Fourth, in this study, we are not able to determine the exact source of differences in the subcortical volumes between reference populations because of the variation in vendor, field strength, and acquisition used in the different populations. However, we are able to demonstrate the magnitude of these differences, indicating the impact of these differences on individual patient assessment in an everyday clinical setting. Fifth, the lack of overlap of the complete age range of all three reference populations is a limitation of

this study, making comparison of the reference curves more difficult. Given the important differences in age ranges, comparison of percentile volumes of healthy participants was performed on the combined healthy participants of the Rotterdam Study, ADNI, and UKBB, although the percentile curves itself were fitted on these same reference populations. Ideally, separate healthy participants from the same reference populations, which are not included in the percentile curve fitting, would be used to test percentile value differences for the different reference populations. Furthermore, this study was limited to the subcortical structure volumes including hippocampus volume, whereas in a clinical setting, other (cortical) brain volumes would be also of importance. The current choice for subcortical volumes was driven on the one hand by an increasing scientific interest into subcortical volumes in neurological diseases (including neurodegenerative diseases in older age) and on the other hand because of the availability of a proven robust segmentation algorithm, which performs population- and vendor-independent, eliminating potential sources of noise. Yet, a logical next step would be to explore the dependence of cortical segmentation algorithms on the choice of reference population. Finally, in this study, the patient population on which the effect of different reference populations were estimated consisted of only patients with AD or participants at higher risk of AD. Next step would be to evaluate reference curves of a broad spectrum of brain structures based on different reference populations and the effect of these differences on diagnostic assessments in different neurological diseases and neuropsychiatric disorders.

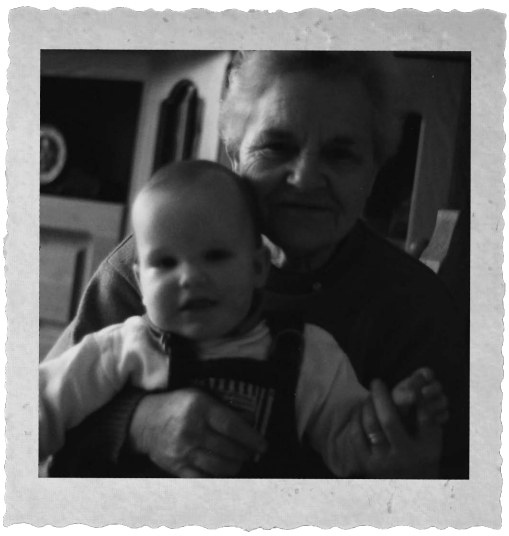
Differences between reference populations

In general, we found slightly lower reference volumes based on the scans from the Rotterdam Study compared with the ADNI and UKBB reference curves. A possible explanation for these differences is that the Rotterdam Study population is a population-based cohort, whereas ADNI has a case-control design. Within the healthy set of the Rotterdam Study, participants with MCI or AD at time of the scan were excluded, whereas the ADNI controls are cognitively normal with no memory complaints and no significant cognitive impairment. Therefore, the control subjects from ADNI are expected to be healthier than the Rotterdam Study population. On the other hand, Rotterdam Study percentile curves also show a slightly lower value than the UKBB percentile curves, which is a population-based cohort as well. This may be due to a lower response rate in the UKBB, with the possibility of healthy selection bias. Furthermore, a slight increase in subcortical volume has been seen in the ADNI reference curves from age 80 to 85. A possible explanation might be that with increasing age (especially from age 80 and older), the ADNI controls become proportionately healthier than control subjects from a population-based study

because of the fact that the higher a subject's age, the more likely he or she is to have memory complaints. This could be interpreted as an increasing healthy selection bias with age. Another possible explanation for differences between reference populations could be the fact that scans from both the Rotterdam Study and UKBB have each been acquired on the same vendor (with only a single scanner for the Rotterdam Study), whereas scans in the ADNI database were collected from different scanners, field strengths, and scanner types. Therefore, characteristics of a single scanner or vendor with an impact on volumetric segmentation might be more dominant in Rotterdam Study and UKBB curves. This effect might have a larger impact on small structures as well as such with subtle contrast boundaries like amygdala and globus pallidus, explaining more pronounced differences between their corresponding percentile curves and additionally the larger percentile differences with higher percentile volumes in these small structures. The study by Potvin et al.[286] that created normative curves for subcortical structures and evaluated the effects of scanner characteristics also showed that for the amygdala structure volume, the effects of scanner characteristics were modest, whereas in the most other structures, the effect was minor compared to age, sex, and intracranial volume.

Summary and conclusion

Overall, we found that the percentile curves of the subcortical structure based on three different reference populations largely overlap, indicating only small differences between the subcortical volumes of these populations, regardless of differences in vendors, field strength, acquisition, and population differences. Therefore, we conclude that the subcortical volume data of these three cohorts are interchangeable, suggesting more flexibility in clinical implementation.



4.2

Progression along APOE-specific data-driven temporal cascades is predictive of Alzheimer's disease in a population-based cohort

V Venkatraghavan*, EJ Vinke*, EE Bron, WJ Niessen, MA Ikram, S Klein, MW Vernooij

NeuroImage. 2021;238

Data-driven disease progression models have provided important insight into the timeline of brain changes in AD phenotypes. However, their utility in predicting the progression of pre-symptomatic AD in a population-based setting has not yet been investigated. In this study, we investigated if the disease timelines constructed in a case-controlled setting, with subjects stratified according to APOE status, are generalizable to a population-based cohort, and if progression along these disease timelines is predictive of AD. Seven volumetric biomarkers derived from structural MRI were considered. We estimated APOE-specific disease timelines of changes in these biomarkers using a recently proposed method called co-initialized discriminative event-based modeling (co-init DEBM). This method can also estimate a disease stage for new subjects by calculating their position along the disease timelines. The model was trained and cross-validated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset, and tested on the population-based Rotterdam Study (RS) cohort. We compared the diagnostic and prognostic value of the disease stage in the two cohorts. Furthermore, we investigated if the rate of change of disease stage in RS participants with longitudinal MRI data was predictive of AD. In ADNI, the estimated disease timelines for ε4 non-carriers and carriers were found to be significantly different from one another ($p<0.001$). The estimate disease stage along the respective timelines distinguished AD subjects from controls with an AUC of 0.83 in both APOE ε4 non-carriers and carriers. In the RS cohort, we obtained an AUC of 0.83 and 0.85 in ε4 non-carriers and carriers, respectively. Progression along the disease timelines as estimated by the rate of change of disease stage showed a significant difference ($p<0.005$) for subjects with pre-symptomatic AD as compared to the general aging population in RS. It distinguished pre-symptomatic AD subjects with an AUC of 0.81 in APOE ε4 non-carriers and 0.88 in carriers, which was better than any individual volumetric biomarker, or its rate of change, could achieve. Our results suggest that co-init DEBM trained on case-controlled data is generalizable to a population-based cohort setting and that progression along the disease timelines is predictive of the development of AD in the general population. We expect that this approach can help to identify at-risk individuals from the general population for targeted clinical trials as well as to provide biomarker based objective assessment in such trials.

Introduction

Alzheimer's disease (AD) is a chronic neurodegenerative disease that affects roughly 3% of the world's elderly population (above 60 years old).[310] A major genetic risk factor for AD is the presence of ε4 allele of APOE.[311] Furthermore, APOE ε4 has also been shown to affect the clinical[312,313] and biological phenotypes of AD[314], making it a key factor in understanding the pathophysiology of AD.

Neuroimaging biomarkers play an important role in disentangling these phenotypes.[191,315] They could also play an important role in finding disease modifying treatments.[316] There has been evidence that selection of the study population at its pre-symptomatic stage is also crucial for the success of potential modifying treatments for AD.[317,318] Hence there is a crucial need for a way to objectively assess the progression of pre-symptomatic AD (or lack thereof).

Biomarkers extracted from neuroimaging data in combination with machine learning approaches have been shown to objectively assess the progression of AD in research cohorts[319] as well as in clinical cohorts[320]. However, machine learning approaches are not explainable by default and the lack of transparency in such approaches could hinder clinical decision making[321].

Disease progression models are data-driven approaches that are interpretable by design and can thus aid not only in predicting AD but also in explaining the decision and facilitating transparency and trust.[322] In recent years, many disease progression models have emerged to provide insight into neurodegenerative diseases such as AD.[323,324] Such insights have also been shown to aid in objective assessment of AD progression.[325] An example of such a model is the discriminative event-based model (DEBM)[326], which estimates a timeline of AD related biomarker abnormality events in a data-driven way. This model was recently extended further to identify APOE genotype-specific differences in AD biomarker progression, where the biomarkers, including volumetric measures obtained from MRI, were found to progress along different timelines depending on APOE status.[327] However, the generalizability of such models to population-based cohorts and their utility in predicting the progression of pre-symptomatic AD in a population-based setting have not yet been investigated.

In this work, we investigate if i) APOE-specific disease timelines constructed in a case-controlled setting are generalizable to a population-based cohort, and ii) if progression along these disease timelines is predictive of AD. For constucting the APOE-specific disease timelines, we use a recently developed approach called co-initialized (co-init) DEBM[327] meant for obtaining disease timelines in stratified

cross-sectional datasets. We demonstrate the potential of the method's fine-grained disease stage estimation in predicting the subjects with pre-symptomatic AD in the general population.

Methods

We first describe the inclusion criteria for participants and the method for obtaining the volumetric biomarkers in the case-controlled Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and the population-based Rotterdam study (RS) dataset. This is followed by the description of co-init DEBM used to construct APOE-specific disease timelines of volumetric biomarkers from baseline scans of the participants in the ADNI. We validated the disease timelines constructed on ADNI by assessing their generalizability to the population-based RS cohort, and by predicting the participants at-risk of becoming symptomatic in the RS cohort.

Participants

ADNI

We considered the baseline measurements of 335 cognitively normal (CN), 565 non-AD, 167 incident-AD and 223 AD participants (prevalent-AD) who had imaging data available in ADNI1, ADNIGO and ADNI2 studies. The non-AD cases were defined as ADNI participants who were either mild cognitively impaired (MCI) or had subjective memory complaints at the time of the baseline MRI scan, and did not develop AD within 3 years of follow-up. The incident-AD cases presented with MCI at baseline but developed AD within 3 years. The prevalent-AD and incident-AD subjects were defined by their clinical diagnosis of AD according to NINCDS-ADRDA's criteria for AD.[328,329] Characteristics of the subjects and their volumetric measures in the ADNI dataset included in our study are shown in Table 1(a).

Rotterdam study

We considered participants from the population-based RS cohort, a prospective longitudinal study among community-dwelling subjects aged 45 years and over.[194] Participants were screened for dementia at baseline and at follow-up examinations with the Mini-Mental State Examination and the Geriatric Mental Schedule organic level. Those with a Mini-Mental State Examination score < 26 or Geriatric Mental Schedule score > 0 underwent further investigation and informant interview, including the Cambridge Examination for Mental Disorders of the Elderly. In addition, the entire cohort was continuously under surveillance for dementia through electronic linkage of the study database with medical records from general

practitioners and the regional institute for outpatient mental health care. Available information on cognitive testing and clinical neuroimaging was used when required for diagnosis of dementia subtype. A consensus panel led by a consultant neurologist established the final diagnosis of AD according to NINCDS-ADRDA criteria for AD.

In this work, we included participants from the RS who had at least one MRI scan, who completed cognitive testing, and were interviewed for the presence of subjective cognitive complaints at the time of the MRI. The included participants were categorized into 4 groups: participants that were cognitively normal at the time of the scan (CN), participants that had subjective memory complaints and/or objective cognitive impairment[330], but who did not develop AD at follow-up (non-AD), participants with AD at the time of the scan (prevalent-AD) and participants who developed AD after the MRI scan (incident-AD). Unlike in ADNI, we did not set a threshold of conversion within 3 years to be included as an incident-AD participant, since we wanted to assess the utility of our method in monitoring the progression of both pre-clinical and prodromal AD subjects. Participants with clinical stroke were excluded.

In our experiments, we used two subsets of the RS cohort: the generalizibility set and the prediction set. The generalizibility set consisted of the last MRI scan available for each partipant in the RS cohort. This subset consisted of 998 CN, 2710 non-AD, 97 incident-AD, and 25 prevalent-AD cases and were used for experiments validating the generalizability of the APOE-specific disease timelines constructed using co-init DEBM. The characteristics of the subjects in this subset are shown in Table 1(b). The prediction set consisted of the last two MRI scans available for each participant, which were used to assess the progression (or lack thereof) of pre-symptomatic AD in the participants. This subset consisted of 183 CN, 852 non-AD and 31 incident-AD cases. For the incident-AD cases, both the included scans were performed before the AD diagnosis. Participants with prevalent-AD were excluded in this subset. The characteristics of the subjects in this subset are shown in Table 1(c). A scatter plot illustrating the longitudinal sampling in this prediction set is shown in Figure 1.

MRI Acquisition and imaging biomarker extraction

The imaging biomarkers used in this study were estimated from T1-weighted (T1w) MRI scans. ADNI participants were scanned on a 1.5T (N = 497) or a 3T (N = 739) MRI system from GE, Philips, or Siemens, using magnetization prepared - rapid gradient echo (MP-RAGE) sequence (voxel size: $1.0 \times 1.0 \times 1.0$ mm^3). RS participants were scanned on a single 1.5T MRI system from GE, using gradient

recalled echo (GRE) sequence (voxel size: $0.49 \times 0.49 \times 1.6$ mm^3). Details of the MRI acquisition protocol can be found in Jack et al.[331] (ADNI) and Ikram et al.[37] (RS). The MRI scans were analyzed with FreeSurfer software v6.0 cross-sectional stream (http://surfer.nmr.mgh.harvard.edu). Outputs were visually checked for the ADNI dataset. In the RS dataset, an automated quality metric was used to exclude scans with insufficient quality, which was visually verified in a randomly selected subset of both selected and rejected scans.[332]

The selected imaging markers were the same markers as that of Archetti et al.[333], namely volumetric measures of: total brain, ventricles, hippocampus, precuneus, middle temporal gyrus, fusiform gyrus and entorhinal cortex. The volumes were defined as the summed volumes of the structure in the left and right hemisphere. To take into account the confounding effects of age, sex, and intracranial volume, linear regressions were performed before constructing the disease timelines. The volumetric measures of CN subjects in ADNI were used to regress against age, sex and intracranial volume to estimate their confounding effects parameterized by their respective slopes and intercepts. These estimates were used for confounding factor correction in the remaining subjects in ADNI as well as in the RS cohort. The resultant volumetric measures will be referred to as biomarkers in the remainder of the manuscript.

Construction of APOE-specific disease timelines using co-init DEBM

The co-init DEBM model introduced by Venkatraghavan et al.[327], constructs genotype-specific AD related disease timelines of biomarker changes, based on cross-sectional datasets. Such an estimation from cross-sectional data is feasible because, in a cohort consisting of subjects encompassing a wide spectrum of severity, early biomarkers have a higher prevalence of abnormal biomarker values as compared to biomarkers that become abnormal later in the disease timeline. The co-init DEBM model estimates this timeline without strictly considering the diagnostic labels of the subjects. The model uses a coupled mixture model to jointly fit normal and abnormal distributions in the dataset stratified by (APOE) genotypes. The model assumes that the normal and abnormal biomarker distributions in the different genotypes can be approximately represented by Gaussians. It also assumes that the different genotypes' abnormal (and normal) biomarker distributions are close to each other.

After the estimation of the normal and abnormal biomarker distributions, the model computes the probability of abnormality of each biomarker for each subject in the training dataset. Based on the assumption that a biomarker that becomes abnormal

earlier in the disease timeline would be more abnormal than the biomarker that becomes abnormal later, it estimates a subject-specific ordering of biomarker changes in each subject of the dataset. A generalized Mallows model is used to average the subject-specific biomarker ordering over the subjects within each genotypic group of the training set, to construct average disease timeline for APOE ε4 non-carriers and carriers. Along with the sequence of the biomarker abnormality events, the model also estimates the relative positioning of such events with respect to each other (event-centers). Absolute magnitudes for these event-centers are irrelevant as they only convey relative (temporal) distances and in this study, they were normalized such that the first event and the last event coincided at a value of 0.1 and 0.9 respectively.

To construct the disease timelines, the co-init DEBM was trained on CN, incident-AD, and prevalent-AD subjects from ADNI. The non-AD subjects in ADNI were excluded for training the model, to reduce the chances of disorders unrelated to AD affecting the estimated timelines. The variance in the estimated disease timeline was computed using 100 independent bootstrap samples. In order to evaluate if the estimated orderings in APOE ε4 non-carriers and carriers were significantly different from one another, we used permutation testing and estimated the distribution of the Kendall's Tau distance under the null hypothesis. To compute this distribution, we generated 1,000 random permutations of the two groups. We then computed the one-sided p-values for the actual Kendall's Tau distances between the orderings of the two groups, calculated as the proportion of sampled permutations where the distance was greater than or equal to the actual distance.

Estimating APOE-specific disease stages

After training the co-init DEBM model, the constructed APOE-specific disease timelines were used to estimate the disease stage at multiple timepoints for subjects of the RS cohort. For estimating the disease stages of ADNI subjects, we used a 10-fold cross validation. The training set was used for constructing the disease timelines and the disease stages were estimated in the test set, including the non-AD subjects excluded in the training phase. Disease stage quantifies the severity of the disease in a subject by positioning them along the pre-constructed disease timelines and is normalized between 0 and 1. The estimated disease stages were used in two sets of experiments.

Experiment 1: Assessing the generalizability of co-init DEBM from ADNI to RS

In this experiment, we tested the generalizability of the co-init DEBM model trained on ADNI by evaluating the diagnostic and prognostic value of its predicted disease stages in the RS cohort. First we performed a visual assessment by constructing normalized histograms of the estimated APOE-specific disease stages for the different diagnostic classes in ADNI and the generalizability set of the RS cohort.

Complementing this visual analysis, for assessing the diagnostic value we used the estimated disease stages to distinguish prevalent-AD from two different reference groups in ADNI and in the generalizability set of the RS cohort. First, only the CN subjects were included in the reference group. To emulate a reference group of participants more representative of the general aging population than the CN group, we used a combined set of CN and non-AD subjects as the second reference group. We computed the area under the receiver operating curve (AUC) for distinguishing the diagnostic classes, and compared the AUCs obtained in ADNI and RS. The confidence intervals of these AUCs were measured using bootstrap resampling while stratifying the diagnostic classes to maintain their relative proportions. For assessing the prognostic value, we used the estimated disease stages to distinguish incident-AD from the aforementioned two reference groups in ADNI and in the generalizability set of RS cohort. We computed the AUCs and their confidence intervals for distinguishing these diagnostic classes and compared values obtained in ADNI and RS.

To compare the generalizability of a model that stratifies based on APOE carriership, with that of a model that does not, we repeated the experiment described above using disease timeline estimated in ADNI subjects, without stratifying for APOE. Furthermore, we computed the correlation of the estimated disease stages with time to dementia diagnosis for incident-AD subjects in ADNI as well as in RS. Lastly, we computed the Spearman correlation of the estimated disease stages with MMSE for subjects in ADNI as well as in RS.

Experiment 2: Predicting AD based on longitudinal data in the RS cohort

In this experiment, we assess if the evolution of the disease stages derived from longitudinal neuroimaging data is predictive of AD in the prediction set of the RS cohort. This experiment is further divided into three parts. In the first part, we build longitudinal trajectories of the disease stages and observe the differences in CN, non-AD and incident-AD subjects. In the second part, we assess the prognostic value of the rate of change of disease stages. Lastly, we assess the marginal utility of the follow-up scans in AD prognostication.

Exp. 2.1: We used the disease stages obtained in the prediction set of the RS cohort for building the trajectories of disease stages in the two APOE ε4 based groups. The trajectories were estimated using linear mixed models with random intercepts and slopes. The time variable in these linear mixed models was follow-up time in years since the first MRI of the subject. To allow different slopes for different diagnostic classes, an interaction between follow-up time and the diagnosis was integrated in the model. Covariates that were accounted for in the model were sex, age at the time of the first MRI, and the interaction of age and follow-up time to allow slope differences for different ages.

Exp. 2.2: We used the rate of change of disease stages (delta disease stage) in the prediction set of the RS cohort to distinguish incident-AD from two different reference groups. As in Experiment 1, the two reference groups selected were CN, and a combined set of CN and non-AD subjects. We computed the AUCs and their confidence intervals for distinguishing these diagnostic classes. For comparison, the AUCs while using the rate of change of the volumetric measures (normalized to their respective intracranial volumes) for distinguishing the same two classes were computed.

Exp. 2.3: Lastly, to evaluate the marginal utility of the follow-up scans for identifying incident-AD subjects, we used the estimated disease stage at the last MRI scan of the subjects in the prediction set of the RS cohort to distinguish incident-AD from the aforementioned two different reference groups. We computed the AUCs and their confidence intervals for distinguishing these diagnostic classes. As a comparison, the AUCs based on participants' age as well as of each individual volumetric imaging biomarker were also computed.

Results

Figure 2 shows the APOE-specific disease timelines constructed for the ε4 non-carriers and carriers in the ADNI dataset. It shows the centers of the biomarker abnormality events along the timeline representing their relative positioning with respect to each other. It can be seen that the disease timelines of APOE ε4 non-carriers and carriers were quite different. The permutation testing further confirmed that the disease timelines of ε4 non-carriers and carriers were indeed significantly different ($p < 0.001$). Most noticeably, ventricular volume and total brain volume were estimated as early biomarkers for APOE ε4 non-carriers, whereas hippocampal volume and volume of the entorhinal cortex were estimated as early biomarkers for APOE ε4 carriers. It can also be seen in Figure 2 that the uncertainty estimates in APOE ε4 non-carriers were greater than in APOE ε4 carriers.

Table 1. Characteristics of the A) ADNI dataset, B) the generalizability set of the RS dataset, and C) the prediction set of the RS dataset. *Indicates values at last scan.

A) ADNI dataset	**CN**	**non-AD**	**incident-AD**	**prevalent-AD**
Number of subjects	335	565	167	223
Number of women, %	174, 51.9	268, 47.4	68, 40.7	104, 46.6
Age (years)	74.3 ± 5.6	71.82 ± 7.2	73.1 ± 7.1	74.0 ± 7.9
Number of *APOE* ε4 carriers, %	92, 27.5	238, 42.1	121, 72.5	151, 67.7
Intracranial volume (ml)	1504.0 ± 155.8	1520.9 ± 152.8	1546.2 ± 180.2	1524.2 ± 183.9
Total brain volume (ml)	1030.7 ± 98.7	1043.3 ± 100.0	1017.7 ± 111.7	991.8 ± 114.1
Ventricle volume (ml)	38.4 ± 18.1	41.0 ± 21.3	49.1 ± 23.9	51.4 ± 21.9
Hippocampus volume (ml)	7.3 ± 0.9	7.1 ± 1.0	6.3 ± 1.0	6.0 ± 1.0
Precuneus volume (ml)	16.7 ± 2.2	17.4 ± 2.4	16.2 ± 2.6	15.4 ± 2.5
Middle temporal gyrus volume (ml)	20.4 ± 2.7	20.4 ± 2.7	18.5 ± 2.9	17.6 ± 3.0
Fusiform gyrus volume (ml)	17.5 ± 2.1	17.6 ± 2.2	16.3 ± 2.4	15.5 ± 2.4
Entorhinal cortex volume (ml)	4.0 ± 0.7	3.9 ± 0.8	3.4 ± 0.8	3.2 ± 0.8
Time before AD diagnosis (years)*			1.4 ± 0.7	

B) RS dataset - generalizability set	**CN**	**non-AD**	**incident-AD**	**prevalent-AD**
Number of subjects	998	2710	97	25
Number of women, %	500, 50.1	1200, 44.3	39, 40.2	10, 40.0
Age (years)	67.4 ± 8.3	70.9 ± 9.3	79.6 ± 5.7	80.2 ± 6.3
Number of *APOE* ε4 carriers, %	255, 25.6	745, 27.5	45, 46.4	11, 44.0
Intracranial volume (ml)	1512.3 ± 157.6	1475.8 ± 155.3	1437.5 ± 156.6	1403.0 ± 163.9
Total brain volume (ml)	1050.3 ± 107.5	1012.6 ± 105.6	936.6 ± 94.9	884.5 ± 105.0
Ventricle volume (ml)	33.7 ± 17.3	36.5 ± 19.3	49.1 ± 21.1	59.9 ± 28.3
Hippocampus volume (ml)	7.9 ± 0.8	7.6 ± 0.8	6.7 ± 0.9	6.0 ± 1.0
Precuneus volume (ml)	18.2 ± 2.1	17.6 ± 2.0	16.8 ± 1.9	15.4 ± 2.2
Middle temporal gyrus volume (ml)	20.6 ± 2.7	19.9 ± 2.7	17.6 ± 2.5	16.2 ± 2.7
Fusiform gyrus volume (ml)	17.7 ± 2.2	17.2 ± 2.1	15.8 ± 2.0	14.5 ± 2.7
Entorhinal cortex volume (ml)	3.7 ± 0.6	3.6 ± 0.7	3.1 ± 0.8	2.6 ± 0.7
Time before AD diagnosis (years)*			2.8 ± 2.3	

C) RS dataset - prediction set	**CN**	**non-AD**	**incident-AD**
Number of subjects	183	852	31
Number of women, %	95, 51.9	412, 48.4	10, 32.3
Age (years)*	73.3 ± 5.5	75.5 ± 6.4	78.4 ± 6.8
Follow-up time (years)	3.5 ± 1.3	3.5 ± 1.4	2.9 ± 0.9
Number of *APOE* ε4 carriers, %	39, 21.3	225, 26.4	13, 41.9
Intracranial volume (ml)*	1522.8 ± 156.6	1478.9 ± 156	1419.4 ± 126.9
Total brain volume (ml)*	1038.7 ± 100.7	998.4 ± 98.3	926.6 ± 91.4
Ventricle volume (ml)*	39.7 ± 20.2	41.1 ± 21.6	44.9 ± 17
Hippocampus volume (ml)*	7.8 ± 0.8	7.4 ± 0.8	6.7 ± 0.9
Precuneus volume (ml)*	18.0 ± 2.0	17.5 ± 1.9	16.5 ± 1.8
Middle temporal gyrus volume (ml)*	20.3 ± 2.6	19.5 ± 2.4	17.5 ± 2.4
Fusiform gyrus volume (ml)*	17.5 ± 2.1	17.0 ± 2.1	15.6 ± 2.1
Entorhinal cortex volume (ml)*	3.7 ± 0.7	3.6 ± 0.7	3.0 ±0.6
Time before AD diagnosis (years)*			2.4 ± 1.8

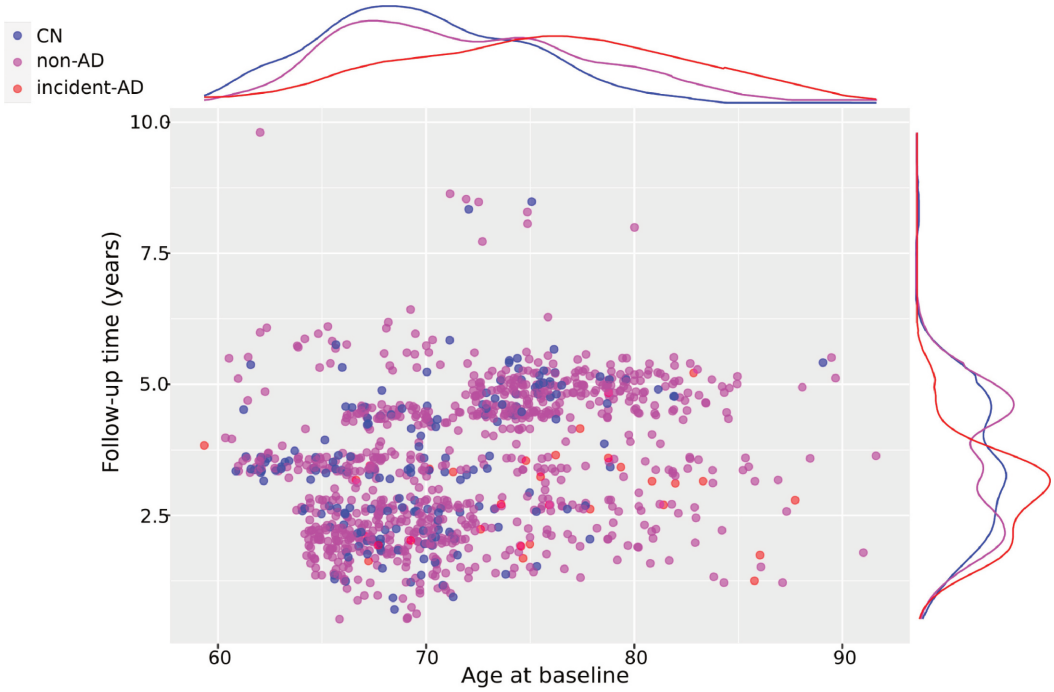


Figure 1. Longitudinal sampling in the prediction set of the RS dataset. The x-axis represents the age of the participant at baseline and the y-axis represents time difference between the baseline and follow-up scan. The plot on the top of the figure shows the kernel density estimates of the age of the participants for the different diagnostic classes and the one on the right shows the kernel density estimates of the follow-up time.

Experiment 1: Assessing the generalizability of co-init DEBM from ADNI to RS

The normalized histograms of the estimated APOE-specific disease stages for the different diagnostic classes in ADNI and the generalizability set of RS are shown in Figure 3. It can be seen that the distributions of the disease stages of the four diagnostic classes in ADNI were largely similar to those in the generalizability set of RS. The CN and non-AD subjects were positioned towards the left side of the spectrum, whereas the prevalent-AD were positioned predominantly towards the right. It can also be seen that for a proportion of prevalent-AD subjects in the APOE ε4 non-carrier group, the model had estimated a low disease stage in both ADNI and RS cohorts. A noticeable difference between ADNI and RS was that a substantial proportion of incident-AD subjects in RS was positioned towards the left side of the histograms in both APOE ε4 non-carriers and carriers.

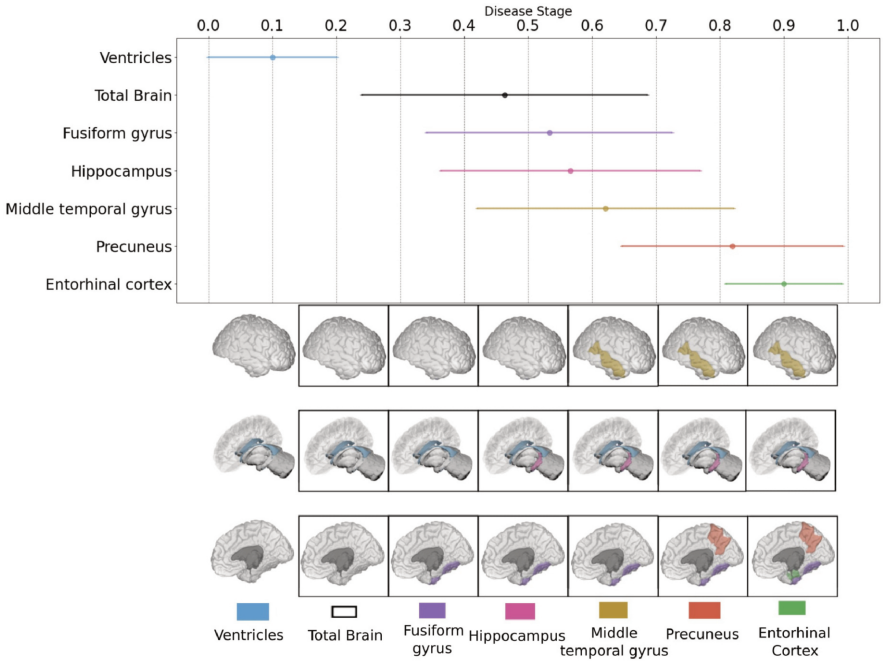
The AUCs for distinguishing the different diagnostic classes using the estimated disease stages are shown in Table 2, along with their confidence intervals. It can be

Table 2. Generalizability assessment: The AUCs for distinguishing the different diagnostic classes using the estimated disease stages and their corresponding 95% confidence intervals. The confidence intervals of the AUCs were determined using bootstrap resampling while stratifying the diagnostic classes to maintain their relative proportions. Co-init DEBM AUC represents the AUCs obtained when separate disease timelines were estimated for APOE ε4 non-carriers and carriers, whereas DEBM AUC represents the AUCs obtained when a combined disease timeline was estimated. N_R and N_C represent the number of subjects in the reference group and number of cases respectively.

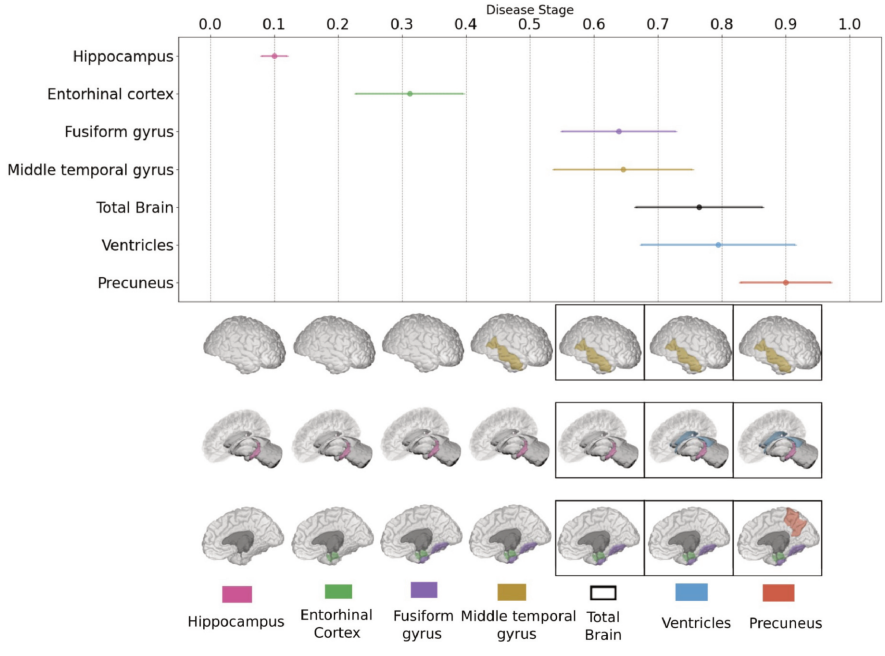
	Cases	No. of Subjects		Co-init DEBM AUC		DEBM AUC	
Reference group		**ADNI (N_R,N_C)**	**RS (N_R,N_C)**	**ADNI**	**RS**	**ADNI**	**RS**
APOE ε4 non carriers							
CN	prevalent-AD	243, 72	743, 14	0.86 (0.81-0.91)	0.85 (0.71-0.98)	0.85 (0.80-0.90)	0.79 (0.64-0.94)
CN + non-AD	prevalent-AD	570, 72	2708, 14	0.83 (0.78-0.88)	0.83 (0.70-0.97)	0.81 (0.76-0.86)	0.79 (0.64-0.94)
CN	incident-AD	243, 46	743, 52	0.83 (0.77-0.90)	0.70 (0.62-0.78)	0.83 (0.77-0.89)	0.63 (0.55-0.72)
CN + non-AD	incident-AD	570, 46	2708, 52	0.81 (0.74-0.88)	0.68 (0.60-0.75)	0.80 (0.73-0.86)	0.64 (0.55-0.72)
APOE ε4 carriers							
CN	prevalent-AD	92, 151	255, 11	0.89 (0.85-0.94)	0.85 (0.74-0.96)	0.92 (0.87-0.96)	0.84 (0.71-0.97)
CN + non-AD	prevalent-AD	330, 151	1000, 11	0.83 (0.79-0.86)	0.85 (0.74-0.95)	0.83 (0.80-0.87)	0.85 (0.72-0.98)
CN	incident-AD	92, 121	255, 45	0.87 (0.82-0.92)	0.63 (0.54-0.72)	0.88 (0.83-0.93)	0.62 (0.52-0.72)
CN + non-AD	incident-AD	330, 121	1000, 45	0.79 (0.74-0.83)	0.62 (0.54-0.71)	0.79 (0.74-0.83)	0.62 (0.52-0.72)

observed that the performance of the disease stages obtained using co-init DEBM in distinguishing prevalent-AD from the set of CN and non-AD subjects in ADNI (AUC = 0.83 for both APOE ε4 non-carriers and carriers) was comparable to that in RS (AUC = 0.83 for APOE ε4 non-carriers and AUC = 0.85 for 4 carriers). It should however be noted that the confidence intervals were larger in the RS cohort. It can also be observed that incident-AD subjects were harder to distinguish than prevalent-AD in the RS cohort (Co-init DEBM: AUC = 0.68 for ε4 non-carriers and AUC = 0.62 for ε4 carriers), but not in ADNI (Co-init DEBM: AUC = 0.81 for ε4 non-carriers and AUC = 0.79 for ε4 carriers). It can also be seen in Table 2 that, while AUCs in ADNI are comparable for both DEBM and Co-init DEBM and in APOE ε4 carriers in RS, the AUCs for distinguishing the different groups in RS APOE ε4 non-carriers is higher for co-init DEBM.

Furthermore, the estimated disease stages showed a significant Pearson correlation with time to diagnosis for APOE ε4 carrier incident-AD subjects in both ADNI ($R = 0.31$, $p = 0.0006$) and RS cohorts ($R = 0.29$, $p = 0.04$). However, the correlation was found to be insignificant for APOE ε4 non-carrier incident-AD subjects in both ADNI ($R = 0.04$, $p = 0.8$) and RS cohorts ($R = 0.1$, $p = 0.4$). Lastly, the obtained



(a)



(b)

Figure 2. Disease timelines of APOE ε4 non-carriers (a) and carriers (b) estimated using co-init DEBM in ADNI. The plot on top of each subfigure shows the event-centers of the different regions and their respective standard deviation estimated from a batch of 100 independent bootstrap samples. The 3D visualization (Marinescu et al., 2019) at the bottom of each subfigure highlights the region that becomes abnormal at the corresponding disease stage. Total brain volume becoming abnormal is depicted by a black-box surrounding the brain at the corresponding disease stage. The vertical positioning of the biomarkers in the event-center part of each subfigure shows the estimated disease timeline in the APOE genotype, which is different for non-carriers and carriers.

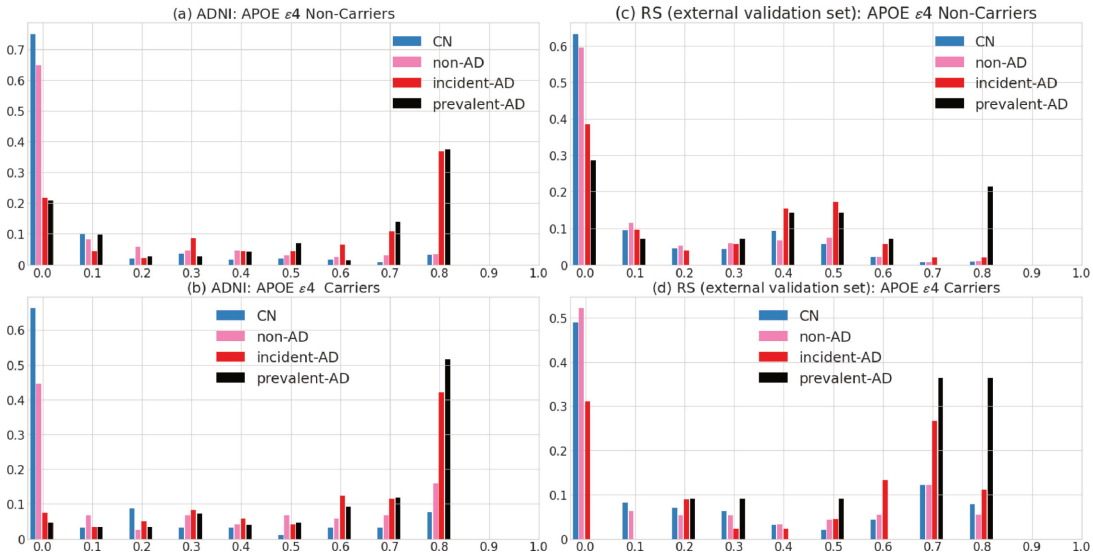


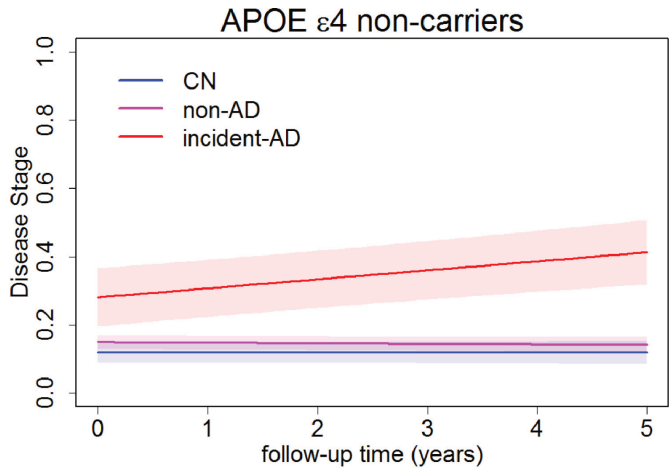
Figure 3. Normalized histograms of the estimated APOE-specific disease stages for the different diagnostic classes in ADNI and the generalizability set of RS. The normalized histograms of disease stages are shown for (a) APOE ε4 non-carriers in ADNI, (b) APOE ε4 carriers in ADNI (c) APOE ε4 non-carriers of the generalizability set in RS, and (d) APOE ε4 carriers of the generalizability set in RS. The x-axis represents the disease stage based on the APOE-specific disease timeline by the co-init DEBM model, and the y-axis represents the relative percentage of subjects in each diagnostic class, meaning that the relative percentages of all disease stages of one diagnostic category add up to one. Estimated disease stage is a continuous variable and was discretized (binned) for visualization purposes only.

disease stages had a significant Spearman correlation with MMSE in both ADNI non-carriers ($R = -0.41$, $p < 0.001$) and carriers ($R = -0.48$, $p < 0.001$) as well as in RS non-carriers ($R = -0.08$, $p < 0.001$) and carriers ($R = -0.06$, $p = 0.05$).

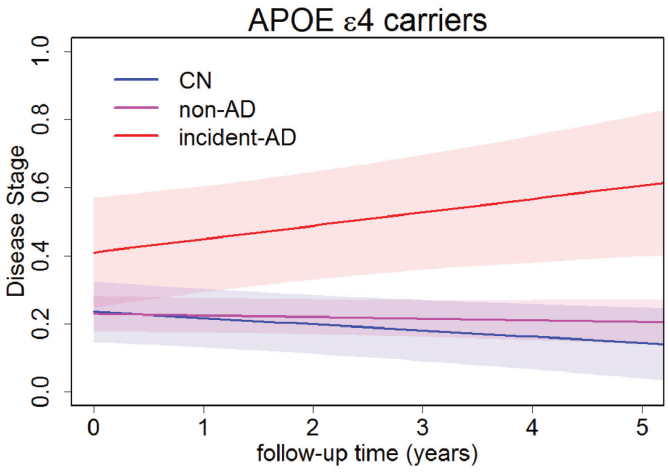
Experiment 2: Predicting AD based on longitudinal data in the RS cohort

Exp. 2.1: In Figure 4, the trajectories of disease stage over time as estimated by linear mixed models are shown for the CN, non-AD and incident-AD groups of the prediction set of RS. The interaction between the incident-AD diagnosis and follow-up time was statistically significant in both APOE ε4 non-carriers and carriers (CN vs. incident-AD $p = 0.0032$ and $p = 0.0041$ respectively; non-AD vs. incident-AD $p = 0.0039$ and $p = 0.0032$ respectively), meaning that incident-AD subjects showed a significant increase in disease stage compared to CN and non-AD subjects.

Exp. 2.2: In the left column of Figure 5, the AUCs and the corresponding 95% confidence intervals for distinguishing incident-AD using two MRI scans based on longitudinal follow-up of participants are shown for APOE ε4 non-carriers and carriers. It can be observed that for distinguishing incident-AD from the reference group, delta disease stage consistently performed the best for both the genotypes. It outperformed the rates of changes of volumetric measures, with respect to the



(a)



(b)

Figure 4. Average disease stage trajectories of participants within the prediction set of RS. The trajectories are shown separately for CN, non-AD and incident-AD subjects within the *APOE* ε4 non-carriers group (a) and the *APOE* ε4 carriers group (b). 95% confidence intervals are shown as shaded regions around the trajectories.

obtained AUC. It can also be observed that distinguishing incident-AD from CN and non-AD subjects in the reference group was harder than distinguishing incident-AD from CN alone, as reflected by the lower AUCs for almost all the measures used.

Exp. 2.3: The right column of Figure 5 shows that age was an important predictor for incident-AD. Age distinguished incident-AD well from CN subjects (AUC of 0.73 for both ε4 non-carriers and carriers), but the performance of age as a predictor

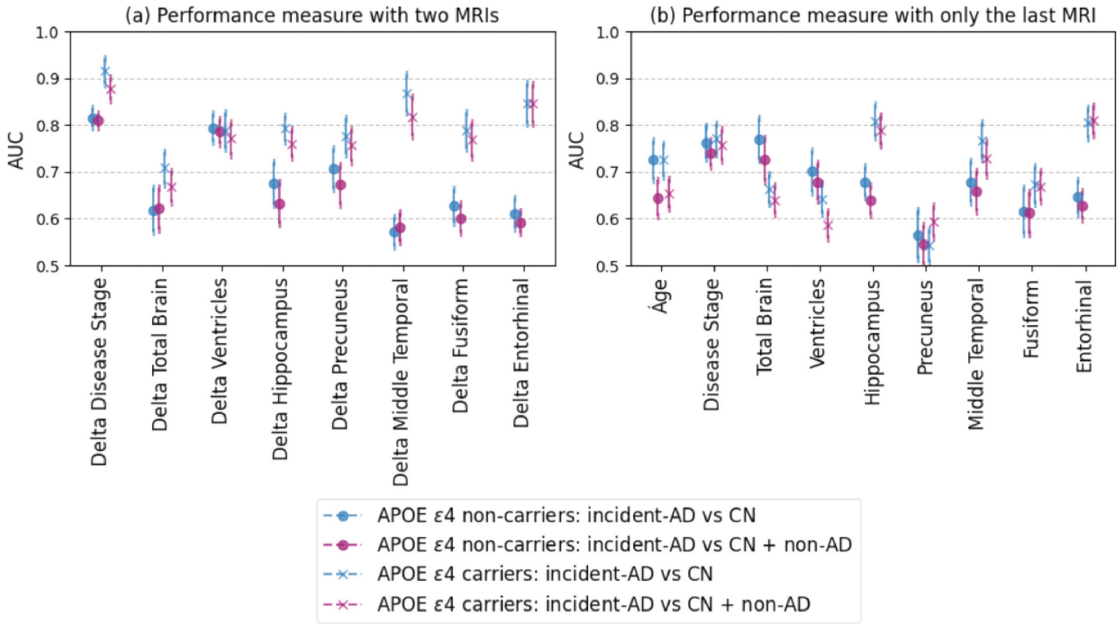


Figure 5. Predicting incident-AD subjects in the RS cohort. Figure (a) shows the AUCs for distinguishing incident-AD while using data from two MRI scans based on longitudinal follow-up of the participants. Figure (b) shows the AUCs for distinguishing incident-AD using only the last MRI scan available for each participant.

dropped substantially when distinguishing incident-AD from CN and non-AD subjects (AUC of 0.64 for ε4 non-carriers and 0.65 for ε4 carriers). When only the last MRI scan was used for incident-AD prediction from a reference group of CN and non-AD subjects, volumes of hippocampus and entorhinal cortex were good indicators in APOE ε4 carriers (AUC of 0.79 and 0.81 respectively) but not for APOE ε4 non-carriers (AUC of 0.64 and 0.63 respectively). Similarly, total brain volume and ventricle volume were good indicators of incident-AD in APOE ε4 non-carriers (AUC of 0.73 and 0.68 respectively), but not for ε4 carriers (AUC of 0.64 and 0.59 respectively). Disease stage estimated using the APOE-specific disease timeline performed well consistently in both the APOE genotypes (AUC of 0.74 for ε4 non-carriers and 0.76 carriers). The marginal utility of an additional MRI scan can be observed by comparing the left column of Figure 5 with the right column of Fig. 5. It can be seen that delta disease stage was much better for incident-AD prediction from a reference group of CN and non-AD subjects (AUC of 0.81 for ε4 non-carriers and 0.88 for carriers) than any measure obtained using only the last MRI scan.

Discussion

In this work, we constructed APOE-specific disease timelines in a case-controlled setting and validated their generalizability to a population-based setting. We assessed that progression along these timelines is predictive of AD in the general population. In this section, we discuss the insights we obtained from our results.

Generalizability of the APOE-specific disease timelines

The disease timelines estimated for APOE ε4 non-carriers and carriers were significantly different from one another and highlighted the APOE-genotype-specific differences in the loss of structural integrity as AD progresses. Ventricular volume and total brain volume were early biomarkers for ε4 non-carriers, and hippocampal volume and volume of the entorhinal cortex were early biomarkers for ε4 carriers. We observed in the normalized histograms that for a proportion of prevalent-AD subjects in the ε4 non-carriers group, the model had estimated a low disease stage. This observation, in combination with the greater uncertainty of the event-centers in that group suggests that there is intra-genotype heterogeneity among the ε4 non-carriers.

The disease timelines were estimated after correcting for the confounding effect of age, assuming a linear relationship of volumetric biomarkers with respect to age. Non-linear biomarker relationship with age such as the one observed in Vinke et al.[83], could have an adverse effect in the generalizability of the model to the RS cohort, particularly due to the observed differences in the mean age of the participants in the reference group and the groups of incident-AD and prevalent-AD. In spite of these differences, we observed that the normalized histograms of disease stages in the different diagnostic classes were visually largely similar for ADNI and RS. An important difference between the two cohorts was that the model estimated a low disease stage for a substantial proportion of incident-AD subjects in RS, but not in ADNI. Complementing the qualitative analysis, we also observed that the disease stages obtained using co-init DEBM could distinguish prevalent-AD subjects from CN and non-AD subjects almost equally well in both ADNI and RS cohorts. However, we noticed a lower performance in distinguishing incident-AD from CN and non-AD subjects in RS as compared to ADNI. Three possible explanations for these differences between ADNI and RS are given below.

First, the incident-AD group in ADNI only consisted of prodromal AD subjects with the mean time to AD diagnosis of 1.4 years, whereas the incident-AD group in RS consisted of prodromal and preclinical AD subjects with the mean time to AD diagnosis of 2.8 years. We observed in Experiment 1 that the obtained disease stages

of incident-AD subjects correlated with time to AD diagnosis for APOE ε4 carriers, making AD harder to detect in the preclinical phase than in the prodromal phase. Hence the difference in the mean time to diagnosis in the two datasets is expected to be a factor contributing to the observed lower performance in the RS cohort.

Secondly, the prodromal AD subjects in ADNI were clinically defined amnestic MCI subjects who have a much higher a priori chance of developing AD symptoms than in the general population, making the prediction in the latter cohort a more difficult problem.

Thirdly, a factor contributing to the performance difference could be that ADNI excluded subjects with severe cardiovascular risk factors whereas the RS did not. Hence the probability of co-morbidity of vascular pathology was higher in the RS incident-AD subjects than in the corresponding ADNI set, which could have led to the drop in performance.

In spite of these factors, biologically, one could expect a Normal distribution of AD severity among incident-AD subjects in a population, whereas the observed distribution in Figure 3 is not. A possible explanation for this apparent anomaly is that, although the biological progression of AD is heterogeneous with differences between subjects even within each genotype, the staging is performed on the basis of a mean disease timeline per genotype. Progression of subjects that is not along the estimated mean disease timeline is not accounted for in our approach, and the corresponding stages are usually an under-estimation of the true biological staging.

We observed that the correlation of MMSE with the obtained disease stages in RS was substantially lower than that in ADNI. One of the reasons for this lower correlation in RS could be that MMSE is a measure of general cognition, not specific to AD and there could be numerous other factors affecting its value in a population-based cohort. Furthermore, in AD the MMSE range is expected to be much broader than in the general population. This in combination with less prevalent-AD cases within RS compared to that of ADNI, could explain the lower correlation as well. Moreover, the correlation of MMSE with the obtained disease stage was similar for non-carriers and carriers as measured along their respective disease timelines.

Given the high AUCs for all other classification tasks, the comparable disease stage histograms in ADNI and RS, and the possible explanations given above for the specific differences related to incident-AD prediction, we conclude that the APOE-specific disease timelines obtained by co-init DEBM are generalizable from a case-controlled to a population-based setting. Moreover, in our experiment, we observed that co-init DEBM was more generalizable to RS cohort than DEBM for APOE non-

carriers, and equally generalizable for ε4 carriers. This could be because ADNI cohort is more enriched for ε4 carriers than the population-based RS cohort and not stratifying based on APOE skewed the estimated timeline more towards ε4 carriers.

However, for precise classification of subjects into either diagnostic category, a cut-off point for disease stage needs to be defined. We expect the cut-off point to be different in a case-controlled setting versus a population-based setting. Estimating this cut-off point in a population should ideally be estimated using an independent validation set taking several factors into consideration such as the a-priori prevalence of AD in the cohort, and the risks associated with false positives and negatives of this classification.

Predicting pre-symptomatic AD in the RS cohort

We observed that a participant's age distinguished incident-AD well from CN. This is in line with earlier studies that identified age as an important predictor.[334,335] However, we also observed that the predictive performance of age deteriorated when the reference group was less healthy, i.e., when distinguishing incident-AD from a combined reference group also consisting of subjects with subjective or objective cognitive decline unrelated to AD. This is in line with the expectation that age is poor in distinguishing cognitive decline due to AD and cognitive decline due to other causes.

The predictive performance of the volumetric biomarkers from a single MRI scan depended on the APOE ε4 carriership. We observed that hippocampus and entorhinal cortex were good predictors in APOE ε4 carriers. Interestingly, those biomarkers were estimated to be early in the corresponding disease timeline. Similarly, total brain volume and ventricle volume were good predictors in APOE ε4 non-carriers which were also the early biomarkers in its disease timeline. These results suggest that for predicting pre-symptomatic AD, early biomarkers play an important role and that it is important to understand the genotype-specific differences. However, it must be noted that in this study, the clinical diagnosis of AD was not confirmed further with the participant's amyloid-β status. Hence part of the differences observed in the disease timelines of APOE ε4 non-carriers and carriers could be attributed to the presence of greater heterogeneity in the non-carriers with respect to participant's pathologic diagnosis.

Lastly, we assessed the marginal utility of longitudinal MRI scans in identifying individuals at-risk of developing AD symptoms. We observed that participants with incident-AD showed a significant increase ($p < 0.005$) in disease stage over time as compared to CN and non-AD participants, in both APOE ε4 non-carriers and

carriers. The rate of change of disease stage distinguished incident-AD subjects better than the disease stage at only the last scan, clearly highlighting the added value of longitudinal MRI scans, particularly in pre-symptomatic subjects. The rate of change of disease stage was also a better predictor of incident-AD than any other volumetric biomarker used in this study. This showed that the progression along the APOE-specific disease timeline can be used to identify subjects in a population at-risk of developing AD.

In this study, we only used imaging biomarkers because cerebrospinal fluid biomarkers in a pre-clinical setting are usually not available. Recent breakthroughs in blood-based biomarkers[336] could help in obtaining fluid biomarkers in the pre-clinical phase of the disease. Previous work on DEBM[326] and co-init DEBM[327] had shown that the model is capable of incorporating biomarkers from multiple modalities for constructing the disease timelines. We expect that our current approach of predicting pre-symptomatic AD in the general population would be applicable also in the presence of fluid biomarkers, should they become available in the future.

Conclusion and future work

We conclude that data-driven disease timelines estimated by co-init DEBM are generalizable to population-based cohorts and that progression of individuals along such timelines is predictive of incident AD. Although the current study only considered volumetric biomarkers as inputs, it can be extended to fluid-based biomarkers, if these would become available in a population based study. Due to its robustness and explainability, we expect that our model can help identify at-risk individuals from the general population for targeted clinical trials as well as provide biomarker based objective assessment in such trials.

4

5

General Discussion

The aim of this thesis was to gain insight in the brain aging patterns, to ultimately improve our understanding of the continuum of aging and age-related diseases. In this chapter I describe the main findings and place them in a broader perspective. Furthermore, I discuss the methodological considerations, implications and future perspectives.

Main findings

In this thesis I investigated and described brain aging based on the following three objectives:

1. Quantification of brain aging trajectories of structural imaging markers, cognitive function and motor function, and the brain aging patterns based on these markers. (**Chapter 2: Quantifying brain aging trajectories**)
2. Determining the association between risk factors and the brain aging trajectories. **(Chapter 3: Determinants of brain aging)**
3. Determining how brain aging can inform disease assessment and prediction. (**Chapter 4: Clinical implications of brain aging**)

Quantifying brain aging trajectories

To understand how different changes in the brain coincide and interact in aging, it is essential to have longitudinal data of a broad range of markers, per subject. To this end, in Chapter 2.1 en 2.2 I estimated trajectories of imaging markers, cognitive and motor function based on longitudinal data and used these to estimate the brain aging patterns. The age at which a marker was considered to have changed, was based on the change of the marker exceeding a threshold based on the variance of that marker. The resulting sequence of changing markers can be interpreted as an average sequence of the broad spectrum of aging in the population. In accordance with literature, the trajectories of brain imaging markers often showed a non-linear curve with accelerating change with advancing age.[16,27,30,32,34,49–52,337] The brain aging pattern, based on a change exceeding 2 standard deviations after the age of 45, showed that total brain volume changes first, followed by mean diffusivity, white matter, hippocampus, and total grey matter and fractional anisotropy being the last markers to change (Chapter 2.1). Cognitive and motor function trajectories generally decline linearly between the ages of 45 and 65 years, followed by a steeper decline after the age of 65–70 years. Test scores for cognitive and motor function declined similarly, with high variation in the rate of decline across age for individual tests. No distinct patterns of an overall decline in cognitive function preceding motor function or vice versa was found (Chapter 2.2). Global gyrification of the cerebral cortex

decreased linearly with age during adulthood and gyrification in the medial prefrontal cortex increases towards the end of life. Gyrification was associated with higher levels of cognitive performance in some local regions of the cortex irrespective of surface area (Chapter 2.3). In Chapter 2.4, the perspective on brain aging changed from estimating the average brain aging pattern, to identifying potential different brain aging patterns present in the population. I found two patterns of structural brain changes with aging, which can be considered as two subtypes of brain aging. The first subtype is characterized by early (midtemporal) cortical atrophy, whereas the second subtype is characterized by early ventricle enlargement and total brain atrophy. Since the white matter subtype showed both more white matter atrophy and a higher burden of focal cerebral small vessel disease markers, as well as lower scores on cognitive function tests, the white matter subtype may present a less successful brain aging subtype.

Determinants of brain aging

In Chapter 3 I used trajectories of different markers to investigate if and how certain determinants alter these brain aging trajectories. In Chapter 3.1 I found that a larger burden of intracranial carotid artery calcification and vertebrobasilar artery calcification is associated with faster increase of cerebral small vessel disease markers over time. In Chapter 3.2 I showed that worse cardiovascular health was associated with a faster increase in white matter hyperintensity volume with advancing age and lower hippocampus volume. Furthermore I found that APOE ε4 genotype was associated with more microbleeds and faster increase in white matter hyperintensities with advancing age. In Chapter 3.3 I investigated whether hearing loss, a promising modifiable risk factor for cognitive decline and dementia,[121,247] accelerates cognitive decline. In contrast to other population-based studies,[250,253] I showed that hearing loss was not associated with accelerated cognitive decline. The association between hearing loss and accelerated cognitive decline was driven by the non-linear relationship between cognitive function and age. This underlines the importance of understanding brain aging and cognitive aging, when investigating the effects of risk factors.

Clinical implications of brain aging

In Chapter 4 I focussed on determining how brain aging can inform disease assessment and prediction. In Chapter 4.1 I estimated percentile curves of subcortical structures based on three different reference populations, using a single segmentation tool. I found that percentile curves of subcortical structures, based on the different reference populations largely overlap. This indicates that there are only small differences between the subcortical volumes of these populations, regardless

of differences in acquisition, field strength, vendors, and population differences. Furthermore, this indicates that in a clinical setting, individual diagnostic assessment based on subcortical volume, may not be biased by the use of a specific reference population. However, the use of the same segmentation tool is crucial here. Several studies have shown differences in segmentation methods where they evaluated accuracy, correlation, absolute or relative differences of different brain regions.[338–345] A recent study investigated the impact of different brain region segmentation methods on single-subject analysis, where they concluded that for single-subject analysis, it is essential that the same method is used to generate normative volume distributions and patient-specific volumes.[346] In Chapter 4.2 I used a data-driven disease progression model to construct APOE-specific Alzheimer's disease related disease timelines based on structural imaging markers in a case-control setting. The data-driven disease timelines were generalizable to population-based cohorts and the progression of individuals along such timelines were predictive of incident Alzheimer's disease.

Methodological considerations

The strengths and shortcomings of the individual studies presented in this thesis have been discussed in the corresponding chapters. Therefore, in this section I discuss four more general topics which are essential in brain aging research and apply to the research presented in this thesis.

Definition of 'normal' aging

The definition of what is normal and what is normal aging in research is highly dependent on the context of the research that is performed. Defining what is normal is often not the main goal, merely a necessity to enable research focussing on disease. However, even within research focussing on aging itself, there is no consensus on what we consider normal aging. Brain aging is especially of interest in the context of neurological and neurodegenerative diseases, therefore a population without neurological diseases as 'normal' brain aging reference is essential. Therefore, in this thesis normal brain aging is defined as brain aging in a population free of overt neurological disease. However a population free of overt neurological disease, may still contain people who are in a pre-clinical phase of the neurological disease. Therefore this definition of 'free of disease' depends on the diagnostic criteria of these diseases. In this thesis, scans of participants with a diagnosis Parkinson's disease (and parkinsonism), dementia or clinical stroke were excluded from analyses in which brain aging trajectories were estimated. These diseases have a large impact on brain structure ánd functioning. Scans from participants that were in the pre-clinical phase were still included in the study population. The impact of

the inclusion of the pre-clinical phase within brain trajectories depends on the prevalence of the disease, the length of the pre-clinical phase, the age of diagnosis and the progression pattern of changes in the marker(s) of interest during the pre-clinical phase. With the pathophysiological process of Alzheimer's disease starting years if not decades before clinical diagnosis, presumably the pre-clinical phase of neurodegenerative diseases increases the width of the normal brain aging spectrum. Furthermore, it could result in an overestimation of the deterioration of brain structure and function in brain aging, close to the average age of diagnosis of these diseases. As was shown in Chapter 2.1, above the age of 70 more than 20% of 'healthy' individuals have one or more microbleeds, and almost 10% have one or more cortical infarcts visible on their brain MRI. Although these participants were free of neurodegenerative disease, these structural brain changes are considered pathological and perhaps a pre-clinical phase of disease, yet remain part of the spectrum of normal brain aging. In Chapter 2.4 these lesions seem to be more present in the white matter brain aging subtype, which we therefore consider a potentially less successful brain aging subtype.

Though I believe that within aging research we will remain dependent on the definition of disease, it is essential to be aware of the dynamic nature of this definition and the impact of this definition on the width of the normal aging spectrum. Subsequently the width of the normal aging spectrum should be kept in mind when choosing methodological approaches to investigate brain aging and interpreting the results.

Spectrum of brain aging

Our view on brain aging goes hand in hand with our definition of normal aging. The broader the definition of normal, the broader the spectrum of normal brain aging. When you assume that brain aging is a single homogeneous process, the spectrum of brain aging is a result of temporal heterogeneity, i.e. different phases in the process occurring at different times within individuals. The single homogeneous process could be captured by estimating the average aging trajectories of a broad range of markers. The variation in the speed in which each of these trajectories occur between individuals result in the broad spectrum of brain aging. The extremes of the spectrum could then be considered as accelerated or delayed brain aging. This perspective of brain aging underlies studies that estimate the biological brain age and the corresponding brain age gap of an individual. However, concerns regarding the assumption that brain aging is a single homogeneous process with some temporal heterogeneity are revealed when comparing the brain aging patterns in Chapter 2.1 and Chapter 2.4. The average brain aging pattern, extracted from brain aging trajectories of a range of imaging markers, is not equal to the most frequent brain

Box 1: Temporal and phenotypic heterogeneity

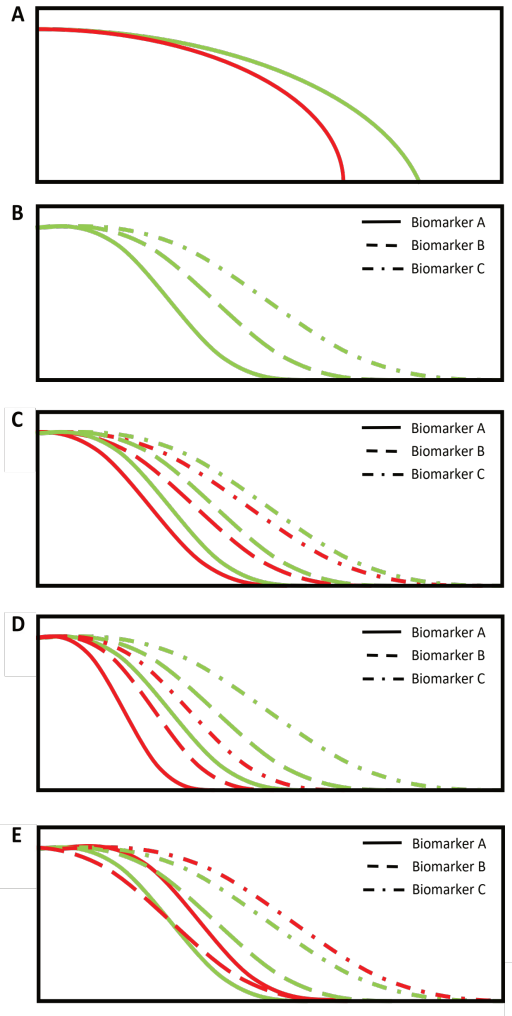


Figure A shows two hypothetical courses of cognitive functioning in aging: 'successful' (green) and 'unsuccessful' (red) cognitive aging. When we assume that these courses of cognitive aging result from changes in 3 different biomarkers (biomarker A, B and C). **Figure B** presents the changes in these biomarkers with age which belong to the 'successful' cognitive aging course. Biomarker A changes the fastest, followed by biomarker B and finally biomarker C. When you assume that the cognitive aging course is driven by a single pattern of changes in biomarkers, with differences in timing and rates of change in these markers, Figure C and D show two forms of temporal heterogeneity. In the biomarker trajectories of unsuccessful aging in **Figure C**, the pattern and the rates at which the biomarker change are identical to those in successful aging. However the age of onset at which these changes start is a few years earlier. In the biomarker trajectories of unsuccessful aging in **Figure D** on the other hand, the pattern and the age of onset of the changes in biomarkers is identical to those in successful aging. However the rate at which the biomarkers change is higher in unsuccessful aging. Note that in both Figure C and D, the relative distances between the biomarkers within successful and unsuccessful aging are identical. **Figure E** on the other hand, presents a situation in which these relative distances between biomarkers are not identical between successful and unsuccessful aging. This is called phenotypic heterogeneity.

aging pattern within the population. The average brain aging pattern estimated in Chapter 2.1, showed most resemblance to the subtype that occurred in approximately 30% of the population (as shown in Chapter 2.4). This elucidates that studying brain aging trajectories of a broad range of markers individually, only provides us a small window into the complex brain aging process. Though considering brain aging as a single homogeneous process seems to be an oversimplification of the brain aging process, it is very likely that the two subtypes described in Chapter 2.4 are still an oversimplification of the actual phenotypic heterogeneity of brain aging. In Box 1 the interpretation of temporal and phenotypic heterogeneity in the context of this thesis is further explained.

The scope of phenotypic heterogeneity of brain aging that we can capture, depends on the types of information on which the brain aging phenotypes are based. For example, Smith et al.[193] identified 62 different brain aging patterns, so called 'modes', based on 6 imaging modalities. These modes represent different aspects of brain aging, showing distinct patterns of structural and functional brain change. In addition, these 62 modes were grouped into 6 clusters, to help understand larger patterns of brain aging. Another example of an approach to understand brain aging patterns, is to examine brain aging patterns based on potential underlying mechanisms. Garbarino et al. investigated the brain atrophy pattern in aging, Alzheimer's disease and multiple sclerosis by linking brain atrophy and connectivity.[347] With these different types of information and approaches to estimate brain aging patterns, each combination of type of information and approach provides one window into the brain aging process. The challenging task is not identifying brain aging patterns in itself, but how to relate the found patterns to each other and to combine these different pieces of information to create a full picture of the brain aging process.

The scope of phenotypic heterogeneity also highly depends on the age range included. One could argue whether brain development and brain aging should be considered as two separate fields of research or that perhaps it is essential to investigate phenotypic and temporal heterogeneity across the lifespan, to understand brain aging. Taking the brain aging subtypes identified in Chapter 2.4 as an example, I cannot exclude that differences in brain structure before the age of 45 have a neurodevelopmental origin. The same holds for the results presented in Chapter 2.1, I cannot conclude whether the sex differences in the imaging marker trajectories are mostly driven by temporal heterogeneity, phenotypic heterogeneity or both. Capturing brain trajectories over the entire lifespan, as was recently done by Bethlehem et al.[337] in the largest synthesis of brain imaging data to date, can help us explain the variation in brain trajectories between brain regions and individuals.

Resilience against brain aging

Three important concepts in brain aging and cognitive aging, which capture the 'resilience' against age- and disease-related changes are: cognitive reserve, brain reserve and brain maintenance.[348] These concepts emerged from the disjunction between the degree of observed brain changes or pathology and the clinical manifestation of those brain changes. Cognitive reserve refers to the adaptability of cognitive processes, which explains differential susceptibility of cognitive functioning to brain aging or pathology. Brain reserve is commonly conceived as neurobiological capital, such as number of neurons or synapses, which allows to better cope with brain aging and pathology, before clinical or cognitive changes emerge. Brain maintenance is defined as reduced development over time of age-related brain changes and pathology based on genetics or lifestyle. In this thesis I have described brain aging trajectories of both brain structure, cognition and motor function, however the interrelations between structural brain changes and cognition or motor function were outside the scope of this thesis. This does not alter the fact that what in this thesis is described as variability within brain aging could also be interpreted as variation in these concepts of resilience.

Brain aging in relation to etiological and prediction research

From an epidemiological perspective, two large research domains in biomedical research are considered: prediction and etiological research. In general, epidemiologists emphasize that these domains should be considered as separate research areas, since they differ in aim, use and research approach. Where prediction research is focused on recognizing individuals with (early or prodromal) disease or at increased risk of disease and etiological research is focused on identifying causes of disease. Recent advances in machine learning methods in biomedical research, such as deep neural networks and random forest, are increasingly becoming major players in the clinical decision making, diagnosis and medical imaging domains. Machine learning methods have proven to be of added value within these fields. With the primary aim of prediction research being recognizing individuals with disease, machine learning methods improving early diagnosis, clearly impact the prediction research domain. Whether these approaches can reach a similar impact within etiological research is under debate.[349,350] Application of machine learning methods in etiological research is not straightforward, for a number of reasons. With the etiological and prediction domains having different aims, the optimization parameters are different. Reaching the highest accuracy in predicting disease, does not (perse) result in the most accurate effect size estimation of a determinant for that disease. Though optimization parameters can be chosen, the key difference between the two research domains is the role of expert knowledge. Which often plays a

limited role in data-driven prediction models, while within etiologic research the definition and identification of the causal question relies on the substantive knowledge of the expert. Claiming that machine learning methods impact etiological research in a similar way as prediction research, in the eyes of many etiological researchers means claiming that the causal relations structure can be determined based on data alone.

Though I agree with the notion that the causal relationships structure cannot be determined from data alone, categorizing studies according to their research aim does not perse result in avoiding misinterpretation of results, since more and more aspects of both the prediction and etiological framework are used to answer one research question. Taking Chapter 2.4 of this thesis as an example, although the overall aim was to describe brain aging patterns within a population, aspects of the predictive and etiological research framework were used. A data-driven disease progression model was used, which is both predictive of disease progression and informative in terms of revealing potential underlying biological patterns.[190] The resulting model was the result of a balance between imposed knowledge and patterns learned from data. The term 'data-driven' in the context of these type of models does not mean that no prior knowledge of the disease was used to for example select the features included in the model, but refers to the fact that these models do not rely on a-priori classification or staging of individuals. These models may be an example of models originating from the prediction research field, but viewing it as a model solely for optimal prediction does not do it justice. Similarly, there are developments in causal inference which are examples of the causal inference framework meets prediction.[351,352]

I believe it is important to be aware of the fundamental differences between etiological and prediction research, however it is just as important to stay openminded. I believe that machine learning methods can have a similar impact on etiological research, but perhaps in an indirect way, where it can potentially improve the exploratory stage and statistical inference of the causal inference framework. Whether or not we can say that with that role machine learning can infer causality is a matter of definition. Let us not linger too long on these definitions, but focus on where machine learning and other computation approaches can improve etiological research. In fact, the intersection at which the two meet each other may be what is needed to accelerate our understanding of human biology, and may even be a match made in heaven. When etiology and prediction meet, or perhaps even to make sure they do, we must prevent being lost in translation.

With brain aging being so intertwined with age-related neurodegenerative disease, it is clear that our understanding of brain aging impacts both research focussing on

identifying causes of disease, as well as prediction research. A better understanding of brain aging will go hand in hand with a better understanding of disease, and could help us gain new insights in potential determinants of disease. Similarly, a better understanding of the sources of variability in brain aging will impact prediction research, perhaps even in such a way that there will be a paradigm shift of our view on brain aging, in which terms like 'healthy controls' or 'normal controls' are replaced by aging signatures that do cover the complexity of brain aging.

Age as a major confounder in longitudinal brain research

Advanced age is associated with a decline in cognitive and motor function, loss of hearing and vision, as well as with numerous concomitant medical disorders, such as cardiovascular diseases and neurodegenerative diseases. Therefore, when the relation between such an outcome and a determinant is assessed, age is often considered as a major confounder. Correction for the confounder age is often done by integrating age as a linear term into the statistical model. However, when the determinant and outcome are continuous variables, there is often a non-linear relation between age and the specific outcome or determinant. In case both the determinant and outcome have a non-linear relationship with age, correcting for age with only a linear term is insufficient and results in residual confounding by age. In cross-sectional analyses, this could largely be avoided by integrating a non-linear term for age (for example using natural cubic splines) into the model.

In longitudinal models in which the time variable is not age, but for example follow-up time from baseline, integrating a non-linear term for age into the model is insufficient. This is illustrated in Figure 2. Imagine the research question: 'Is determinant X associated with accelerated cognitive decline?' and we want to investigate this by using longitudinal cognition data, and determinant X is measured at baseline. In scenario A of Figure 2, we assume that there is a linear relationship between age and the cognition outcome. If we then look at the cognition outcome of three participants, with three different ages at baseline (magenta, blue and yellow, with age at baseline: 45, 60, 75), we see that with a correction for age at baseline as a linear term, the estimated trajectories of the cognition outcome are linear and parallel to one another. Furthermore, the intercept difference between the participants is based on the age-difference between them (namely $\beta_{\text{Age at baseline}}$*ΔAge at baseline), since the age differences are both 15 years, the intercept difference (at follow-up time=0) in scenario A between magenta and blue, and blue and yellow is equal.

In scenario B of Figure 2 we assume the relation between the cognition outcome and age is non-linear. Here you see that with the non-linear relation between cognition

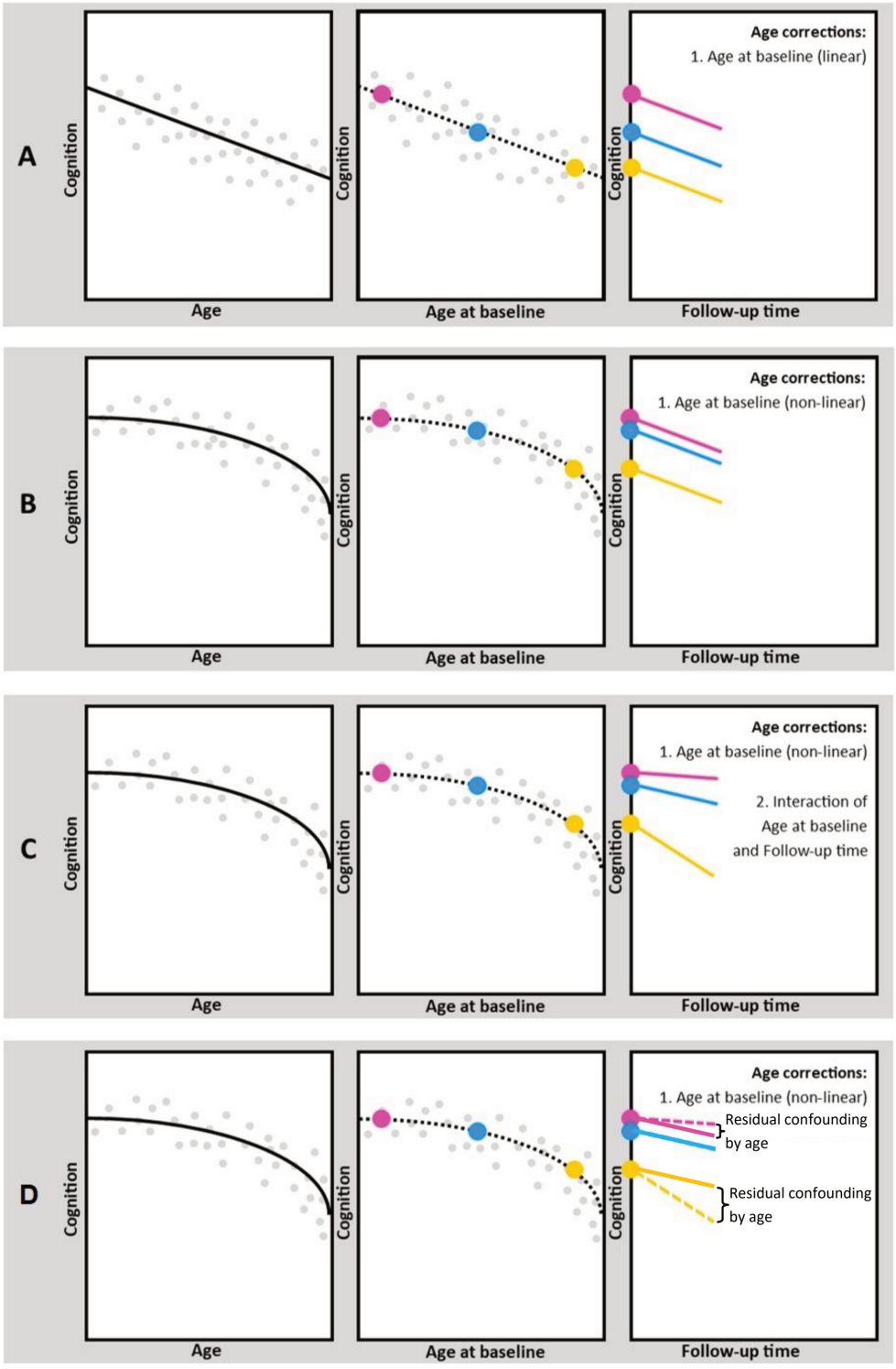


Figure 2. Illustration of residual confounding by age in longitudinal studies.

outcome and age, the cognition outcome value of magenta and blue are very close together, which is visible in the intercept differences of the trajectories with follow-up time. Though the non-linear term for Age at baseline is integrated, the trajectories with follow-up time are still linear and parallel to each other. In scenario C, similarly to scenario B, we assume that there is a non-linear relationship between the cognition outcome and age. However in scenario C we also integrate an interaction term between Age at baseline and Follow-up time. The addition of this interaction term enables the cognition trajectories with Follow-up time to have a different slope, depending on the Age at baseline (where the slope difference per year Follow-up time between two cognition trajectories is then defined as $\beta_{interactionterm}$ *ΔAge at baseline). With a non-linear relationship between the cognition outcome and age, it is plausible that within subjects the speed at which the cognition outcome changes over time, depends on the age, in this example meaning that the cognition outcomes decreases faster at a higher age.

Imagine that there is indeed a non-linear relationship between the cognitive outcome and age, and only a non-linear term for Age at baseline is integrated, as shown in scenario B, while in reality the speed at which the cognition outcome decreases depends on Age at baseline (as shown in scenario C). If in that case the determinant X, for example hearing loss, is strongly and non-linearly related to age, the (residual) association between age and cognition (see scenario D) would be attributed to hearing loss, since especially the oldest have hearing loss, and the residual aging effects on cognition is also the largest in these older participants.

Next to considering age as an confouder, age could also be considered as a proxy for the cummulative damage that has occurred until that specific time. Because of the multifactoral nature of age-related diseases, the determinantion of those causes of disease, is a major challenge. In summary, I believe that taking into account the non-linear relationships with age in longitudinal brain research is crucial, to preclude that potentially causal relations are missed and non-causal relationships are indicated as potentially causal.

Implications and Future research

Brain aging trajectories of structural imaging markers, cognitive function and motor function are essential background information for studies into age-related neurological diseases. Furthermore these trajectories can also play an important role in clinical translation, for example as reference values. Especially in studies looking at differences between age-related pathology and normal aging, it is important to take into account the non-linear relationship with age. Apart from serving as a reference for disease, brain aging in itself is an important topic of research. Brain

aging trajectories of a broad range of markers provide insights in the concurrency of changing markers. Furthermore these trajectories can be used to investigate the effects of risk factors on the individual markers over time. A challenging but crucial next step towards understanding brain aging is to not only show the concurrency of changes, or the effects of risk factors on the individual marker in aging, but to also unravel interactions between different markers and risk factors.

I believe that the field of brain aging can take example of research performed in the setting of multifactorial diseases with a very heterogeneous clinical presentation, such as dementia or multiple sclerosis. Timing of events such as the onset of certain symptoms or presence of disease markers are already being used in these diseases to investigate and identify subtypes of disease and to predict disease progression. Considering that we may never be able to draw a clear line between normal aging and abnormal aging, I believe that with a special focus on the timing and sequence of events in brain aging, we may also be able to identify different patterns of aging in a similar way, which could potentially explain the sources of the temporal and phenotypic heterogeneity of brain aging, its impact on brain functioning, and ultimately improve our understanding of age-related diseases as well.

Concluding remarks/Conclusion

In this thesis I presented brain aging trajectories of structural imaging markers, cognitive function and motor function in a population free of overt neurological disease. I have used these trajectories to investigate brain aging patterns and to investigate the effects different risk factors on brain aging. This thesis highlights the complexity and the importance of understanding brain aging, as well as the importance to take into account the often non-linear relationship of these markers with age.

Based on my findings and reflections, I would like to challenge researchers in the field of aging and brain aging to first of all dare to create a more holistic perspective on aging. Though aging is a very complex process, we must avoid overspecialisation, in which we potentially fail to learn from each other. Secondly, I call for brain researchers to be aware of advances in etiological, prediction and other types of methodological research which could provide us with tools that bring us one step closer towards understanding the aging process.

6

Summary & Samenvatting

English Summary

The aging process is biologically complex and there is a large variation in the effects of aging on the human body between as well as within individuals. Similarly, the aging brain is known for its large intra- and inter-individual variability in functioning and structural changes that occur with advancing age. In a nutshell, the brain shrinks with advancing age as a result of brain atrophy and age-related changes in cognition vary considerable across cognitive domains and across individuals. An important challenge in brain aging research in particular is the overlap in changes in brain structure and function in aging and neurodegenerative disease, such as Alzheimer's disease. Though challenging, this overlap also underlines the importance of understanding brain aging to ultimately improve our understanding of neurodegenerative disease. Therefore, the aim of this thesis was to quantify brain aging trajectories and brain aging patterns. These trajectories were used to investigate the association between risk factors and brain aging. Furthermore I assessed how brain aging can inform disease assessment and prediction.

In **Chapter 2.1** I presented an overview of the concurrency of changing imaging markers with aging. I showed trajectories of volumetric, microstructural and focal imaging markers. In accordance with literature, the trajectories showed often a non-linear curve with accelerating change with advancing age. The brain aging pattern showed that total brain volume changes first, followed by mean diffusivity, white matter, hippocampus, and total grey matter and fractional anisotropy being the last markers to change. Several studies have investigated sex differences in brain structure volume in the aging brain, however they have not yielded consistent results. I showed sex differences in trajectories of all volumetric MRI markers in aging, after correcting for head size differences. In general, earlier acceleration of change in global and lobar volumetric and microstructural markers in men compared with women were found. In **Chapter 2.2** I presented cognitive and motor function trajectories, which generally decline linearly between the ages of 45 and 65 years, followed by a steeper decline after the age of 65–70 years. Test scores for cognitive and motor function declined similarly, with high variation in the rate of decline across age for individual tests. Importantly, whereas a higher level of education was associated with higher cognitive function, there was no association between level of education and function on the majority of the motor tests. No distinct patterns of an overall decline in cognitive function preceding motor function or vice versa was found. However, in accordance with the white matter changes preceding grey matter changes with aging as described in Chapter 2.1, the pattern of changes in the cognitive and motor function test with aging showed that changes in cognitive and motor domains that depend on white matter integrity (i.e. information processing speed, executive function and the gait domain "phases") did occur earlier compared

to the domains that depend on grey matter volume (i.e. memory and gait domain "base of support"). In **Chapter 2.3** I presented the relation between cortical gyrification and age and cognition. Global gyrification of the cerebral cortex decreased linearly with age during adulthood and gyrification in the medial prefrontal cortex increases towards the end of life. Changes in cortical gyrification with age are expected, since changes in brain volumes affects the gyrification. Grey matter atrophy results in a reduction of surface area of the cortex, which leads to more shallow sulci and consequently a decrease in gyrification. Furthermore, gyrification increases with higher levels of cognitive performance in some local regions of the cortex irrespective of surface area. These findings indicate that gyrification may be an imaging marker of interest within the perspective of brain aging and cognitive aging. In **Chapter 2.4** I identified two patterns of structural brain changes with aging, which I refer to as brain aging subtypes, among cognitively normal community-dwelling subjects, using a data-driven approach based on volumetric imaging markers from brain MRI. The cortical brain aging subtype was characterized by early (midtemporal) cortical atrophy, whereas the white matter brain aging subtype was characterized by early ventricle enlargement and total brain atrophy. Participants assigned to the white matter subtype showed both more white matter atrophy and a higher burden of focal cerebral small vessel disease markers than the cortical subtype, as well as lower scores on cognitive function tests. These results are suggestive of the white matter subtype representing a less successful brain aging subtype, while the pattern seen in the white matter subtype shows most resemblance with the average brain aging pattern presented in Chapter 2.1. This underlines the importance of disentangling this phenotypic heterogeneity, as almost 70% of the population actually depicts the cortical subtype, which is not adequately reflected in the previously established average sequence. In **Chapter 2.5** I highlighted the importance and potential of using longitudinal data for unravelling the brain aging process.

In Chapter 3 trajectories of different markers were determined to investigate if and how certain determinants alter these brain aging trajectories. In **Chapter 3.1** I found that a larger burden of intracranial carotid artery calcification and vertebrobasilar artery calcification is associated with faster increase of cerebral small vessel disease markers with increasing age. In **Chapter 3.2** I showed that worse cardiovascular health was associated with a faster increase in white matter hyperintensity volume with advancing age and lower hippocampus volume. Furthermore I found that APOE ε4 genotype was associated with more microbleeds and faster increase in white matter hyperintensities with advancing age. In **Chapter 3.3** I investigated whether hearing loss is associated with accelerates cognitive decline. Hearing loss has been put forward as a promising modifiable risk factor for cognitive decline and

dementia. In contrast to other population-based studies, I showed that hearing loss was not associated with accelerated cognitive decline. The association between hearing loss and accelerated cognitive decline was driven by the non-linear relationship between cognitive function and age. This underlines the importance of understanding brain aging and cognitive aging, when investigating the effects of risk factors.

In Chapter 4 I focussed on determining how brain aging can inform disease assessment and prediction. In **Chapter 4.1** I estimated percentile curves of subcortical structures based on three different reference populations, using a single segmentation tool. I found that percentile curves of subcortical structures, based on the different reference populations largely overlap. This indicates that there are only small differences between the subcortical volumes of these populations, regardless of differences in acquisition, field strength, vendors, and population differences. Furthermore, this indicates that in a clinical setting, individual diagnostic assessment based on subcortical volume, may not be biased by the use of a specific reference population, as long as the same imaging processing tool is used. Using a data-driven disease progression model in **Chapter 4.2**, I constructed APOE-specific Alzheimer's disease related disease timelines based on structural imaging markers in a case-control setting. I showed that the data-driven disease timelines are generalizable to population-based cohorts and that progression of individuals along such timelines is predictive of incident Alzheimer's disease.

In **Chapter 5** I summarized the main findings of this thesis, I discussed methodological considerations, implications and future directions. This thesis highlights the complexity and the importance of understanding brain aging, as well as the importance to take into account the often non-linear relationship of these markers with age.

Nederlandse Samenvatting

Veroudering van het menselijk lichaam is een zeer complex biologisch proces met een grote variatie in de mate van vatbaarheid van het menselijk lichaam voor dit verouderingsproces, zowel tussen als binnen individuen. Kijkend naar het brein is er een grote variabiliteit in veranderingen die plaatsvinden in brein structuur met het ouder worden. Brein regio's verschillen in de gevoeligheid voor verandering door veroudering, maar over het algemeen krimpt het brein naarmate we ouder worden, genaamd brein atrofie. Tegelijkertijd varieert de mate waarin ons cognitief functioneren verandert met veroudering per domein en tussen individuen. Een belangrijke uitdaging in het onderzoek naar brein veroudering is de overlap tussen veranderingen in brein structuur en functie in veroudering én in neurodegeneratieve ziekten, zoals de ziekte van Alzheimer. Deze overlap benadrukt, dat met het beter begrijpen van brein veroudering, we uiteindelijk ook neurodegeneratieve ziekten beter zullen begrijpen. De doelstelling van dit proefschrift was daarom ook het kwantificeren van brein veroudering curves en patronen van brein veroudering. Deze curves heb ik vervolgens gebruikt om de associatie tussen risicofactoren en brein veroudering te onderzoeken. Vervolgens heb ik onderzocht hoe brein veroudering ons kan helpen in de beoordeling en predictie van ziekte.

In **Hoofdstuk 2.1** presenteer ik curves van brein volumes, microstructuur en focale laesies en beschrijf ik de samenloop van deze brein veranderingen met het ouder worden. In overeenstemming met de literatuur zijn deze curves vaak non-lineair, met snellere brein veranderingen naarmate leeftijd toeneemt. Het patroon in brein veroudering laat zien dat totale brein volume het eerst verandert, gevolgd door de gemiddelde diffusie, witte stof, hippocampus, met als laatste de grijze stof en fractionele anisotropie. Verscheidene studies hebben onderzoek gedaan naar verschillen in brein structuur volume tussen mannen en vrouwen, maar deze hebben geen consistente resultaten laten zien. Ik laat zien dat de brein veroudering curves verschillen tussen mannen en vrouwen, na correctie voor verschillen in hoofd grootte. Over het algemeen is er een vroegere toename in snelheid van de veranderingen in globale en lobaire brein volume en microstructuur in mannen, in vergelijking met vrouwen. In **Hoofdstuk 2.2** presenteer ik veroudering curves van cognitieve functies en motor functies, waarin in het algemeen een lineaire afname tussen de leeftijden 45 en 65 is te zien, wat gevolgd wordt door een sterkere afname na de leeftijd van 65-70 jaar. Test scores van cognitieve functies en motor functies nemen op een vergelijkbare manier af met leeftijd, met een grote variatie in de mate van afname over leeftijd voor de individuele testen. Belangrijk daarbij is dat waar een hoger educatie niveau geassocieerd was met een hoger cognitieve functie, er geen associatie was tussen educatie en de meerderheid van de motor functie testen. Er was geen duidelijk patroon waarin cognitieve functie voorafgaand aan motor

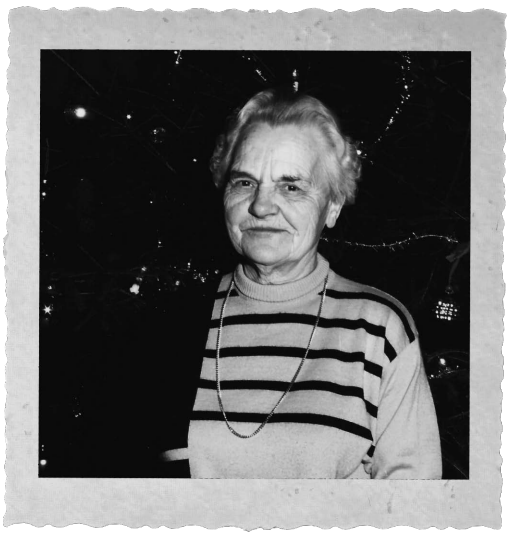
functie afneemt of vice versa. Echter, in overeenstemming met de witte stof afname voorafgaand aan grijze stof afname met veroudering zoals beschreven in Hoofdstuk 2.1, was te zien dat cognitieve en motor functie testen die afhankelijk zijn van de witte stof integriteit (zoals de snelheid van informatieverwerking, uitvoerend functioneren en het looppatroon domein "fases") eerder veranderden dan testen die afhankelijk zijn van grijze stof volume (zoals geheugen en looppatroon domein "basis van ondersteuning"). In **Hoofdstuk 2.3** presenteer ik de relatie tussen corticale gyrificatie en leeftijd en cognitie. Globale gyrificatie van de cerebrale cortex neemt bij volwassenen lineair af met leeftijd en gyrificatie in de mediale prefrontale cortex neemt toe op hoge leeftijd. Veranderingen in corticale gyrificatie met leeftijd werden verwacht, aangezien veranderingen in brein volume invloed heeft op gyrificatie. Grijze stof atrofie resulteert in een afname van de oppervlakte van de cortex, wat leid tot ondiepere sulci, met als gevolg een afname van gyrificatie. Gyrificatie neemt toe met een hoger niveau van cognitieve prestaties in sommige regio's van de cortex, onafhankelijk van de oppervlakte van deze regio's. Deze bevindingen tonen aan dat gyrificatie mogelijk een interessante marker is in het domein van brein veroudering en cognitieve veroudering. In **Hoofdstuk 2.4** heb ik twee patronen van structurele brein veroudering geïdentificeerd, oftewel subtypes van brein veroudering, onder cognitief normale thuiswonende personen, gebruik makend van een data gedreven benadering op basis van brein volumetrie. Het corticale subtype kenmerkt zich door vroege (midtemporale) corticale atrofie, terwijl het witte stof subtype zich kenmerkt door vroege toename van het ventrikel volume en totale brein atrofie. Personen met een verouderingspatroon passend bij het witte-stof subtype hadden meer witte stof atrofie en meer focale laesies dan het corticale subtype, en daarnaast slechtere scores op cognitieve testen. Deze resultaten wijzen erop dat het witte stof subtype mogelijk een minder succesvol brein veroudering subtype is, dit terwijl het patroon van het corticale subtype de meeste gelijkenis vertoond met het brein verouderingspatroon in Hoofdstuk 2.1. Dit benadrukt het belang van het ontrafelen van de fenotypische heterogeniteit, aangezien bijna 70% van de populatie een brein verouderingspatroon heeft wat meer lijkt op het corticale subtype, wat niet overeenkomt met het gemiddelde patroon beschreven in Hoofdstuk 2.1. In **Hoofdstuk 2.5** benadruk ik het belang en de potentie van het gebruiken van longitudinale data voor het ontrafelen van het brein verouderingsproces.

In Hoofdstuk 3 zijn de curves van verschillende markers bepaald om te onderzoeken hoe bepaalde determinanten deze brein veroudering curves beïnvloeden. In **Hoofdstuk 3.1** heb ik geconstateerd dat meer intracraniële carotis calcificatie en vertebrobasilaris calcificatie geassocieerd is met een snellere toename van focale laesies met veroudering. In **Hoofdstuk 3.2** heb ik laten zien dat een slechtere cardiovasculaire gezondheid geassocieerd is met een snellere toename in witte stof

hyperintensiteiten met veroudering en een lager hippocampus volume. Daarbij zag ik dat het APOE ε4 genotype geassocieerd is met meer microbloedingen en een snellere toename van witte stof hyperintensiteiten met veroudering. In **Hoofdstuk 3.3** heb ik onderzocht of gehoorverlies geassocieerd is met een versnelde cognitieve achteruitgang. Gehoorverlies is naar voren geschoven as een veelbelovende aanpasbare risicofactor voor cognitieve achteruitgang en dementie. In tegenstelling tot andere populatie studies, heb ik laten zien dat gehoorverlies niet geassocieerd is met versnelde cognitieve achteruitgang. De associatie tussen gehoorverlies en versnelde achteruitgang werd gedreven door de non-lineaire relatie tussen cognitie en leeftijd. Dit benadrukt het belang van het begrijpen van brein veroudering en cognitieve veroudering in onderzoek naar de effecten van risico factoren.

In Hoofdstuk 4 focus ik op het bepalen hoe brein veroudering ons kan bijdragen in de beoordeling en predictie van ziekte. In **Hoofdstuk 4.1** heb ik percentiel curves bepaald van subcorticale structuren, gebaseerd op drie verschillende referentie populaties. Ik concludeerde dat percentiel curves van subcorticale structuren, gebaseerd op de verschillende referentie populaties, grotendeels overlappen. Dit betekent dat er mogelijk slechts kleine verschillen zijn in subcorticale volumes, ondanks de verschillen in populatie, acquisitie, veld sterkte, en merk van de MRI scanners. Daarbij impliceert dit dat in een klinische setting, waarbij individuele diagnose gebaseerd wordt op subcorticale volume, er mogelijk geen bias ontstaat op basis van de referentie populatie, zolang dezelfde imaging processing is toegepast. Gebruik makend van een data gedreven ziekte progressie model in **Hoofdstuk 4.2**, heb ik het effect van het APOE genotype op het ziekteprogressie patroon van de ziekte van Alzheimer bepaald aan de hand van structurele brein volumes in een case-control setting. De resulterende ziekteprogressietijdslijnen waren generaliseerbaar naar een populatie cohort en de progressie van een individu over deze tijdslijn is voorspellend voor het krijgen van de ziekte van Alzheimer.

In **Hoofdstuk 5** heb ik de belangrijkste bevindingen van dit proefschrift samengevat, methodologische overwegingen, implicaties en aanbevelingen voor toekomstig onderzoek bediscussieerd. Dit proefschrift benadrukt de complexiteit en het belang van het begrijpen van brein veroudering, en het belang van het rekening houden met een vaak niet-lineaire relatie tussen determinanten en uitkomsten met leeftijd.



Bibliography

1. Rowe, J. W. & Kahn, R. L. Human Aging: Usual and Successful. *Science (80-.).* **237**, 143–149 (1987).

2. Depp, C. A. & Jeste, D. V. Definitions and Predictors of Successful Aging: A Comprehensive Review of Larger Quantitative Studies. *Am. J. Geriatr. Psychiatry* **14**, 6–20 (2006).

3. Glisky, E. L. Changes in Cognitive Function in Human Aging. in *Brain Aging: Models, Methods, and Mechanisms* (ed. Riddle, D. R.) 18 (CRC Press/Taylor & Francis, 2007).

4. Johnson, K. A., Fox, N. C., Sperling, R. A. & Klunk, W. E. Brain imaging in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **2**, a006213–a006213 (2012).

5. Jack, C. R. *et al.* Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* **7**, 257–262 (2011).

6. Fjell, A. M. *et al.* What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Prog. Neurobiol.* **117**, 20–40 (2014).

7. Gur, R. C. *et al.* Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. *Proc. Natl. Acad. Sci.* **88**, 2845–2849 (1991).

8. Jernigan, T. L., Press, G. A. & Hesselink, J. R. Methods for Measuring Brain Morphologic Features on Magnetic Resonance Images: Validation and Normal Aging. *Arch. Neurol.* **47**, 27–32 (1990).

9. Jernigan, T. L. *et al.* Cerebral structure on MRI, Part I: Localization of age-related changes. *Biol. Psychiatry* **29**, 55–67 (1991).

10. Ranga Krishnan, K. *et al.* In vivo stereological assessment of caudate volume in man: Effect of normal aging. *Life Sci.* **47**, 1325–1329 (1990).

11. Pfefferbaum, A. *et al.* A Quantitative Magnetic Resonance Imaging Study of Changes in Brain Morphology From Infancy to Late Adulthood. *Arch. Neurol.* **51**, 874–887 (1994).

12. Sullivan, E. V, Marsh, L., Mathalon, D. H., Lim, K. O. & Pfefferbaum, A. Age-related decline in MRI volumes of temporal lobe gray matter but not hippocampus. *Neurobiol. Aging* **16**, 591–606 (1995).

13. Coffey, C. E. *et al.* Sex Differences in Brain Aging: A Quantitative Magnetic Resonance Imaging Study. *Arch. Neurol.* **55**, 169–179 (1998).

14. Good, C. D. *et al.* A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *Neuroimage* **14**, 21–36 (2001).

15. Mu, Q., Xie, J., Wen, Z., Weng, Y. & Shuyun, Z. A Quantitative MR Study of the Hippocampal Formation, the Amygdala, and the Temporal Horn of the Lateral Ventricle in Healthy Subjects 40 to 90 Years of Age. *Am. J. Neuroradiol.* **20**, 207 LP – 211 (1999).

16. Coupé, P., Catheline, G., Lanuza, E., Manjón, J. V. & Initiative, for the A. D. N.

Towards a unified analysis of brain maturation and aging across the entire lifespan: A MRI analysis. *Hum. Brain Mapp.* **38**, 5501–5518 (2017).

17. Cox, S. R. *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat. Commun.* **7**, 13629 (2016).

18. de Groot, M. *et al.* Tract-specific white matter degeneration in aging: The Rotterdam Study. *Alzheimer's Dement.* **11**, 321–330 (2015).

19. Lebel, C. *et al.* Diffusion tensor imaging of white matter tract evolution over the lifespan. *Neuroimage* **60**, 340–352 (2012).

20. Sullivan, E. V *et al.* Equivalent disruption of regional white matter microstructure in ageing healthy men and women. *Neuroreport* **12**, (2001).

21. van Velsen, E. F. S. *et al.* Brain cortical thickness in the general elderly population: The Rotterdam Scan Study. *Neurosci. Lett.* **550**, 189–194 (2013).

22. Vermeer, S. E., Koudstaal, P. J., Oudkerk, M., Hofman, A. & Breteler, M. M. B. Prevalence and Risk Factors of Silent Brain Infarcts in the Population-Based Rotterdam Scan Study. *Stroke* **33**, 21–25 (2002).

23. Barrick, T. R., Charlton, R. A., Clark, C. A. & Markus, H. S. White matter structural decline in normal ageing: A prospective longitudinal study using tract-based spatial statistics. *Neuroimage* **51**, 565–577 (2010).

24. Driscoll, I. *et al.* Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology* **72**, 1906–1913 (2009).

25. de Groot, M. *et al.* White Matter Degeneration with Aging: Longitudinal Diffusion MR Imaging Analysis. *Radiology* **279**, 532–541 (2015).

26. Du, A.-T. *et al.* Age effects on atrophy rates of entorhinal cortex and hippocampus. *Neurobiol. Aging* **27**, 733–740 (2006).

27. Fjell, A. M. *et al.* Critical ages in the life course of the adult brain: nonlinear subcortical aging. *Neurobiol. Aging* **34**, 2239–2247 (2013).

28. M., F. A. *et al.* Development and aging of cortical thickness correspond to genetic organization patterns. *Proc. Natl. Acad. Sci.* **112**, 15462–15467 (2015).

29. Fraser, M. A., Shaw, M. E. & Cherbuin, N. A systematic review and meta-analysis of longitudinal hippocampal atrophy in healthy human ageing. *Neuroimage* **112**, 364–374 (2015).

30. Narvacan, K., Treit, S., Camicioli, R., Martin, W. & Beaulieu, C. Evolution of deep gray matter volume across the human lifespan. *Hum. Brain Mapp.* **38**, 3771–3790 (2017).

31. Pfefferbaum, A. *et al.* Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85years) measured with atlas-based parcellation of MRI. *Neuroimage* **65**, 176–193 (2013).

32. Raz, N. *et al.* Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cereb. Cortex* **15**, 1676–1689 (2005).

33. Sexton, C. E. *et al.* Accelerated Changes in White Matter Microstructure during Aging: A Longitudinal Diffusion Tensor Imaging Study. *J. Neurosci.* **34**, 15425 LP – 15436 (2014).

34. Storsve, A. B. *et al.* Differential Longitudinal Changes in Cortical Thickness, Surface Area and Volume across the Adult Life Span: Regions of Accelerating and Decelerating Change. *J. Neurosci.* **34**, 8488–8498 (2014).

35. Sullivan, E. V, Rohlfing, T. & Pfefferbaum, A. Longitudinal Study of Callosal Microstructure in the Normal Adult Aging Brain Using Quantitative DTI Fiber Tracking. *Dev. Neuropsychol.* **35**, 233–256 (2010).

36. Ikram, M. A. *et al.* The Rotterdam Study: 2018 update on objectives, design and main results. *Eur. J. Epidemiol.* **32**, 807–850 (2017).

37. Ikram, M. A. *et al.* The Rotterdam Scan Study: design update 2016 and main findings. *Eur. J. Epidemiol.* **30**, (2015).

38. de Boer, R. *et al.* White matter lesion extension to automatic brain tissue segmentation on MRI. *Neuroimage* **45**, 1151–1161 (2009).

39. Vrooman, H. A. *et al.* Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification. *Neuroimage* **37**, 71–81 (2007).

40. Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **17**, 87–97 (1998).

41. Bokde, A. L. W. *et al.* Reliable manual segmentation of the frontal, parietal, temporal, and occipital lobes on magnetic resonance images of healthy subjects. *Brain Res. Protoc.* **14**, 135–145 (2005).

42. Ikram, M. A. *et al.* Kidney Function Is Related to Cerebral Small Vessel Disease. *Stroke* **39**, 55–61 (2008).

43. Fischl, B. *et al.* Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**, S69–S84 (2004).

44. Vernooij, M. W. *et al.* Incidental Findings on Brain MRI in the General Population. *N. Engl. J. Med.* **357**, 1821–1828 (2007).

45. Adams, H. H. H. *et al.* Rating Method for Dilated Virchow–Robin Spaces on Magnetic Resonance Imaging. *Stroke* **44**, 1732–1735 (2013).

46. Pinheiro, J. & Bates, D. NLME: Linear and nonlinear mixed effects models. *R Packag. version 3.1-158* (2006).

47. Højsgaard, S., Halekoh, U. & Yan, J. The R Package geepack for Generalized Estimating Equations. *J. Stat. Softw.* **15**, 1–11 (2005).

48. Poels, M. M. F. *et al.* Incidence of Cerebral Microbleeds in the General Population. *Stroke* **42**, 656–661 (2011).

49. DeCarli, C. *et al.* Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. *Neurobiol. Aging* **26**, 491–510 (2005).

50. Lockhart, S. N. & DeCarli, C. Structural Imaging Measures of Brain Aging. *Neuropsychol. Rev.* **24**, 271–289 (2014).

51. Fotenos, A. F., Snyder, A. Z., Girton, L. E., Morris, J. C. & Buckner, R. L. Normative estimates of cross-sectional and longitudinal brain volume decline in aging and AD. *Neurology* **64**, 1032 LP – 1039 (2005).

52. Ge, Y. *et al.* Age-Related Total Gray Matter and White Matter Changes in Normal Adult Brain. Part I: Volumetric MR Imaging Analysis. *Am. J. Neuroradiol.* **23**, 1327 LP – 1333 (2002).

53. Fjell Anders M. & Walhovd Kristine B. Structural Brain Changes in Aging: Courses, Causes and Cognitive Consequences. *Rev. Neurosci.* **21**, 187–222 (2010).

54. Pfefferbaum, A. & Sullivan, E. V. Cross-sectional versus longitudinal estimates of age-related changes in the adult brain: overlaps and discrepancies. *Neurobiol. Aging* **36**, 2563–2567 (2015).

55. SULLIVAN, E. V & PFEFFERBAUM, A. Neuroradiological characterization of normal adult ageing. *Br. J. Radiol.* **80**, S99–S108 (2007).

56. Raz, N. *et al.* Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. *Cereb. Cortex* **7**, 268–282 (1997).

57. Ikram, M. A. *et al.* Brain tissue volumes in the general elderly population: The Rotterdam Scan Study. *Neurobiol. Aging* **29**, 882–890 (2008).

58. Leonard, C. M. *et al.* Size Matters: Cerebral Volume Influences Sex Differences in Neuroanatomy. *Cereb. Cortex* **18**, 2920–2931 (2008).

59. Beaulieu, C. The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR Biomed.* **15**, 435–455 (2002).

60. de Groot, M. *et al.* Changes in Normal-Appearing White Matter Precede Development of White Matter Lesions. *Stroke* **44**, 1037–1042 (2013).

61. Poels, M. M. F. *et al.* Prevalence and Risk Factors of Cerebral Microbleeds. *Stroke* **41**, S103–S106 (2010).

62. Vermeer, S. E., Longstreth, W. T. & Koudstaal, P. J. Silent brain infarcts: a systematic review. *Lancet Neurol.* **6**, 611–619 (2007).

63. Goldstein, L. B. *et al.* Primary Prevention of Ischemic Stroke . *Stroke* **32**, 280–299 (2001).

64. Darweesh, S. K. L. *et al.* Association Between Poor Cognitive Functioning and Risk of Incident Parkinsonism: The Rotterdam Study. *JAMA Neurol.* **74**, 1431–1438 (2017).

65. Verruca, L. *et al.* Cognitive Impairment and Risk of Stroke in the Older Population. *J. Am. Geriatr. Soc.* **44**, 237–241 (1996).

66. Darweesh, S. K. L. *et al.* Quantitative gait, cognitive decline, and incident dementia: The Rotterdam Study. *Alzheimer's Dement.* **15**, 1264–1273 (2019).

67. Darweesh, S. K. L. *et al.* Trajectories of prediagnostic functioning in Parkinson's disease. *Brain* **140**, 429–441 (2017).

68. Licher, S. *et al.* Development and Validation of a Dementia Risk Prediction Model in the General Population: An Analysis of Three Longitudinal Studies. *Am. J. Psychiatry* **176**, 543–551 (2018).

69. Hoogendam, Y. Y., Hofman, A., van der Geest, J. N., van der Lugt, A. & Ikram, M. A. Patterns of cognitive function in aging: the Rotterdam Study. *Eur. J. Epidemiol.* **29**, 133–140 (2014).

70. Hoogendam, Y. Y. *et al.* Older age relates to worsening of fine motor skills: a population-based study of middle-aged and elderly persons. *Front. Aging Neurosci.* **6**, 259 (2014).

71. Singh-Manoux, A. *et al.* Timing of onset of cognitive decline: results from Whitehall II prospective cohort study. *BMJ* **344**, d7622 (2012).

72. Cooper, R. *et al.* Age and Gender Differences in Physical Capability Levels from Mid-Life Onwards: The Harmonisation and Meta-Analysis of Data from Eight UK Cohort Studies. *PLoS One* **6**, e27899 (2011).

73. Downer, B., Chen, N.-W., Raji, M. & Markides, K. S. A longitudinal study of cognitive trajectories in Mexican Americans age 75 and older. *Int. J. Geriatr. Psychiatry* **32**, 1122–1130 (2017).

74. Mielke, M. M. *et al.* Assessing the Temporal Relationship Between Cognition and Gait: Slow Gait Predicts Cognitive Decline in the Mayo Clinic Study of Aging. *Journals Gerontol. Ser. A* **68**, 929–937 (2013).

75. Watson, N. L. *et al.* Executive Function, Memory, and Gait Speed Decline in Well-Functioning Older Adults. *Journals Gerontol. Ser. A* **65A**, 1093–1100 (2010).

76. Gale, C. R., Allerhand, M., Sayer, A. A., Cooper, C. & Deary, I. J. The dynamic relationship between cognitive function and walking speed: the English Longitudinal Study of Ageing. *Age (Omaha).* **36**, 9682 (2014).

77. Krall, J. R., Carlson, M. C., Fried, L. P. & Xue, Q.-L. Examining the Dynamic, Bidirectional Associations Between Cognitive and Physical Functioning in Older Adults. *Am. J. Epidemiol.* **180**, 838–846 (2014).

78. Atkinson, H. H. *et al.* The Relationship Between Cognitive Function and Physical Performance in Older Women: Results From the Women's Health Initiative Memory Study. *Journals Gerontol. Ser. A* **65A**, 300–306 (2010).

79. Clouston, S. A. P. *et al.* The Dynamic Relationship Between Physical Function and Cognition in Longitudinal Aging Cohorts. *Epidemiol. Rev.* **35**, 33–50 (2013).

80. Montero-Odasso, M., Verghese, J., Beauchet, O. & Hausdorff, J. M. Gait and Cognition: A Complementary Approach to Understanding Brain Function and the Risk of Falling. *J. Am. Geriatr. Soc.* **60**, 2127–2136 (2012).

81. Elbaz, A. *et al.* Motor function in the elderly. *Neurology* **81**, 417 LP – 426 (2013).

82. Atkinson, H. H. *et al.* Cognitive Function, Gait Speed Decline, and Comorbidities:

The Health, Aging and Body Composition Study. *Journals Gerontol. Ser. A* **62**, 844–850 (2007).

83. Vinke, E. J. *et al.* Trajectories of imaging markers in brain aging: the Rotterdam Study. *Neurobiol. Aging* **71**, 32–40 (2018).

84. Licher, S. *et al.* Lifetime risk of common neurological diseases in the elderly population. *J. Neurol. Neurosurg. & Psychiatry* **90**, 148 (2019).

85. Welsh, K. A. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part V. A normative study of the neuropsychological battery. *Neurology* **44**, 609 LP – 609 (1994).

86. Smith, A. & JHelmuth, J. The symbol-digit modalities test: a neuropsychologic test of learning and other cerebral disorders. Learning disorders. *Spec. child Publ. Seattle* 91 (1968).

87. Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology* vol. 18 643–662 (1935).

88. Brand, N. & Jolles, J. Learning and Retrieval Rate of Words Presented Auditorily and Visually. *J. Gen. Psychol.* **112**, 201–210 (1985).

89. Killgore, W. D. S., Glahn, D. C. & Casasanto, D. J. Development and Validation of the Design Organization Test (DOT): A Rapid Screening Instrument for Assessing Visuospatial Ability. *J. Clin. Exp. Neuropsychol.* **27**, 449–459 (2005).

90. Van der Elst, W., van Boxtel, M. P. J., van Breukelen, G. J. P. & Jolles, J. The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24–81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. *J. Clin. Exp. Neuropsychol.* **28**, 998–1009 (2006).

91. Franzen, M. D., Tishelman, A. C., Sharp, B. H. & Friedman, A. G. An investigation of the test-retest reliability of the stroop colorword test across two intervals. *Arch. Clin. Neuropsychol.* **2**, 265–272 (1987).

92. Lezak, M. D. Neuropsychological assessment in behavioral toxicology — Developing techniques and interpretative issues. *Scand. J. Work. Environ. Health* **10**, 25–29 (1984).

93. Van Der Elst, W. I. M., Van Boxtel, M. P. J., Van Breukelen, G. J. P. & Jolles, J. Rey's verbal learning test: normative data for 1855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of presentation. *J. Int. Neuropsychol. Soc.* **11**, 290–302 (2005).

94. MacKINNON, D. P., Geiselman, R. E. & Woodward, J. A. The effects of effort on Stroop interference. *Acta Psychol. (Amst).* **58**, 225–235 (1985).

95. Desrosiers, J., Hébert, R., Bravo, G. & Dutil, E. The Purdue Pegboard Test: Normative data for people aged 60 and over. *Disabil. Rehabil.* **17**, 217–224 (1995).

96. Verlinden, V. J. A., van der Geest, J. N., Hofman, A. & Ikram, M. A. Cognition and gait show a distinct pattern of association in the general population. *Alzheimer's Dement.* **10**, 328–335 (2014).

97. Verlinden, V. J. A. *et al.* Gait patterns in a community-dwelling population aged 50 years and older. *Gait Posture* **37**, 500–505 (2013).

98. Wingbermühle, R., Wen, K., Wolters, F. J., Ikram, M. A. & Bos, D. Smoking, APOE genotype, and cognitive decline: the Rotterdam study. *J. Alzheimer's Dis.* **57**, 1191–1195 (2017).

99. Slooter, A. J. C. *et al.* Risk estimates of dementia by apolipoprotein E genotypes from a population-based incidence study: the Rotterdam Study. *Arch. Neurol.* **55**, 964–968 (1998).

100. Van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).

101. Camicioli, R., Howieson, D., Oken, B., Sexton, G. & Kaye, J. Motor slowing precedes cognitive impairment in the oldest old. *Neurology* **50**, 1496–1498 (1998).

102. Deshpande, N., Metter, E. J., Bandinelli, S., Guralnik, J. & Ferrucci, L. Gait speed under varied challenges and cognitive decline in older persons: a prospective study. *Age Ageing* **38**, 509–514 (2009).

103. Buracchio, T., Dodge, H. H., Howieson, D., Wasserman, D. & Kaye, J. The trajectory of gait speed preceding mild cognitive impairment. *Arch. Neurol.* **67**, 980–986 (2010).

104. Morris, R. *et al.* Gait rather than cognition predicts decline in specific cognitive domains in early Parkinson's disease. *Journals Gerontol. Ser. A Biomed. Sci. Med. Sci.* **72**, 1656–1662 (2017).

105. Callisaya, M. L. *et al.* Longitudinal relationships between cognitive decline and gait slowing: the Tasmanian Study of Cognition and Gait. *Journals Gerontol. Ser. A Biomed. Sci. Med. Sci.* **70**, 1226–1232 (2015).

106. Tabbarah, M., Crimmins, E. M. & Seeman, T. E. The relationship between cognitive and physical performance: MacArthur Studies of Successful Aging. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **57**, M228–M235 (2002).

107. Soumaré, A., Tavernier, B., Alpérovitch, A., Tzourio, C. & Elbaz, A. A cross-sectional and longitudinal study of the relationship between walking speed and cognitive function in community-dwelling elderly people. *Journals Gerontol. Ser. A Biomed. Sci. Med. Sci.* **64**, 1058–1065 (2009).

108. Inzitari, M. *et al.* Gait speed predicts decline in attention and psychomotor speed in older adults: the health aging and body composition study. *Neuroepidemiology* **29**, 156–162 (2007).

109. Verghese, J., Wang, C., Lipton, R. B., Holtzer, R. & Xue, X. Quantitative gait dysfunction and risk of cognitive decline and dementia. *J. Neurol. Neurosurg. Psychiatry* **78**, 929–935 (2007).

110. Cremers, L. G. M. *et al.* Altered tract-specific white matter microstructure is related to poorer cognitive performance: The Rotterdam Study. *Neurobiol. Aging* **39**, 108–117 (2016).

111. Ikram, M. A. *et al.* Brain tissue volumes in relation to cognitive function and risk of dementia. *Neurobiol. Aging* **31**, 378–386 (2010).

112. Rosano, C. *et al.* Special Article: Gait Measures Indicate Underlying Focal Gray Matter Atrophy in the Brain of Older Adults. *Journals Gerontol. Ser. A* **63**, 1380–1388 (2008).

113. Verlinden, V. J. A. *et al.* Tract-specific white matter microstructure and gait in humans. *Neurobiol. Aging* **43**, 164–173 (2016).

114. Alattar, A. A. *et al.* Hearing impairment and cognitive decline in older, community-dwelling adults. *Journals Gerontol. Ser. A* **75**, 567–573 (2020).

115. Schubert, C. R. *et al.* Sensorineural impairments, cardiovascular risk factors, and 10-year incidence of cognitive impairment and decline in midlife: The Beaver Dam Offspring Study. *Journals Gerontol. Ser. A* **74**, 1786–1792 (2019).

116. van der Willik, K. D., Schagen, S. B. & Ikram, M. A. Cancer and dementia: Two sides of the same coin? *Eur. J. Clin. Invest.* **48**, e13019 (2018).

117. Federman, A. D. *et al.* Diminished cognitive function among chronic obstructive pulmonary disease patients during periods of acute illness exacerbation. *Journals Gerontol. Ser. A Biomed. Sci. Med. Sci.* **71**, 279–280 (2016).

118. Scherer, S. A., Bainbridge, J. S., Hiatt, W. R. & Regensteiner, J. G. Gait characteristics of patients with claudication. *Arch. Phys. Med. Rehabil.* **79**, 529–531 (1998).

119. Stern, Y. *et al.* Influence of education and occupation on the incidence of Alzheimer's disease. *Jama* **271**, 1004–1010 (1994).

120. Valenzuela, M. J. & Sachdev, P. Brain reserve and cognitive decline: a non-parametric systematic review. *Psychol. Med.* **36**, 1065–1073 (2006).

121. Livingston, G. *et al.* Dementia prevention, intervention, and care. *Lancet* **390**, 2673–2734 (2017).

122. Euser, S. M., Schram, M. T., Hofman, A., Westendorp, R. G. J. & Breteler, M. M. B. Measuring cognitive function with age: the influence of selection by health and survival. *Epidemiology* 440–447 (2008).

123. Docherty, A. R. *et al.* Does degree of gyrification underlie the phenotypic and genetic associations between cortical surface area and cognitive ability? *Neuroimage* **106**, 154–160 (2015).

124. Zilles, K., Palomero-Gallagher, N. & Amunts, K. Development of cortical folding during evolution and ontogeny. *Trends Neurosci.* **36**, 275–284 (2013).

125. Duret, P. *et al.* Gyrification changes are related to cognitive strengths in autism. *NeuroImage Clin.* **20**, 415–423 (2018).

126. Blanken, L. M. E. *et al.* Cortical morphology in 6-to 10-year old children with autistic traits: a population-based neuroimaging study. *Am. J. Psychiatry* **172**, 479–486 (2015).

127. Matsuda, Y. & Ohi, K. Cortical gyrification in schizophrenia: current perspectives. *Neuropsychiatr. Dis. Treat.* **14**, 1861 (2018).

128. Cao, B. *et al.* Lifespan gyrification trajectories of human brain in healthy individuals and patients with major psychiatric disorders. *Sci. Rep.* **7**, 1–8 (2017).

129. White, T., Su, S., Schmidt, M., Kao, C.-Y. & Sapiro, G. The development of gyrification in childhood and adolescence. *Brain Cogn.* **72**, 36–45 (2010).

130. Gregory, M. D. *et al.* Regional variations in brain gyrification are associated with general cognitive ability in humans. *Curr. Biol.* **26**, 1301–1305 (2016).

131. Hogstrom, L. J., Westlye, L. T., Walhovd, K. B. & Fjell, A. M. The structure of the cerebral cortex across adult life: age-related patterns of surface area, thickness, and gyrification. *Cereb. cortex* **23**, 2521–2530 (2013).

132. Schaer, M. *et al.* A surface-based approach to quantify local cortical gyrification. *IEEE Trans. Med. Imaging* **27**, 161–170 (2008).

133. Battaglini, M. *et al.* Lifespan normative data on rates of brain volume changes. *Neurobiol. Aging* **81**, 30–37 (2019).

134. Bleecker, M. L., Bolla-Wilson, K., Agnew, J. & Meyers, D. A. Age-related sex differences in verbal memory. *J. Clin. Psychol.* **44**, 403–411 (1988).

135. Houx, P. J., Jolles, J. & Vreeling, F. W. Stroop interference: aging effects assessed with the Stroop Color-Word Test. *Exp. Aging Res.* **19**, 209–224 (1993).

136. Tiffin, J. & Asher, E. J. The Purdue Pegboard: norms and studies of reliability and validity. *J. Appl. Psychol.* **32**, 234 (1948).

137. Deary, I. J. The stability of intelligence from childhood to old age. *Curr. Dir. Psychol. Sci.* **23**, 239–245 (2014).

138. Deary, I. J. Intelligence. *Annual Review of Psychology* vol. 63 453–482 (2012).

139. Marcus, D. S., Olsen, T. R., Ramaratnam, M. & Buckner, R. L. The extensible neuroimaging archive toolkit. *Neuroinformatics* **5**, 11–33 (2007).

140. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).

141. Schaer, M. *et al.* How to measure cortical folding from MR images: a step-by-step tutorial to compute local gyrification index. *JoVE (Journal Vis. Exp.* e3417 (2012).

142. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).

143. White, T. *et al.* Automated quality assessment of structural magnetic resonance images in children: Comparison with visual inspection and surface-based reconstruction. *Hum. Brain Mapp.* **39**, 1218–1231 (2018).

144. Vliegenthart, R. *et al.* Alcohol consumption and risk of peripheral arterial disease: the Rotterdam study. *Am. J. Epidemiol.* **155**, 332–338 (2002).

145. R Core Team, Rf. R: A language and environment for statistical computing. (2018).

146. Hagler Jr, D. J., Saygin, A. P. & Sereno, M. I. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *Neuroimage* **33**, 1093–1103 (2006).

147. Greve, D. N. & Fischl, B. False positive rates in surface-based anatomical analysis. *Neuroimage* **171**, 6–14 (2018).

148. Taylor, P., Hobbs, J. N., Burroni, J. & Siegelmann, H. T. The global landscape of cognition: hierarchical aggregation as an organizational principle of human cortical networks and functions. *Sci. Rep.* **5**, 1–18 (2015).

149. Cox, S. R. *et al.* Brain cortical characteristics of lifetime cognitive ageing. *Brain Struct. Funct.* **223**, 509–518 (2018).

150. Rubin, D. B. Multiple imputation for nonresponse in surveys. *Stat. Pap.* **31**, 180–180 (1990).

151. Greenland, S. *et al.* Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur. J. Epidemiol.* **31**, 337–350 (2016).

152. Yang, H. *et al.* Study of brain morphology change in Alzheimer's disease and amnestic mild cognitive impairment compared with normal controls. *Gen. psychiatry* **32**, (2019).

153. Ossenkoppele, R. *et al.* Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease. *Neurology* **92**, e601–e612 (2019).

154. Dickerson, B. C. *et al.* Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiol. Aging* **30**, 432–440 (2009).

155. Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816–847 (2012).

156. Jeneson, A. & Squire, L. R. Working memory, long-term memory, and medial temporal lobe function. *Learn. Mem.* **19**, 15–25 (2012).

157. Kohli, J. S., Kinnear, M. K., Martindale, I. A., Carper, R. A. & Müller, R.-A. Regionally decreased gyrification in middle-aged adults with autism spectrum disorders. *Neurology* **93**, e1900 LP-e1905 (2019).

158. Palaniyappan, L. & Liddle, P. F. Aberrant cortical gyrification in schizophrenia: a surface-based morphometry study. *J. Psychiatry Neurosci.* **37**, 399–406 (2012).

159. Nesvåg, R. *et al.* Reduced brain cortical folding in schizophrenia revealed in two independent samples. *Schizophr. Res.* **152**, 333–338 (2014).

160. Madre, M. *et al.* Structural abnormality in schizophrenia versus bipolar disorder: a whole brain cortical thickness, surface area, volume and gyrification analyses. *NeuroImage Clin.* **25**, 102131 (2020).

161. Tan, P. K., Ananyev, E. & Hsieh, P.-J. Distinct genetic signatures of cortical and subcortical regions associated with human memory. *Eneuro* **6**, (2019).

162. Ersland, K. M. *et al.* Gene-based analysis of regionally enriched cortical genes in

GWAS data sets of cognitive traits and psychiatric disorders. *PLoS One* **7**, e31687 (2012).

163. van der Lee, S. J. *et al.* A genome-wide association study identifies genetic loci associated with specific lobar brain volumes. *Commun. Biol.* **2**, 285 (2019).

164. Richman, D. P., Stewart, R. M., Hutchinson, J. & Caviness Jr, V. S. Mechanical Model of Brain Convolutional Development: Pathologic and experimental data suggest a model based on differential growth within the cerebral cortex. *Science (80-.).* **189**, 18–21 (1975).

165. Taki, Y. *et al.* Correlations among brain gray matter volumes, age, gender, and hemisphere in healthy individuals. *PLoS One* **6**, e22734 (2011).

166. Schippling, S. *et al.* Global and regional annual brain volume loss rates in physiological aging. *J. Neurol.* **264**, 520–528 (2017).

167. Fjell, A. M. *et al.* Accelerating cortical thinning: unique to dementia or universal in aging? *Cereb. cortex* **24**, 919–934 (2014).

168. Essen, D. C. van. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature* **385**, 313–318 (1997).

169. Inano, S., Takao, H., Hayashi, N., Abe, O. & Ohtomo, K. Effects of age and gender on white matter integrity. *Am. J. Neuroradiol.* **32**, 2103–2109 (2011).

170. Burzynska, A. Z. *et al.* White matter integrity declined over 6-months, but dance intervention improved integrity of the fornix of older adults. *Front. Aging Neurosci.* 59 (2017).

171. Xu, G. *et al.* Axons pull on the brain, but tension does not drive cortical folding. *J. Biomech. Eng.* **132**, (2010).

172. Madan, C. R. & Kensinger, E. A. Test–retest reliability of brain morphology estimates. *Brain informatics* **4**, 107–121 (2017).

173. Madan, C. R. Age differences in head motion and estimates of cortical morphology. *PeerJ* **2018**, (2018).

174. Wood, R. L. Accelerated cognitive aging following severe traumatic brain injury: A review. *Brain Inj.* **31**, 1270–1278 (2017).

175. Cole, J. H., Leech, R., Sharp, D. J. & Initiative, A. D. N. Prediction of brain age suggests accelerated atrophy after traumatic brain injury. *Ann. Neurol.* **77**, 571–581 (2015).

176. Mende, M. A. Alcohol in the aging brain–the interplay between alcohol consumption, cognitive decline and the cardiovascular system. *Front. Neurosci.* **13**, 713 (2019).

177. Park, D. C. & Reuter-Lorenz, P. The Adaptive Brain: Aging and Neurocognitive Scaffolding. *Annu. Rev. Psychol.* **60**, 173–196 (2009).

178. Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M. & Lindenberger, U. Trajectories of brain aging in middle-aged and older adults: Regional and individual

differences. *Neuroimage* **51**, 501–511 (2010).

179. Schnack, H. G. *et al.* Accelerated Brain Aging in Schizophrenia: A Longitudinal Pattern Recognition Study. *Am. J. Psychiatry* **173**, 607–616 (2016).

180. Wang, J. *et al.* Gray Matter Age Prediction as a Biomarker for Risk of Dementia. *Proc. Natl. Acad. Sci.* **116**, 21213 LP – 21218 (2019).

181. Franke, K., Luders, E., May, A., Wilke, M. & Gaser, C. Brain maturation: Predicting individual BrainAGE in children and adolescents using structural MRI. *Neuroimage* **63**, 1305–1312 (2012).

182. Liem, F. *et al.* Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage* **148**, 179–188 (2017).

183. Kolbeinsson, A. *et al.* Accelerated MRI-predicted brain ageing and its associations with cardiometabolic and brain disorders. *Sci. Rep.* **10**, 19940 (2020).

184. Dinsdale, N. K. *et al.* Learning patterns of the ageing brain in MRI using deep convolutional networks. *Neuroimage* **224**, 117401 (2021).

185. Chen, C.-L. *et al.* Validation of neuroimaging-based brain age gap as a mediator between modifiable risk factors and cognition. *Neurobiol. Aging* **114**, 61–72 (2022).

186. Dobson, R. & Giovannoni, G. Multiple sclerosis – a review. *Eur. J. Neurol.* **26**, 27–40 (2019).

187. Murray, M. E. *et al.* Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol.* **10**, 785–796 (2011).

188. Lublin, F. D. *et al.* Defining the clinical course of multiple sclerosis. *Neurology* **83**, 278 LP – 286 (2014).

189. Lam, B., Masellis, M., Freedman, M., Stuss, D. T. & Black, S. E. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers. Res. Ther.* **5**, 1 (2013).

190. Oxtoby, N. P. & Alexander, D. C. Imaging plus X: multimodal models of neurodegenerative disease. *Curr. Opin. Neurol.* **30**, 371–379 (2017).

191. Young, A. L. *et al.* Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat. Commun.* **9**, 4273 (2018).

192. Eshaghi, A. *et al.* Identifying multiple sclerosis subtypes using unsupervised machine learning and MRI data. *Nat. Commun.* **12**, 2078 (2021).

193. Smith, S. M. *et al.* Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *Elife* **9**, e52677 (2020).

194. Ikram, M. A. *et al.* Objectives, design and main findings until 2020 from the Rotterdam Study. *Eur. J. Epidemiol.* **35**, (2020).

195. Tombaugh, T. N. & McIntyre, N. J. The Mini-Mental State Examination: A Comprehensive Review. *J. Am. Geriatr. Soc.* **40**, 922–935 (1992).

196. Koppelmans, V. *et al.* Global and focal white matter integrity in breast cancer survivors 20 years after adjuvant chemotherapy. *Hum. Brain Mapp.* **35**, 889–899 (2014).

197. de Bruijn, R. F. A. G. *et al.* The potential for prevention of dementia across two decades: the prospective, population-based Rotterdam Study. *BMC Med.* **13**, 132 (2015).

198. Wieberdink, R. G., Ikram, M. A., Hofman, A., Koudstaal, P. J. & Breteler, M. M. B. Trends in stroke incidence rates and stroke risk factors in Rotterdam, the Netherlands from 1990 to 2008. *Eur. J. Epidemiol.* **27**, 287–295 (2012).

199. Darweesh, S. K. L., Koudstaal, P. J., Stricker, B. H., Hofman, A. & Ikram, M. A. Trends in the Incidence of Parkinson Disease in the General Population: The Rotterdam Study. *Am. J. Epidemiol.* **183**, 1018–1026 (2016).

200. Hofman, A. *et al.* The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.* **30**, 661–708 (2015).

201. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).

202. Honea, R. A., Vidoni, E., Harsha, A. & Burns, J. M. Impact of APOE on the healthy aging brain: a voxel-based MRI and DTI study. *J. Alzheimers. Dis.* **18**, 553–564 (2009).

203. O'Donoghue, M. C., Murphy, S. E., Zamboni, G., Nobre, A. C. & Mackay, C. E. APOE genotype and cognition in healthy individuals at risk of Alzheimer's disease: A review. *Cortex* **104**, 103–123 (2018).

204. Gottesman, R. F. & Seshadri, S. Risk Factors, Lifestyle Behaviors, and Vascular Brain Health. *Stroke* **53**, 394–403 (2022).

205. Tan, L., Jiang, T., Tan, L. & Yu, J.-T. Toward precision medicine in neurological diseases. *Ann. Transl. Med.* **4**, 104 (2016).

206. Collij, L. E. *et al.* Spatial-Temporal Patterns of β-Amyloid Accumulation. *Neurology* **98**, e1692 LP-e1703 (2022).

207. Vogel, J. W. *et al.* Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat. Med.* **27**, 871–881 (2021).

208. Bos, D. *et al.* Atherosclerotic calcification is related to a higher risk of dementia and cognitive decline. *Alzheimer's Dement.* **11**, 639–647 (2015).

209. Dolan, H. *et al.* Atherosclerosis, dementia, and Alzheimer disease in the Baltimore Longitudinal Study of Aging cohort. *Ann. Neurol.* **68**, 231–240 (2010).

210. Della-Morte, D. *et al.* Carotid intima-media thickness is associated with white matter hyperintensities: the Northern Manhattan Study. *Stroke* **49**, 304–311 (2018).

211. Devantier, T. A. *et al.* White matter lesions, carotid and coronary atherosclerosis in late-onset depression and healthy controls. *Psychosomatics* **57**, 369–377 (2016).

212. Gustavsson, A. *et al.* Midlife atherosclerosis and development of Alzheimer or vascular dementia. *Ann. Neurol.* **87**, 52–62 (2020).

213. Moroni, F. *et al.* Carotid atherosclerosis, silent ischemic brain damage and brain atrophy: a systematic review and meta-analysis. *Int. J. Cardiol.* **223**, 681–687 (2016).

214. Muller, M. *et al.* Carotid atherosclerosis and progression of brain atrophy: the SMART-MR study. *Ann. Neurol.* **70**, 237–244 (2011).

215. Vidal, J.-S. *et al.* Coronary artery calcium, brain function and structure: the AGES-Reykjavik Study. *Stroke* **41**, 891–897 (2010).

216. Odink, A. E. *et al.* Association between calcification in the coronary arteries, aortic arch and carotid arteries: the Rotterdam study. *Atherosclerosis* **193**, 408–413 (2007).

217. Folsom, A. R. *et al.* Coronary artery calcification compared with carotid intima-media thickness in the prediction of cardiovascular disease incidence: the Multi-Ethnic Study of Atherosclerosis (MESA). *Arch. Intern. Med.* **168**, 1333–1339 (2008).

218. Bos, D. *et al.* Calcification in major vessel beds relates to vascular brain disease. *Arterioscler. Thromb. Vasc. Biol.* **31**, 2331–2337 (2011).

219. Bos, D. *et al.* Atherosclerotic calcification relates to cognitive function and to brain changes on magnetic resonance imaging. *Alzheimer's Dement.* **8**, S104–S111 (2012).

220. Romero, J. R. *et al.* Carotid artery atherosclerosis, MRI indices of brain ischemia, aging, and cognitive impairment: the Framingham study. *Stroke* **40**, 1590–1596 (2009).

221. Rosano, C., Naydeck, B., Kuller, L. H., Longstreth Jr, W. T. & Newman, A. B. Coronary artery calcium: associations with brain magnetic resonance imaging abnormalities and cognitive status. *J. Am. Geriatr. Soc.* **53**, 609–615 (2005).

222. Erbay, S. *et al.* Is intracranial atherosclerosis an independent risk factor for cerebral atrophy? A retrospective evaluation. *BMC Neurol.* **8**, 1–8 (2008).

223. van der Toorn, J. E. *et al.* Vertebrobasilar artery calcification: prevalence and risk factors in the general population. *Atherosclerosis* **286**, 46–52 (2019).

224. Bos, D. *et al.* Intracranial carotid artery atherosclerosis: prevalence and risk factors in the general population. *Stroke* **43**, 1878–1884 (2012).

225. Dubost, F. *et al.* Enlarged perivascular spaces in brain MRI: Automated quantification in four regions. *Neuroimage* **185**, 534–544 (2019).

226. Cole, T. J. & Green, P. J. Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat. Med.* **11**, 1305–1319 (1992).

227. Chung, P.-W. *et al.* Intracranial internal carotid artery calcification: a representative for cerebral artery calcification and association with white matter hyperintensities. *Cerebrovasc. Dis.* **30**, 65–71 (2010).

228. Manolio, T. A. *et al.* Relationships of cerebral MRI findings to ultrasonographic carotid atherosclerosis in older adults: the Cardiovascular Health Study. *Arterioscler. Thromb. Vasc. Biol.* **19**, 356–365 (1999).

229. Adams, H. H. H. *et al.* Heritability and genome-wide association analyses of intracranial carotid artery calcification: the Rotterdam study. *Stroke* **47**, 912–917 (2016).

230. Caplan, L. R. The intracranial vertebral artery: a neglected species. *Cerebrovasc. Dis.* **34**, 20–30 (2012).

231. Yaghi, S., Prabhakaran, S., Khatri, P. & Liebeskind, D. S. Intracranial atherosclerotic disease: mechanisms and therapeutic implications. *Stroke* **50**, 1286–1293 (2019).

232. Wu, X. H., Chen, X.-Y., Wang, L. J. & Wong, K. S. Intracranial artery calcification and its clinical significance. *J. Clin. Neurol.* **12**, 253–261 (2016).

233. Sangiorgi, G. *et al.* Arterial calcification and not lumen stenosis is highly correlated with atherosclerotic plaque burden in humans: a histologic study of 723 coronary artery segments using nondecalcifying methodology. *J. Am. Coll. Cardiol.* **31**, 126–133 (1998).

234. Huffman, M. D. *et al.* Cardiovascular Health Behavior and Health Factor Changes (1988–2008) and Projections to 2020. *Circulation* **125**, 2595–2602 (2012).

235. Lloyd-Jones, D. M. *et al.* Defining and Setting National Goals for Cardiovascular Health Promotion and Disease Reduction. *Circulation* **121**, 586–613 (2010).

236. Voortman, T. *et al.* Adherence to the 2015 Dutch dietary guidelines and risk of non-communicable diseases and mortality in the Rotterdam Study. *Eur. J. Epidemiol.* **32**, 993–1005 (2017).

237. Abraham, H. M. A. *et al.* Cardiovascular risk factors and small vessel disease of the brain: Blood pressure, white matter lesions, and functional decline in older persons. *J. Cereb. Blood Flow Metab.* **36**, 132–142 (2015).

238. Pasi, M. & Cordonnier, C. Clinical Relevance of Cerebral Small Vessel Diseases. *Stroke* **51**, 47–53 (2020).

239. Evans, L. E. *et al.* Cardiovascular comorbidities, inflammation, and cerebral small vessel disease. *Cardiovasc. Res.* **117**, 2575–2588 (2021).

240. Fotuhi, M., Do, D. & Jack, C. Modifiable factors that alter the size of the hippocampus with ageing. *Nat. Rev. Neurol.* **8**, 189–202 (2012).

241. Flowers, S. A. & Rebeck, G. W. APOE in the normal brain. *Neurobiol. Dis.* **136**, 104724 (2020).

242. Lyall, D. M. *et al.* Association between APOE e4 and white matter hyperintensity volume, but not total brain volume or white matter integrity. *Brain Imaging Behav.* **14**, 1468–1476 (2020).

243. Lind, L., Sundström, J., Ärnlöv, J. & Lampa, E. Impact of Aging on the Strength of Cardiovascular Risk Factors: A Longitudinal Study Over 40 Years. *J. Am. Heart*

Assoc. **7**, e007061 (2018).

244. Deal, J. A. *et al.* Hearing impairment and incident dementia and cognitive decline in older adults: the Health ABC Study. *J Gerontol A Biol Sci Med Sci* **72**, (2017).

245. Lin, F. R. *et al.* Hearing loss and cognition in the Baltimore Longitudinal Study of Aging. *Neuropsychology* **25**, (2011).

246. Lin, F. R. *et al.* Hearing loss and incident dementia. *Arch Neurol* **68**, (2011).

247. Thomson, R. S., Auduong, P., Miller, A. T. & Gurgel, R. K. Hearing loss as a risk factor for dementia: A systematic review. *Laryngoscope Investig. Otolaryngol.* **2**, 69–79 (2017).

248. Bernabei, R. *et al.* Hearing loss and cognitive decline in older adults: questions and answers. *Aging Clin Exp Res* **26**, (2014).

249. Gates, G. A. & Mills, J. H. Presbycusis. *Lancet* **366**, (2005).

250. Deal, J. A. *et al.* Hearing Impairment and Cognitive Decline: A Pilot Study Conducted Within the Atherosclerosis Risk in Communities Neurocognitive Study. *Am. J. Epidemiol.* **181**, 680–690 (2015).

251. Fortunato, S. *et al.* A review of new insights on the association between hearing loss and cognitive decline in ageing. Ipoacusia e declino cognitivo: revisione della letteratura. *Acta Otorhinolaryngol Ital* **36**, (2016).

252. Fritze, T. *et al.* Hearing impairment affects dementia incidence. An analysis based on longitudinal health claims data in Germany. *PLoS One* **11**, (2016).

253. Lin, F. R. *et al.* Hearing Loss and Cognitive Decline in Older Adults. *JAMA Intern. Med.* **173**, 293–299 (2013).

254. Wayne, R. V & Johnsrude, I. S. A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Res Rev* **23**, (2015).

255. Fischer, M. E. *et al.* Age-related sensory impairments and risk of cognitive impairment. *J Am Geriatr Soc* **64**, (2016).

256. Gallacher, J. *et al.* Auditory threshold, phonologic demand, and incident dementia. *Neurology* **79**, (2012).

257. Gates, G. A. *et al.* Central auditory dysfunction, cognitive dysfunction, and dementia in older people. *Arch Otolaryngol Head Neck Surg* **122**, (1996).

258. Gillingham, S. M., Vallesi, A., Pichora-Fuller, M. K. & Alain, C. Older adults with hearing loss have reductions in visual, motor and attentional functioning. *Front Aging Neurosci* **10**, (2018).

259. Gurgel, R. K. *et al.* Relationship of hearing loss and dementia: a prospective, population-based study. *Otol Neurotol* **35**, (2014).

260. Heywood, R. *et al.* Hearing loss and risk of mild cognitive impairment and dementia: findings from the Singapore Longitudinal Ageing Study. *Dement Geriatr Cogn Disord* **43**, (2017).

261. Lin, M. Y. *et al.* Vision impairment and combined vision and hearing impairment predict cognitive and functional decline in older women. *J Am Geriatr Soc* **52**, (2004).

262. Merten, N., Fischer, M. E., Tweed, T. S., Breteler, M. M. B. & Cruickshanks, K. J. Associations of hearing sensitivity, higher-order auditory processing, and cognition over time in middle-aged adults. *J Gerontol A Biol Sci Med Sci.* **75**, (2020).

263. Ray, J., Popli, G. & Fell, G. Association of Cognition and Age-Related Hearing Impairment in the English Longitudinal Study of Ageing. *JAMA Otolaryngol. Neck Surg.* **144**, 876 (2018).

264. Tay, T. *et al.* Sensory and cognitive association in older persons: findings from an older Australian population. *Gerontology* **52**, (2006).

265. Thomas, P. D. *et al.* Hearing acuity in a healthy elderly population: effects on emotional, cognitive, and social status. *J Gerontol* **38**, (1983).

266. Jagger, C., Clarke, M., Anderson, J. & Battcock, T. Misclassification of dementia by the mini-mental state examination–are education and social class the only factors? *Age Ageing* **21**, (1992).

267. Rabbitt, P. M. A. *et al.* The University of Manchester Longitudinal Study of Cognition in Normal Healthy Old Age, 1983 through 2003. *Aging Neuropsychol Cogn.* **11**, (2004).

268. Rutherford, B. R., Brewster, K., Golub, J. S., Kim, A. H. & Roose, S. P. Sensation and psychiatry: linking age-related hearing loss to late-life depression and cognitive decline. *Am J Psychiatry* **175**, (2018).

269. Moore, D. R. *et al.* Relation between speech-in-noise threshold, hearing loss and cognition from 40–69 years of age. *PLoS One* **9**, (2014).

270. Hood, J. D. The principles and practice of bone conduction audiometry: A review of the present position. *Laryngoscope* **70**, (1960).

271. Stevens, G. *et al.* Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. *Eur J Public Heal.* **23**, (2013).

272. Koole, A. *et al.* Using the digits-in-noise test to estimate age-related hearing loss. *Ear Hear* **37**, (2016).

273. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, (1950).

274. Scarpina, F. & Tagini, S. The Stroop color and word test. *Front Psychol* **8**, (2017).

275. Swerdlik, M. E. & Cohen, R. J. *Psychological testing and assessment: an introduction to tests and measurement*. (McGraw-Hill, 2005).

276. Armstrong, N. M., Deal, J. A. & Betz, J. Associations of hearing loss and depressive symptoms with incident disability in olderadults: health, aging, and body composition study. *J Gerontol A Biol Sci Med Sci.* **75**, (2020).

277. Yoo, M. *et al.* Moderate hearing loss is related with social frailty in a community-dwelling older adults: The Korean Frailty and Aging Cohort Study (KFACS). *Arch*

Gerontol Geriatr **83**, (2019).

278. Gates, G. A. *et al.* Executive dysfunction and presbycusis in older persons with and without memory loss and dementia. *Cogn Behav Neurol* **23**, (2010).

279. Robertson, D. A., Savva, G. M. & Kenny, R. A. Frailty and cognitive impairment–a review of the evidence and causal mechanisms. *Ageing Res Rev* **12**, (2013).

280. Deal, J. A. *et al.* Hearing treatment for reducing cognitive decline: design and methods of the aging and cognitive health evaluation in elders randomized controlled trial. *Alzheimer's Dement (New York, NY)* **4**, (2018).

281. Kilander, L., Nyman, H., Boberg, M., Hansson, L. & Lithell, H. Hypertension is related to cognitive impairment: a 20-year follow-up of 999 men. *Hypertension* **31**, (1998).

282. Kilander, L., Nyman, H., Boberg, M. & Lithell, H. The association between low diastolic blood pressure in middle age and cognitive function in old age. A population-based study. *Age Ageing* **29**, (2000).

283. Launer, L. J., Masaki, K., Petrovitch, H., Foley, D. & Havlik, R. J. The association between midlife blood pressure levels and late-life cognitive function. The Honolulu-Asia Aging Study. *JAMA* **274**, (1995).

284. Walker, K. A., Power, M. C. & Gottesman, R. F. Defining the relationship between hypertension, cognitive decline, and dementia: a review. *Curr Hypertens Rep* **19**, (2017).

285. Levine, D. A. *et al.* Blood pressure and cognitive decline over 8 years in middle-aged and older black and white Americans. *Hypertension* **73**, (2019).

286. Potvin, O., Mouiha, A., Dieumegarde, L., Duchesne, S. & Initiative, A. D. N. Normative data for subcortical regional volumes over the lifetime of the adult human brain. *Neuroimage* **137**, 9–20 (2016).

287. Potvin, O., Dieumegarde, L., Duchesne, S. & Initiative, A. D. N. Normative morphometric data for cerebral cortical areas over the lifetime of the adult human brain. *Neuroimage* **156**, 315–339 (2017).

288. Peterson, M., Warf, B. C. & Schiff, S. J. Normative human brain volume growth. *J. Neurosurg. Pediatr.* **21**, 478–485 (2018).

289. Rummel, C. *et al.* A fully automated pipeline for normative atrophy in patients with neurodegenerative disease. *Front. Neurol.* **8**, 727 (2018).

290. Tutunji, R. *et al.* Thalamic volume and dimensions on MRI in the pediatric population: Normative values and correlations: (A cross sectional study). *Eur. J. Radiol.* **109**, 27–32 (2018).

291. Brewer, J. B. Fully-automated volumetric MRI with normative ranges: translation to clinical practice. *Behav. Neurol.* **21**, 21–28 (2009).

292. Ross, D. E., Ochs, A. L., Seabaugh, J. M., Shrader, C. R. & Initiative, A. D. N. Man versus machine: comparison of radiologists' interpretations and NeuroQuant® volumetric analyses of brain MRIs in patients with traumatic brain injury. *J.*

Neuropsychiatry Clin. Neurosci. **25**, 32–39 (2013).

293. Ross, D. E. *et al.* Man versus machine part 2: comparison of radiologists' interpretations and NeuroQuant measures of brain asymmetry and progressive atrophy in patients with traumatic brain injury. *J. Neuropsychiatry Clin. Neurosci.* **27**, 147–152 (2015).

294. Vernooij, M. W. *et al.* Automatic normative quantification of brain tissue volume to support the diagnosis of dementia: A clinical evaluation of diagnostic accuracy. *NeuroImage Clin.* **20**, 374–379 (2018).

295. Cavedo, E. *et al.* Fully automatic MRI-based hippocampus volumetry using FSL-FIRST: Intra-scanner test-retest stability, inter-field strength variability, and performance as enrichment biomarker for clinical trials using prodromal target populations at risk for Alzheimer's . *J. Alzheimer's Dis.* **60**, 151–164 (2017).

296. Heinen, R. *et al.* Robustness of automated methods for brain volume measurements across different MRI field strengths. *PLoS One* **11**, e0165719 (2016).

297. Maclaren, J., Han, Z., Vos, S. B., Fischbein, N. & Bammer, R. Reliability of brain volume measurements: a test-retest dataset. *Sci. data* **1**, 1–9 (2014).

298. Tudorascu, D. L. *et al.* Reproducibility and Bias in Healthy Brain Segmentation: Comparison of Two Popular Neuroimaging Platforms . *Frontiers in Neuroscience* vol. 10 (2016).

299. Velasco-Annis, C., Akhondi-Asl, A., Stamm, A. & Warfield, S. K. Reproducibility of Brain MRI Segmentation Algorithms: Empirical Comparison of Local MAP PSTAPLE, FreeSurfer, and FSL-FIRST. *J. Neuroimaging* **28**, 162–172 (2018).

300. Tondelli, M. *et al.* Structural MRI changes detectable up to ten years before clinical Alzheimer's disease. *Neurobiol. Aging* **33**, 825.e25-825.e36 (2012).

301. Kälin, A. M. *et al.* Subcortical shape changes, hippocampal atrophy and cortical thinning in future Alzheimer's disease patients. *Front. Aging Neurosci.* **9**, 38 (2017).

302. Roh, J. H. *et al.* Volume reduction in subcortical regions according to severity of Alzheimer's disease. *J. Neurol.* **258**, 1013–1020 (2011).

303. Štepán-Buksakowska, I. *et al.* Cortical and Subcortical Atrophy in Alzheimer Disease: Parallel Atrophy of Thalamus and Hippocampus. *Alzheimer Dis. Assoc. Disord.* **28**, (2014).

304. Wenzel, F. *et al.* Rapid fully automatic segmentation of subcortical brain structures by shape-constrained surface adaptation. *Med. Image Anal.* **46**, 146–161 (2018).

305. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Med.* **12**, e1001779 (2015).

306. Yeo, I. & Johnson, R. A. A new family of power transformations to improve normality or symmetry. *Biometrika* **87**, 954–959 (2000).

307. Yee, T. W. The VGAM Package for Categorical Data Analysis. *J. Stat. Softw.* **32**,

1–34 (2010).

308. Carpenter, J. & Bithell, J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Stat. Med.* **19**, 1141–1164 (2000).

309. Rousselet, G. A., Pernet, C. R. & Wilcox, R. R. Beyond differences in means: robust graphical methods to compare two groups in neuroscience. *Eur. J. Neurosci.* **46**, 1738–1748 (2017).

310. Organization, W. H. Global action plan on the public health response to dementia 2017–2025. (2017).

311. Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet. Med.* **18**, 421–430 (2016).

312. Holmes, C. Genotype and phenotype in Alzheimer's disease. *Br. J. Psychiatry* **180**, 131–134 (2002).

313. Weintraub, S. *et al.* APOE is a correlate of phenotypic heterogeneity in Alzheimer disease in a national cohort. *Neurology* **94**, e607–e612 (2020).

314. Ferreira, D., Nordberg, A. & Westman, E. Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology* **94**, 436–448 (2020).

315. Ryan, J., Fransquet, P., Wrigglesworth, J. & Lacaze, P. Phenotypic heterogeneity in dementia: a challenge for epidemiology and biomarker studies. *Front. public Heal.* **6**, 181 (2018).

316. Devi, G. & Scheltens, P. Heterogeneity of Alzheimer's disease: consequence for drug trials? *Alzheimers. Res. Ther.* **10**, 1–3 (2018).

317. Sevigny, J. *et al.* The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* **537**, 50–56 (2016).

318. Sperling, R. A., Karlawish, J. & Johnson, K. A. Preclinical Alzheimer disease—the challenges ahead. *Nat. Rev. Neurol.* **9**, 54–58 (2013).

319. Marinescu, R. V *et al.* The alzheimer's disease prediction of longitudinal evolution (TADPOLE) challenge: Results after 1 year follow-up. *arXiv Prepr. arXiv2002.03419* (2020).

320. Klöppel, S. *et al.* Applying automated MR-based diagnostic methods to the memory clinic: a prospective study. *J. Alzheimer's Dis.* **47**, 939–954 (2015).

321. Wachter, S., Mittelstadt, B. & Russell, C. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harv. JL Tech.* **31**, 841 (2017).

322. Holzinger, A., Biemann, C., Pattichis, C. S. & Kell, D. B. What do we need to build explainable AI systems for the medical domain? *arXiv Prepr. arXiv1712.09923* (2017).

323. Donohue, M. C. *et al.* Estimating long-term multivariate progression from short-term data. *Alzheimer's Dement.* **10**, S400–S410 (2014).

324. Fonteijn, H. M. *et al.* An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *Neuroimage* **60**, 1880–1889 (2012).

325. Koval, I. *et al.* Spatiotemporal propagation of the cortical atrophy: Population and individual patterns. *Front. Neurol.* 235 (2018).

326. Venkatraghavan, V., Bron, E. E., Niessen, W. J. & Klein, S. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. *Neuroimage* **186**, 518–532 (2019).

327. Venkatraghavan, V. *et al.* Analyzing the effect of APOE on Alzheimer's disease progression using an event-based model for stratified populations. *Neuroimage* **227**, 117646 (2021).

328. Dubois, B. *et al.* Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *Lancet Neurol.* **6**, 734–746 (2007).

329. Petersen, R. C. *et al.* Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology* **74**, 201–209 (2010).

330. de Bruijn, R. F. A. G. *et al.* Determinants, MRI Correlates, and Prognosis of Mild Cognitive Impairment: The Rotterdam Study. *J. Alzheimer's Dis.* **42**, (2014).

331. Jack, C. R. *et al.* The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* **27**, (2008).

332. Lamballais, S., Vinke, E. J., Vernooij, M. W., Ikram, M. A. & Muetzel, R. L. Cortical gyrification in relation to age and cognition in older adults. *Neuroimage* **212**, 116637 (2020).

333. Archetti, D. *et al.* Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. *NeuroImage Clin.* **24**, (2019).

334. Park, K. M. *et al.* Population-based dementia prediction model using Korean public health examination data: A cohort study. *PLoS One* **14**, (2019).

335. Stephan, B. C. M. *et al.* Usefulness of data from magnetic resonance imaging to improve prediction of dementia: population based cohort study. *BMJ* **350**, h2863–h2863 (2015).

336. Palmqvist, S. *et al.* Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *JAMA* **324**, (2020).

337. Bethlehem, R. A. I. *et al.* Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).

338. Grimm, O. *et al.* Amygdalar and hippocampal volume: A comparison between manual segmentation, Freesurfer and VBM. *J. Neurosci. Methods* **253**, 254–261 (2015).

339. Fischl, B. *et al.* Whole Brain Segmentation: Automated Labeling of Neuroanatomical Structures in the Human Brain. *Neuron* **33**, 341–355 (2002).

340. Morey, R. A. *et al.* A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. *Neuroimage* **45**, 855–866 (2009).

341. Babalola, K. O. *et al.* Comparison and Evaluation of Segmentation Techniques for Subcortical Structures in Brain MRI. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008* (eds. Metaxas, D., Axel, L., Fichtinger, G. & Székely, G.) 409–416 (Springer Berlin Heidelberg, 2008).

342. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J. V & Rueckert, D. Classifier Selection Strategies for Label Fusion Using Large Atlas Databases BT - Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007. in (eds. Ayache, N., Ourselin, S. & Maeder, A.) 523–531 (Springer Berlin Heidelberg, 2007).

343. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907–922 (2011).

344. Murgasova, M. *et al.* Segmentation of Brain MRI in Young Children. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006* (eds. Larsen, R., Nielsen, M. & Sporring, J.) 687–694 (Springer Berlin Heidelberg, 2006).

345. Perlaki, G. *et al.* Comparison of accuracy between FSL's FIRST and Freesurfer for caudate nucleus and putamen segmentation. *Sci. Rep.* **7**, 2418 (2017).

346. Huizinga, W. *et al.* Differences Between MR Brain Region Segmentation Methods: Impact on Single-Subject Analysis. *Front. big data* **4**, 577164 (2021).

347. Garbarino, S. *et al.* Differences in topological progression profile among neurodegenerative diseases from imaging data. *Elife* **8**, (2019).

348. Stern, Y. *et al.* Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimer's Dement.* **16**, 1305–1311 (2020).

349. Lin, S.-H. & Ikram, M. A. On the relationship of machine learning with causal inference. *Eur. J. Epidemiol.* **35**, 183–185 (2020).

350. Keil, A. P. & Edwards, J. K. You are smarter than you think: (super) machine learning in context. *Eur. J. Epidemiol.* **33**, 437–440 (2018).

351. Hernán, M. A., Hsu, J. & Healy, B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE* **32**, 42–49 (2019).

352. van Geloven, N. *et al.* Prediction meets causal inference: the role of treatment in clinical prediction models. *Eur. J. Epidemiol.* **35**, 619–630 (2020).

Appendices

Dankwoord

Het is zover, het dankwoord, waarschijnlijk het meest gelezen hoofdstuk van een proefschrift. Ik wil graag iedereen enorm bedanken die bijgedragen heeft aan dit proefschrift en de reis er naartoe! Graag wil ik nog een aantal mensen in het bijzonder bedanken.

Allereerst mijn promotoren Prof. dr. Meike Vernooij en Prof. dr. Arfan Ikram. Ik herinner me nog goed mijn solicitatiegesprek op de 25ste verdieping, een gesprek waarin ik me verassend genoeg zowel uitgedaagd als thuis voelde, een voorbode voor de jaren die daarop volgden! Meike, de manier waarop jij in een fractie van een seconde door hebt hoe de vork in de steel zit én dit direct zeer treffend onder woorden weet te brengen is een talent die niet te evenaren is. Heel veel dank voor je begeleiding van mijn promotie traject en ook voor je betrokkenheid daarbuiten! Jouw leiderschaps- en begeleidingsstijl is een voorbeeld voor velen! Arfan, jouw altijd scherpe blik heeft mij steeds opnieuw gemotiveerd om me de stof en methodologie nog meer eigen te maken. Met als resultaat zeer levendige discussies (of zoals Meike het laatst noemde: de box-ring). Eerlijk is eerlijk, 'Verdorie, Arfan heeft wéér gelijk' was vaak mijn conclusie, welliswaar met een grijns op mijn gezicht. Ontzettend bedankt voor je begeleiding!

Graag wil ik de promotiecommissie bedanken voor het lezen en het beoordelen van mijn proefschrift. Ik kijk er naar uit met jullie van gedachten te wisselen tijdens mijn verdediging.

My dear EuroPOND'ers, first of all the BIGR team, Vikram, Stefan, Esther, and Wiro. Thank you all for the great collaborations, the lively discussions and of course the good company during the consortium meetings. Vikram, I really enjoyed working with you. Our projects were definitely challenging, but with your patience, open mindedness, and drive to get it done, we did! I hope that more collaborations lie ahead!

To all EuroPOND'ers, it was an honour to be part of such a successful multidisciplinary consortium, where acquiring knowledge and collaborations was always central, you set the bar high! A special thanks to Daniel Alexander and Neil Oxtoby for welcoming me at UCL and letting me take part in the CMIC group during my stay in London.

Dank aan alle (oud-)neuro-epi en PopI collega's, ik denk nu al met heimwee terug aan die heerlijke jaren op de 28ste verdieping. Dank voor al jullie steun, gezelligheid, inspiratie, inhoudelijke (en niet inhoudelijke) discussies en goede gesprekken, wat voelde ik mij thuis!

Aan alle imaging helden van de radiologie, in het bijzonder Marcel, Hakim, Henri en Marius. Marius, in het begin van mijn PhD heb jij me leren navigeren in het doolhof genaamd image processing en mij geïnspireerd om mijn programeer-skills uit te breiden, waarvoor veel dank! Marcel, Hakim en Henri, heel veel dank voor al jullie inzet, hulp en geduld in onze reis naar het gebruik van de ViewR. Dank voor de fijne samenwerking en niet te vergeten dank dat jullie mij kennis hebben laten maken met de wondere wereld van o.a. object oriented programming en de betekenis van TLDR.

I also would like to thank my collaborators Fabian Wenzel, Martin Berghtoldt, Wyke Huizinga, and others I perhaps forgot to mention. Thank you for all for your help and discussions, I enjoyed our collaborations.

Beste Aad, heel veel dank dat ik onderdeel uit mocht maken van de imaging infrastructuur groep. Ik heb veel geleerd en met veel plezier deel uit gemaakt van het team, zelfs al was het grotendeels in tijden van lock-downs.

Lieve paranimfen, Pauline en Christel, wat een geluk dat ik jullie aan mijn zijde heb! Pauline, elke dag dat jij niet meer op de afdeling werkt, realiseer ik me opnieuw wat heerlijk het was om met jou op een kamer te zitten. Met jou is elke dag een feestje: meerdere keren per dag de slappe lach, een goede dosis zelfspot, maar ook goede inhoudelijke discussies, en altijd oog voor een ander en een luisterend oor kunnen bij jou niet ontbreken. Ik kijk nu al uit naar hoe jij mijn zenuwen kunt onderbreken met wat goeie (of slechte) grappen in het zweetkamertje!

Lieve Christel, mijn allerliefste en allerstoerste zusje, wat een mazzelaar ben ik dat ik jou in mijn leven heb. Mijn PhD tijd omvat een aantal mooie en bijzondere mijlpalen van ons leven, wat fijn dat jij bij deze mijlpaal naast me staat! Op naar nog vele mooie mijlpalen in de toekomst!

Lieve familie en schoonfamilie, ontzettend bedankt voor al jullie steun! Pap en mam, hier ligt het dan, het boekje waar ik het al zo'n 6 jaar over heb. Iets wat toch altijd ongrijpbaar blijft, wanneer jullie naar de juiste woorden zoeken als iemand vraagt: "Wat doet Eline nou eigenlijk?". Het is tastbaar geworden, met oma voorop!

Lieve Roelant, terwijl ik dit schrijf staat jouw bul achter me te schitteren op de hoek van de kast, ik hoor Rutger en Merle het huis op z'n kop zetten, terwijl jij na weer een veel te korte nacht een poging doet om een kop koffie te zetten. Ik kan zoveel redenen bedenken waar ik je dankbaar voor ben, teveel op hier op te noemen. Jij maakt het verschil… en zonder jou lag dit boekje er niet.

List of Publications

Manuscripts that form the basis of this thesis

Chapter 2

Vinke EJ, de Groot M, Venkatraghavan V, Klein S, Niessen WJ, Ikram MA, Vernooij MW. Trajectories of imaging markers in brain aging: The Rotterdam Study. Neurobiology of Aging. 2018;71:32-40

Licher S*, van der Willik KD*, **Vinke EJ**, Knol MJ, Darweesh SKL, van der Geest JN, Schagen SB, Ikram MK, Luik AI, Ikram MA. Trajectories of cognitive and motor function between ages 45 and 90 years: a population-based study. J Gerontol A Biol Sci Med Sci. 2021;76(2):297-306.

Lamballais S, **Vinke EJ**, Vernooij MW, Ikram MA, Muetzel RL. Cortical gyrification in relation to age and cognition in older adults. NeuroImage. 2020;212

Vinke EJ, Young A, Oxtoby NP, Alexander DC, Venkatraghavan V, Ikram MA, Vernooij MW. Identifying subtypes in 'normal' brain ageing through data-driven disease progression models. *Submitted to Brain*

Vinke EJ, Ikram MA, Vernooij MW. Brain Aging: more of the same!? Aging (Albany NY). 2019;11(3):849

Chapter 3

Vinke EJ*, Yilmaz P*, van der Toorn JE, Fakhry R, Frenzen K, Dubost F, Licher S, de Bruijne M, Kavousi M, Ikram MA, Vernooij MW, Bos D. Intracranial arteriosclerosis is related to cerebral small vessel disease and accelerated brain atrophy: a prospective cohort study. Neurobiology of Aging. 2021;105:16-24

Vinke EJ, Bron EE, Voortman T, Ikram MA, Vernooij MW. Cardiovascular health in relation to trajectories of imaging markers in brain ageing. *Submitted to Age and Ageing*

Croll PH*, **Vinke EJ***, Armstrong NM, Licher S, Vernooij MW, Baatenburg de Jong RJ, Goedegebure A, Ikram MA. Hearing loss and cognitive decline in the general population: a prospective cohort study. Journal of Neurology. 2021;268(3):860-871

Chapter 4

Vinke EJ, Huizinga W, Bergtholdt M, Adams HH, Steketee RME, Papma JM, de Jong FJ, Niessen WJ, Ikram MA, Wenzel F, Vernooij MW, Alzheimer's Disease Neuroimaging Initiative. Normative brain volumetry derived from different reference populations: impact on single-subject diagnostic assessment in dementia. Neurobiology of Aging. 2019;84:9-16

Venkatraghavan V*, **Vinke EJ***, Bron EE, Niessen WJ, Ikram MA, Klein S, Vernooij MW, Alzheimer's Disease Neuroimaging Initiative. Progression along APOE-specific data-driven temporal cascades is predictive of Alzheimer's disease in a population-based cohort. NeuroImage. 2021;238

Other publications

Huizinga W, Poot DHJ, **Vinke EJ**, Wenzel F, Bron EE, Toussaint N, Ledig C, Vrooman H, Ikram MA, Niessen WJ, Vernooij MW*, Klein S*. Differences Between MR Brain Region Segmentation Methods: Impact on Single-Subject Analysis. Frontiers in Big Data. 2021;4

Season of birth and the risk of dementia in the population-based Rotterdam Study. Mooldijk SS, Licher S, **Vinke EJ**, Vernooij MW, Ikram MK, Ikram MA. European journal of epidemiology. 2021;35(5):497-506

van der Willik KD, Hauptmann M, Jóźwiak, **Vinke EJ**, Ruiter R, Stricker BH, Compter A, Ikram MA, Schagen SB. Trajectories of cognitive function prior to cancer diagnosis: a population-based study. JNCI: Journal of the National Cancer Institute. 2020;112(5):480-488

Van Arendonk J, Wolters FJ, Neitzel J, **Vinke EJ**, Vernooij MW, Ikram MA, Ghanbari M. Plasma Neurofilament light chain (NfL) relates to preclinical changes in cognition, structural white matter integrity and markers of cerebral small-vessel disease: A population-based study. Alzheimer's & Dementia. 2021;17:e053611

Vilor-Tejedor N, Ikram MA, Roshchupkin R, **Vinke EJ**, Vernooij MW, Adams HH. Aging-Dependent Genetic Effects Associated to ADHD Predict Longitudinal Changes of Ventricular Volumes in Adulthood. Frontiers in psychiatry. 2020;11:574

Licher S, van der Willik KD, **Vinke EJ**, Yilmaz P, Fani L, Schagen SB, Ikram MA, Ikram MK. Alzheimer's disease as a multistage process: an analysis from a population-based cohort study. Aging (Albany NY). 2019;11(4)

Garbarino S, Lorenzi M, Oxtoby NP, **Vinke EJ**, Marinescu RV, Eshaghi A, Ikram MA, Niessen WJ, Ciccarelli O, Barkhof F, Schott JM, Vernooij MW, Alexander DC. Differences in topological progression profile among neurodegenerative diseases from imaging data. Elife. 2019;8

Van den Brule JMD, Stolk R, **Vinke EJ**, van Loon LM, Pickkers P, van der Hoeven JG, Kox M, Hoedemaekers CWE. Vasopressors do not influence cerebral critical closing pressure during systemic inflammation evoked by experimental endotoxemia and sepsis in humans. Schock: Injury, Inflammation, and Sepsis: Laboratory and Clinical Approaches. 2018;49(5):529-535

van den Brule JMD, **Vinke EJ**, van Loon LM, van der Hoeven JG, Hoedemaekers CWE. Low spontaneous variability in cerebral blood flow velocity in non-survivors after cardiac arrest. Resuscitation. 2017;111:110-115

Middle cerebral artery flow, the critical closing pressure, and the optimal mean arterial pressure in comatose cardiac arrest survivors-An observational study. van den Brule JMD, **Vinke EJ**, van Loon LM, van der Hoeven JG, Hoedemaekers CWE. Resuscitation. 2017;110:85-89

Vinke EJ, Kortenbout AJ, Eyding J, Slump CH, van der Hoeven JG, de Korte CL, Hoedemaekers CWE. The potential of contrast enhanced ultrasound as a bedside monitoring technique of the cerebral perfusion: a systematic review. Ultrasound in Medicine & Biology. 2017;43(12):2751-2757

Vinke EJ, Eyding J, de Korte CL, Slump CH, van der Hoeven JG, Hoedemaekers CWE. Repeatability of bolus kinetics ultrasound perfusion imaging for the quantification of cerebral blood flow. Ultrasound in Medicine & Biology. 2017;43(12):2758-2764

Vinke EJ, Eyding J, de Korte C, Slump CH, van der Hoeven JG*, Hoedemaekers CWE*. Quantification of macrocirculation and microcirculation in brain using ultrasound perfusion imaging. Part of the Acta Neurochirurgica Supplement book series (NEUROCHIRURGICA volume 126), Intracranial Pressure & Neuromonitoring XVI

Kruse RR, **Vinke EJ**, Poelmann FB, Rohof D, Holewijn S, Slump CH, Reijnen MMPJ. Computation of blood flow through collateral circulation of the superficial femoral artery. Vascular. 2016;24(2):126-133

*These authors contributed equally to the respective manuscript.

Portfolio

Name	Elisabeth Janine Vinke
Departments	Radiology and Nuclear Medicine, Epidemiology Erasmus MC - University Medical Center Rotterdam
Research school	Netherlands Institute for Health Sciences (NIHES)
PhD period	2016-2022
Promotors	Prof. dr. MW Vernooij Prof. dr. MA Ikram

PhD training	**Year**	**ECTS***
Master of Science in Clinical Epidemiology (NIHES)	2016-2018	70
The course on R (Molecular Medicine, Rotterdam)	2016	1.8
Praktische Neuroanatomie en Neuroradiologie (VUmc)	2018	0.6
Scientific Integrity Course (Erasmus MC)	2018	0.3
Department research seminars	2016-2022	2.0
Journal club seminars	2016-2022	0.7
Conferences		
VasCog, Amsterdam, The Netherlands	2016	0.6
European Congress of Radiology (ECR), Vienna, Austria: oral presentation & pitch	2018	1.5
Dutch Epidemiology Conference (WEON), Groningen, The Netherlands: oral presentation	2019	1.2
Health-Holland Visitors Programme, Rotterdam - the ageing society track: Population Imaging workshop leader		0.6
Computational approaches of ageing and age-related diseases (CompAge): poster presentation		1.2
Teaching activities		
Supervising Junior Med School students	2017-2018	2.0
Teaching assistant ESP01 Principles of Research in Medicine and Epidemiology	2021	0.5
Supervision Bachelor Assignment Clinical Technology	2021	2.0
Jury Bachelor Assignment Clinical Technology	2021-2022	0.3
Other activities		
Reviewing activities for peer-reviewed journals	2018-2022	1.0
Secretary, PR and Vice-chairman of Technical Innovation in Medicne (TiiM) conference 2018	2017-2018	
Logistics manager of Technical Innovation in Medicine (TiiM) conference 2019	2018-2019	

*ECTS: European Credit Transfer System equals a workload of 28 hours

About the Author

Eline Vinke was born on the 7th of November, 1989 in Arnhem, The Netherlands. She graduated from secondary school in 2008 at the Isala College in Silvolde. In September that year she started her bachelors in Technical Medicine at the University Twente. After obtaining her bachelor degree she persued her master Technical Medicine in 'Medical Signalling and Stimulation'. After her master graduation project on the development of a contrast enhanced ultrasound (CEUS) technique to measure cerebral blood flow in patients with acute brain injury, she started her technical medicine fellowship in 2016 focused on neuromonitoring at the neuro intensive care unit at the Radboud UMC in Nijmegen. During this fellowship she discovered that her passion for research outreached that of working in the clinic. In August 2016 she therefore started her work described in this thesis, under the supervision of prof. dr. MW Vernooij and prof. dr. MA Ikram, at the department of Radiology and Nuclear Medicine and the department of Epidemiology of the Erasmus MC. As part of her research training she obtained her master degree in Clinical Epidemiology from the Netherlands Institute for Health Sciences. During her PhD, Eline was intensively involved in imaging data management of the Rotterdam Study. In 2019, Eline started a part-time position as imaging coordinator of Biobanking and BioMolecular resources Research Infrastructure The Netherlands (BBMRI.nl), where she used her expertise in population imaging and clinical epidemiology to contribute to the development of a unified imaging data infrastructure, which would support development of biomedical imaging technologies and the dissemination of knowledge, with the ultimate goal of improving diagnosis, treatment and prevention of disease. Besides research, Eline enjoys contributing to (science) education in the form of organizing conferences, teaching activities and in January 2022 she started as a member of the Curriculum Committee of the bachelor Clinical Technology, responsible for the bachelor curriculum update, in which Eline focuses on Academic Education.