Memory and Neurodegeneration Across the Lifespan

Doctoral thesis

to obtain a doctorate (PhD)

from the Faculty of Medicine

of the University of Bonn

Meta Miriam Bönniger

from Cologne, Germany

2021

Written with authorization of

the Faculty of Medicine of the University of Bonn

First reviewer: Prof Dr Dr Monique M.B. Breteler

Second reviewer: Prof Dr Carole Dufouil

Day of oral examination: 30.08.2021

For the German Center for Neurodegenerative Diseases (DZNE e.V.), Population Health Sciences

Director: Prof Dr Dr Monique M.B. Breteler

Table of content

List of abbreviations

AC-PC	anterior commissure-posterior commissure
AD	Alzheimer's disease
ADCS-PACC	Alzheimer's Disease Cooperative Study Preclinical Alzheimer's Cognitive Composite
ATC	Anatomical Therapeutic Chemical
AVLT	Rey's Auditory Verbal Learning Test
BMBF	Federal Ministry of Education and Research
BOLD	blood oxygen level-dependent
c	response bias
CI	confidence interval
d'	d-prime
DZNE	German Center for Neurodegenerative Diseases
EPI	echo-planar imaging
ESA	encoding success activity
FA	false alarm
FDR	false discovery rate
FLAIR	fluid-attenuated inversion recovery
fMRI	functional magnetic resonance imaging
FOV	field of view
FWHM	full-width-half-maximum
GCL	ganglion cell layer
GCP	Good Clinical Practice
HCC	hair cortisol concentration

HPA	hypothalamic-pituitary-adrenal
ICC	intraclass correlation coefficient
ICH	International Council for Harmonization
ICH-GCP	International Council for Harmonization Good Clinical Practice
IMBIE	Institute for Medical Biometry, Informatics and Epidemiology
IQR	interquartile range
ISCED	International Standard Classification of Education
LC-MS/MS	liquid chromatography tandem mass spectrometry
M	mean
mGCL	macular ganglion cell layer
MMSE	Mini-Mental State Examination
MNI	Montreal Neurological Institute
MoCA	Montreal Cognitive Assessment
MRI	magnetic resonance imaging
MWT-B	Mehrfachwahl-Wortschatz Intelligenztest
OCT	optical coherence tomography
OR	odds ratio
OxVoc	Oxford Vocal
pRNFL	peripapillary retinal nerve fiber layer
PSS	Cohen’s Perceived Stress Scale
RNFL	retinal nerve fiber layer
RR	rate ratio
SD	standard deviation
SD-OCT	spectral-domain optical coherence tomography

SUBTLEX-DE	German subtitle-based word frequency
SWI	susceptibility-weighted imaging
TE	echo time
TMT	Trail-making test
TR	repetition time
VLMT	Verbaler Lern- und Merkfähigkeitstest

1 Abstract

The prevention and treatment of age-related memory impairment requires the early detection of underlying pathological changes and the availability of effective risk modifying intervention. This, in turn, demands research into biomarkers and risk factors of cognitive performance over the life course. Prospective cohort studies are well suited for examining biomarkers and risk factors. However, life-course approaches to cognitive performance require cognition tests that are applicable to a wide age range of participants, provide reliable and valid results after multiple administration, require limited examination time, and yield rich data to allow exploring a broad range of potential research hypotheses. When setting up the cognitive assessment battery for the Rhineland Study, a brain focused prospective cohort study which aims to investigate causes and biomarkers of healthy aging and neurodegeneration, the need for memory tasks that meet these requirements became apparent. Therefore, I aimed to develop sensitive memory measurements for use in prospective cohort studies. I successfully created nine equally difficult versions of Rey's Auditory Verbal Learning Test (AVLT), a sensitive, well-known and widely used measure of episodic verbal memory which is subject to learning effects in case of multiple testing. Furthermore, I developed a functional magnetic resonance imaging (fMRI) task that shows visual and auditory specific sensory brain activity as well as sensory-specific and -unspecific memory-encoding-associated brain activity within only ten minutes of fMRI acquisition time. Finally, we compiled the Rhineland Study cognitive assessment battery, which consists of the AVLT as our main memory examination together with other tests that examine other cognitive domains.

To illustrate the use of such a cognitive assessment battery, in the second part of my thesis we examined the association between retinal layers as potential biomarkers and chronic stress as a potential risk factor of neurodegenerative processes. We found small but significant associations between the macular ganglion cell layer (mGCL) volume and global cognitive function, processing speed and episodic verbal memory independent of age and other influencing factors. Perceived stress also showed associations with all cognitive domains, especially working memory and executive functions. We conclude that mGCL volume may be a potential biomarker and perceived stress a potential risk factor for memory decline, but longitudinal studies are needed.

2 General introduction

To find prevention and treatment methods for age-related memory changes, the identification of biomarkers and risk factors is necessary. However, in order to explore potential biomarkers and risk factors, methods for early detection of differences and changes in memory functions are required. Population-based prospective studies, like the Rhineland Study, are an important method for identifying such potential biomarkers and risk factors, as they can follow a large sample of participants over time. However, such studies have special requirements due to their design for memory and cognition tasks. Since there is usually only a limited amount of study time per participant, tests must provide as much information as possible in the shortest possible time, and they must be applicable to a wide age range. They should also provide reliable and valid results after repeated administration to capture change in performance over time. Versatility in results is important, as prospective studies usually address a broad range of questions that may change over time. When compiling the cognitive assessment battery for the Rhineland Study, a prospective population-based study with a focus on brain aging, we noticed that no memory task met all these criteria, so I decided to create and adapt two tasks measuring memory performance with different methods.

A common, inexpensive and relatively easy to use method to assess memory function is a neuropsychological examination. Rey's Auditory Verbal Learning Test (AVLT),[1,2] for example, is one of the most popular tests of episodic verbal memory. It measures different levels of memory and learning using lists of semantically unrelated words that participants must learn and remember. Over five repetitions of learning and free recall, participants learn a list of semantically unrelated nouns. Then, another list is read to the participant who freely recalls it once, followed by a single recall of the previous list. Finally, after 20-30 minutes, a delayed free recall is performed. General free recall episodic memory performance, and specifically AVLT performance, has been shown to be a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease (AD) and an early indicator for the risk of developing dementia.[3–8] Further, the test provides a wide range of information about learning and memory performance and is relatively time efficient. However, in case of repeated measurements it is affected by a memory bias and therefore requires parallel forms for reexaminations.[9,10] Up to now, there was mainly one test version available in Germany which provided two parallel versions. However, these

versions have only been rudimentarily tested for parallelism.[11] To allow for multiple repeated measures of an individual's episodic verbal memory, as we plan to do in the Rhineland Study, I developed ten German word list versions and tested them for equal difficulty (Chapter 3.1).

An alternative to traditional methods is the combination of imaging techniques (e.g. functional magnetic resonance imaging (fMRI)) with specifically developed neuropsychological tests. Imaging techniques offer the possibility of investigating brain structures as the biological basis of neuropsychological functions.[12] Most task fMRI protocols however take too long to be easily implemented at a large scale and they usually very specified to one specific cognitive function (e.g. face-name associative memory). We considered that a fMRI task that would measure memory encoding using at least two sensory systems and would take at maximum ten minutes fMRI acquisition time, could possibly be included in the Rhineland Study core protocol. Two sensory systems were to be included in order to be able to investigate sensory independent memory encoding activity. In addition, this would also allow to obtain partial results for participants who have a mild impairment in one of the sensory systems (e.g. hearing or visual impairment). Since we did not find a task that time efficiently measured memory encoding related brain activity across and within sensory conditions, I developed an fMRI task that included two types of auditory (vocal and environmental) and two types of visual (face and scene) stimuli (Chapter 3.2). This enabled us to measure sensory-specific brain activity, as well as sensory-specific and -nonspecific memory-encoding brain activity. In order to show as many stimuli as possible in as little time as possible without reducing time for each trial too much, we used the mixed block event-related design. This allowed us to calculate additional rather experimental contrasts, such as differences between sustained and transient brain activity. In summary, the encoding part of the task, which required fMRI study time, took only ten minutes. The additional recognition part of the task, which is necessary to be able to create memory contrasts required an additional 10 to 19 minutes. However, the recognition task can be performed either inside or outside fMRI. In our study we ran the recognition task during structural and diffusion scans which entertained participants and even improved scan quality with regard to head motion in comparison to resting state condition.[13]

By compiling the cognitive battery for the Rhineland Study, which is described in detail in Chapter 3.3, we had the criteria for prospective cohort studies in mind, but we also aimed to include tests measuring a variety of cognitive domains to measure a participant's neuropsychological functions as complete as possible. Each domain was aimed to be represented by at least two different tests. Similar to our rational for the fMRI task, we selected tests with different sensory stimulus material and behavioural responses to limit the effects of possible sensory and motor impairments on test performance. We also preferred tests which outcomes represent only one domain, are well-studied to increase comparability to other studies, and ease interpretation of test results. Further, to reduce feelings of frustration and failure in participants we preferred tests that have no predefined response scale so participants could not identify the maximum level of performance. Unfortunately, the fMRI task was not included in the final protocol of the Rhineland Study as the more global measure of resting state fMRI was preferred over the rather specific memory task fMRI and both could not be included in the protocol due to time reasons.

The cognitive test battery in the Rhineland Study is used both to identify potential early markers of neurodegeneration and to identify causes of cognitive decline. To illustrate the use of our memory and cognitive measures in identifying potential biomarkers and investigating risk factors we performed two studies. The first examined thickness of two retinal layers in relation to cognitive performance (Chapter 4.1). In the second, we examined physiological and subjective/psychological chronic stress as a potential risk factor for worse cognitive function (Chapter 4.2).

Retinal layers such as the retinal nerve fiber layer (RNFL) and the ganglion cell layer (GCL) are examined as potential biomarkers for neurodegeneration as the retinal tissue has physiological similarities to brain tissue because it is embryologically derived from the cranial part of the neural tube.[14] Retinal measurements are non-invasive and relatively easy to assess using state-of-the-art imaging techniques such as spectral domain optical coherence tomography (SD-OCT). Previous studies reported that thinner RNFL and GCL are associated with an increased risk of cognitive decline and dementia.[14–16] However, the association of these markers with memory and cognitive function has not been studied in depth in the general population and across the adult life-span. Using our cognitive domain scores, we therefore quantified the association between SD-OCT derived retinal

measurements, peripapillary RNFL (pRNFL) thickness and macular GCL (mGCL) volume, with memory and cognitive functions and identified factors influencing these relationships.

A prominent modifiable potential risk factor of neurodegeneration is chronic stress. Long term stress has been shown to have detrimental effects on cognitive function and may also promote the onset and progression of memory and cognitive decline and dementia.[17–19] However, measuring stress is not easy as it is a very complex construct that includes individual and environmental factors that also interact with each other.[20] Previous larger population-based studies examining the association between various stress measures and cognitive functions reported heterogeneous results.[e.g.21–23] Those studies included only one dimension of stress (physiological or subjective). Therefore, we used in 2,000 participants our memory and cognitive examination scores to investigate the association between both subjective/psychological and physiological measures of chronic stress with one another and with memory and cognitive functions.

THESIS AIM AND OUTLINE

The Rhineland Study is a prospective cohort study designed to investigate causes of healthy and unhealthy aging. A particular focus is on brain aging, which requires studying how memory and cognitive performance changes across adulthood.

A main aim of my thesis was therefore to develop memory tasks which can be utilized in the context of prospective cohort studies such as the Rhineland Study, to measure cognitive performance. The designed memory tasks had to address several challenges that are posed by a prospective cohort study, including being applicable to a wide age range of participants, providing reliable and valid results after multiple administration, requiring limited examination time, and yielding rich data to allow addressing a broad range of potential research hypotheses.

For the Rhineland Study cognitive assessment battery, I developed two sensitive memory measurements (Chapter 3). The first study presents ten list versions of the AVLT including detailed age and sex effect estimates for a large number of outcomes (Chapter 3.1). The second study presents a fMRI paradigm measuring sensory-specific and sensory-unspecific memory encoding and visual and auditory sensory encoding within just ten

minutes of fMRI acquisition time (Chapter 3.2). Following these two studies I describe the cognitive assessment battery that we developed and implemented in the Rhineland Study (Chapter 3.3).

In the second part of this thesis (Chapter 4) I present two studies that illustrate the use of our memory and cognitive measures in a large non-demented cross-sectional population-based sample of adults (30-95 years), to first identify potential biomarkers of neurodegeneration, and second investigate risk factors for cognitive decline. In the first project we quantify the association between the thickness of two retinal layers as potential biomarkers, the RNFL and the GCL, and memory and other cognitive functions (Chapter 4.1). The second project presents the association between chronic stress, a modifiable potential risk factor, and memory and other cognitive functions (Chapter 4.2).

In the final chapter, I briefly summarise and discuss my main findings, and provide an outlook for further research (Chapter 5).

REFERENCES

1. Rey A. Mémorisation d'une série de 15 mots en 5 répetitions. In: Rey A, ed. *L'examen Clinique En Psychologie*. 3rd ed. Presses Universitaries de France; 1970:141-193.
2. Taylor EM. 8. Learning and Memory. In: Taylor EM, ed. *Psychological Appraisal of Children with Cerebral Defects*. 3rd ed. Harvard University Press; 1968:417-437.
3. Boraxbekk C-J, Lundquist A, Nordin A, Nyberg L, Nilsson L-G, Adolfsson R. Free recall episodic memory performance predicts dementia ten years prior to clinical diagnosis: Findings from the Betula Longitudinal Study. *Dement Geriatr Cogn Dis Extra*. 2015;5(2):191-202. doi:10.1159/000381535
4. Ding X, Charnigo RJ, Schmitt FA, Kryscio RJ, Abner EL. Evaluating trajectories of episodic memory in normal cognition and mild cognitive impairment: Results from ADNI. *PLoS One*. 2019;14(2):e0212435. doi:10.1371/journal.pone.0212435

5. Chen J, Chen G, Shu H, et al. Predicting progression from mild cognitive impairment to Alzheimer's disease on an individual subject basis by applying the CARE index across different independent cohorts. *Aging (Albany NY)*. 2019;11(8):2185-2201. doi:10.18632/aging.101883

6. Estévez-González A, Kulisevsky J, Boltes A, Otermín P, García-Sánchez C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *Int J Geriatr Psychiatry*. 2003;18(11):1021-1028. doi:10.1002/gps.1010

7. Landau SM, Harvey D, Madison CM, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*. 2010;75(3):230-238. doi:10.1212/WNL.0b013e3181e8e8b8

8. Chen P, DeKosky ST, Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*. 2000;55(12):1847-1853. doi:10.1212/wnl.55.12.1847

9. Lezak MD. Memory I: Tests. In: *Neuropsychological Assessment*. Oxford University Press; 1995:429-489.

10. Crawford JR, Stewart LE, Moore JW. Demonstration of savings on the AVLT and development of a parallel form. *J Clin Exp Neuropsychol*. 1989;11(6):975-981. doi:10.1080/01688638908400950

11. Helmstaedter C, Lendt M, Lux S. *VLMT - Verbaler Lern- Und Merkfähigkeitstest*. Beltz Test GmbH; 2001.

12. Chen JJ. Functional MRI of brain physiology in aging and neurodegenerative diseases. *Neuroimage*. 2019;187:209-225. doi:10.1016/j.neuroimage.2018.05.050

13. Huijbers W, Van Dijk KRA, Boenniger MM, Stirnberg R, Breteler MMB. Less head motion during MRI under task than resting-state conditions. *Neuroimage*. 2017;147:111-120. doi:10.1016/j.neuroimage.2016.12.002

14. den Haan J, Verbraak FD, Visser PJ, Bouwman FH. Retinal thickness in Alzheimer's disease: A systematic review and meta-analysis. *Alzheimer's Dement Diagnosis, Assess Dis Monit*. 2017;6:162-170. doi:10.1016/j.dadm.2016.12.014

15. Ko F, Muthy ZA, Gallacher J, et al. Association of retinal nerve fiber layer thinning with current and future cognitive decline: A study using optical coherence tomography. *JAMA Neurol*. 2018;75(10):1198-1205. doi:10.1001/jamaneurol.2018.1578

16. Mutlu U, Colijn JM, Ikram MA, et al. Association of retinal neurodegeneration on optical coherence tomography with dementia: A population-based study. *JAMA Neurol*. 2018;75(10):1256-1263. doi:10.1001/jamaneurol.2018.1563

17. Dong H, Csernansky JG. Effects of stress and stress hormones on amyloid-β protein and plaque deposition. *J Alzheimer's Dis*. 2009;18(2):459-469. doi:10.3233/JAD-2009-1152

18. Justice NJ. The relationship between stress and Alzheimer's disease. *Neurobiol Stress*. 2018;8:127-133. doi:10.1016/j.ynstr.2018.04.002

19. Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci*. 2009;10(6):434-445. doi:10.1038/nrn2639

20. Epel ES, Crosswell AD, Mayer SE, et al. More than a feeling: A unified view of stress measurement for population science. *Front Neuroendocrinol*. 2018;49:146-169. doi:10.1016/j.yfrne.2018.03.001

21. Aggarwal NT, Wilson RS, Beck TL, et al. Perceived stress and change in cognitive function among adults 65 years and older. *Psychosom Med*. 2014;76(1):80-85. doi:10.1097/PSY.0000000000000016

22. Korten NCM, Comijs HC, Penninx BWJH, Deeg DJH. Perceived stress and cognitive function in older adults: Which aspect of perceived stress is important? *Int J Geriatr Psychiatry*. 2017;32(4):439-445. doi:10.1002/gps.4486

23. McLennan SN, Ihle A, Steudte-Schmiedgen S, Kirschbaum C, Kliegel M. Hair cortisol and cognitive performance in working age adults. *Psychoneuroendocrinology*. 2016;67:100-103. doi:10.1016/j.psyneuen.2016.01.029

3 Memory and cognitive assessments for population-based studies

3.1 Ten German versions of Rey's Auditory Verbal Learning Test: Age and sex effects in 4,000 adults of the Rhineland Study

Meta M. Boenniger,[1] Christian Staerk,[2] Annabell Coors,[1] Willem Huijbers,[1] Ulrich Ettinger,[3] and Monique M. B. Breteler[1,2]

[1]Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[2]Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany

[3]Department of Psychology, University of Bonn, Bonn, Germany

Journal: Journal of Clinical and Experimental Neuropsychology 2021 (accepted)

ABSTRACT

Introduction: Detecting early pathological cognitive decline is critical for dementia and aging-related research and clinical diagnostics. Rey's Auditory Verbal Learning Test (AVLT) is commonly used to measure episodic verbal memory. The test requires participants to learn a list of 15 words over several trials. Since multiple testing is often required to detect cognitive decline, but repeating the same test can bias results, we developed 10 German AVLT word lists.

Method: We randomly assigned the lists to 4,000 participants (aged 30-94 years) from a population-based cohort to test their comparability, as well as aging effects and sex differences.

Results: Nine lists were highly comparable, with only one being slightly more difficult. Recall performance decreased on average by 0.6-1.1 words per trial per decade of age. Perseveration errors decreased with increasing age. Women remembered on average between 0.8-1.5 words per trial more than men, regardless of age. Women also outperformed men in the sum of trials 1 to 5, learning over trials, retroactive inhibition, and false positive and interference errors. Proactive inhibition remained stable across age and was unaffected by sex.

Conclusion: This German AVLT version presents comparable lists including detailed age and sex references and therefore allows test repetition excluding training effects. These versions are a valuable resource for research and clinical application.

INTRODUCTION

Neurodegenerative diseases and cognitive aging are among the main challenges in society, as they result in a decline in a multitude of cognitive functions, such as episodic memory. Neuropsychological assessments are the main method to assess and monitor cognitive aging, and tests such as the Rey Auditory-Verbal Learning Test (AVLT)[1,2] can detect pathological changes before disease onset.[3,4]

The AVLT is one of the most commonly used research and clinical instruments to examine declarative episodic verbal memory and learning. It is a serial learning task in which 15 semantically unrelated nouns are learned and recalled over multiple trials including an interference trial, a time delay trial and sometimes a recognition trial. The basic test idea and implementation goes back to Claparède who described a one-trial 15 word list recall at the beginning of the 20th century.[5] Rey later introduced a modification of Claparède's task including five trials and a recognition trial.[1] The first German version of the AVLT called Verbaler Lern- und Merkfähigkeitstest (VLMT) was published in 1990,[6] based on the original AVLT described by Rey and Taylor.[1,2] Helmstaedter and Durwen retained the general testing procedure and translated the word lists from the French and English versions, including some linguistic modifications. In 2001, the latest and commonly used version of the German test version was published.[7] Additional norm studies for this test version have since been published with sample sizes of n = 407 and n = 92 on elderly participants and including sex-adjustments.[8,9]

Due to its sensitivity for detecting early cognitive decline, its cost efficiency and the simplicity of the design, the AVLT has been widely used in children, adolescents and elderly participants in many languages.[e.g.9–17] It has also been applied in numerous clinical samples, including temporal lobe epilepsy, depression, borderline personality disorder, schizophrenia, suspected dementia, and dementia.[7,8,18–20] In participants with either subjective memory complaints or mild cognitive impairment, the AVLT predicted significantly, and more than other neuropsychological tests, whether participants progressed to Alzheimer's disease within the next two to three years.[3,21]

AVLT recall performance declines with age, especially for delayed recall.[10,11,22] However, the reported rate of age-related decline on the AVLT differs between test versions,

countries and languages. Therefore, it is important to examine aging effects in large samples covering a wide age range, especially for new test versions.[23]

Sex differences in the AVLT performance have also been observed, with women on average outperforming men.[9,11,24,25] However, some studies did not find sex differences.[7,26] One reason for these inconsistencies might be that samples differed in social status indicators. For example, Asperholm, Nagar, Dekhtyar, and Herlitz found that verbal episodic memory performance benefits from higher education and employment status in women more than in men.[27] Thus, educational level has to be considered when examining sex differences.

An important issue in research and clinical practice concerns the availability of parallel test versions. These are required to enable re-examinations of the same participant or patient cohort[28–30] and to lower the risk that participants exchange information about study materials. Two additional parallel German lists were published within the test version by Helmstaedter.[7] However, to our knowledge, the comparability of those list versions was investigated only in a small clinical sample and those parallel versions are not often reported to be used.[7] Therefore, there is a clear need for further development and thorough testing of additional versions with equal difficulty for the German AVLT.

This study presents a German version of the AVLT that contains 10 newly developed test versions that were tested for equal difficulty in a large population-based sample with a wide age range (30 to 94 years). The ten versions each have their own learning list, but the interference list has been kept the same for all versions so as not to introduce additional differences between versions due to the difficulty of this list. We examined the effects of age and sex in depth, and also considered the influence of education. Normative data are provided in the supplementary material for a large number of outcomes, both across all list versions and separately for each list version. Comparability of our scores with other reference studies[7,9,11] is discussed.

MATERIALS AND METHODS

Instrument development and testing procedure

We developed 10 parallel word lists of 15 words each in three steps, i.e. word collection, category attribution including word frequency assignment, and list definition. Word collection was done on the basis of previous publications that reported different versions of word learning tests similar or equivalent to the AVLT. We used already published words to achieve similar difficulty and to include words that have been shown to be memorable. We identified 11 sources which reported detailed information on their testing procedure and word stimuli.[1,6,7,26,28,29,31–35] For one article we contacted the authors for the word lists.[31] We translated the words into German, as the source material was mostly English, one source Spanish, one Dutch with English translation, and one Persian with English translation.

In the second step, two scientists categorized words into semantic groups (animals, body parts, buildings, clothing, food, furniture, musical instruments, objects, outdoor, plants, profession, relatives, sports, and vehicles), plus a miscellaneous category. The latter category contained words that could not be translated into meaningful German words, had double meaning in German or that did not fit into one of the other categories. Similar to Crawford et al.,[29] we assigned each word a frequency using the German film and television subtitle-based word frequency (SUBTLEX-DE).[36] For categories that contained fewer than 10 words, we chose additional ones with a similar frequency as the other words in that category to have one word per list from that category. Word frequency was considered to provide information on the familiarity of each word for participants, to ensure that all words were well-known and not so rare that participants could memorise them exceptionally well.

In the third step, one word from each category was included in each list, with exception of the large and heterogeneous category “outdoor” from which we selected two words. If categories contained more words than needed for our ten list versions, we selected words that were already used in other German lists to maximize comparability in learning difficulty with established versions. Next, we selected additional words based on high SUBTLEX-DE frequency. We also chose not to use words with a lot of syllables in the lists. Moreover, we asked native German speakers across different age groups to check

all lists for word combinations that could simplify memorization processes, for example combination of words that are linked to well-known phrases, songs or nursery rhymes. When such combinations were found, words were exchanged across lists within the same category. In a last step we randomized the order of words within each list.

The interference list was adopted from previous reports.[2,6,9] We used the same interference list for all 10 list versions, similar to Helmstaedter et al..[7] As mentioned in the introduction we did not expect a single recall to be remembered longer. We also aimed to keep the difficulty of this list as similar as possible to produce a uniformly strong interference.

Participants

Cross-sectional data were obtained from 4,000 participants of the Rhineland Study, a single-center population-based prospective cohort study of community dwelling individuals from two geographically defined areas in Bonn (Germany), aged 30 to 94 years. The study aims to investigate the etiology and prediction of neurodegenerative and other age-related diseases, and emphasizes deep phenotyping. The study is approved by the ethics committee of the Faculty of Medicine of the University of Bonn and is carried out in accordance with the recommendations of the International Council for Harmonisation (ICH) *Good Clinical Practice* (*GCP*) standards (ICH-*GCP*). All participants provided written informed consent in accordance with the Declaration of Helsinki. The present analysis contained all participants assessed between March 2016 and August 2019. Participants received no financial incentives for their participation.

Of the initial 4,000 participants, 43 persons had missing or incomplete AVLT data. Reasons for missing/incomplete data included data quality (n = 19), refusal to complete the test (n = 10), failed acquisition due to technical or organizational issues (n = 5), insufficient time (n = 3), and unknown (n = 2) or multiple reasons (n = 4). From the remaining 3,957 participants we excluded 93 participants who reported at least one medical diagnosis that could possibly affect AVLT performance, i.e. dementia, stroke including transient ischemic attacks or severe traumatic brain injury. We further excluded 34 participants with missing or incomplete data on education. Most of these missing cases were due to participants' refusal to provide this information or insufficient information to

clearly define educational level. Our final sample for analyses thus comprised 3,830 participants.

Procedure and measurement

The AVLT was carried out as the first part of a 50-minute cognitive test battery. For each participant, study assistants selected the list version from a predefined table in which each list version was represented equally often, but in randomized order.[37] In five successive trials, this list was read to the participant with about one second per word in a fixed order. Each of the trials 1 to 5 were immediately followed by free recall, for which the study assistant took notes on the computer. To avoid mistakes and to improve the writing speed, we applied a word completion aid for the words in the respective list. Study assistants were instructed to note down all words given as answers by the participants. Repetition of words within memorization strategies were not noted down. Subsequently, another, also semantically unrelated, word list (interference list) was read out and had to be immediately recalled. After this interference list, the participant was asked to recall the first list again without the study assistant repeating the words first (Trial 6). Then, 20-30 minutes later, a delayed recall (Trial 7) was carried out, asking the participant for the words learned in the first list, again without prior repetition of the list. Although the order in which the word lists were presented stayed the same, the participant was allowed to recall the words in any order.[1,2,7,28,35] We did not include a recognition trial. In the delay between Trial 6 and Trial 7, participants performed a fixed set of non-verbal cognitive tests, comprising oculomotor tasks,[38] a Corsi block-tapping test,[39] and a trail-making test.[40]

As one of the last tests in the cognitive battery, participants performed a German verbal intelligence Test (Mehrfachwahl-Wortschatz-Intelligenztest; MWT-B)[41] to allow an estimate of crystallized intelligence. In this test, participants were asked to identify a correct German word among four imaginary words in each of 37 items. We included MWT-B performance in the participant characteristics to provide a comprehensive description of our sample.

Information on educational status and medical history was obtained using structured interviews and self-administered questionnaires. Education was defined as the highest self-reported educational attainment. Based on the International Standard Classification

of Education (ISCED) 2011, we classified *low education* as the completion of lower and secondary education or below, *middle education* as completed upper secondary education up to completed Bachelor's degree or equivalent and *high education* as completed Master's degree or equivalent up to completed doctoral degree or equivalent. All examinations were carried out or supervised by specifically trained and certified study assistants to ensure data quality.

Definition of AVLT scores

Definitions and calculations of AVLT variables are listed in Table 1. Primary AVLT variables were the number of correctly recalled words for each trial (*Trial 1*, *Trial 2*, *Trial 3*, *Trial 4*, *Trial 5*, *Interference*, *Trial 6*, *Trial 7*). As secondary variables, we selected commonly used summary scores which give valuable additional information on episodic verbal memory performance. The *sum of trials 1 to 5* is a frequently used score to evaluate learning performance. The *learning over trials* variable shows learning performance independent of recall in Trial 1. In both variables, sum of trials 1 to 5 and learning over trials, higher values represent better learning performance. The *proactive inhibition* variable shows the strength of the inhibition effect of a previous learning task on new word list learning. The *retroactive inhibition* variable shows the strength of inhibition effect of the interference list on the previously learned list. In both inhibition variables positive numbers show an inhibition effect on performance, whereas negative numbers occur if participants recalled more words in Trial 6 than Trial 5 (retroactive inhibition) or more words in the interference trial than in Trial 1 (proactive inhibition). To also include incorrectly recalled words in our analysis we calculated three error variables (*false positive errors*, *interference errors* and *perseveration errors*). As numbers per trial were usually very small we summed the values over all trials. Still, numbers remained very small especially for false positive and interference errors. To check whether results were affected by the total number of recalled words, we additionally calculated the *percentage of perseveration errors*, as this was the most frequent error.

Table 1. AVLT score definitions

AVLT score	Description
Trial 1	Number of correctly recalled words after first learning trial.
Trial 2	Number of correctly recalled words after second learning trial.
Trial 3	Number of correctly recalled words after third learning trial.
Trial 4	Number of correctly recalled words after fourth learning trial.
Trial 5	Number of correctly recalled words after fifth learning trial.
Interference	Number of correctly recalled words from the interference list after interference list learning trial.
Trial 6	Number of correctly recalled words from first list without learning.
Trial 7	Number of correctly recalled words from first list without learning after 20 to 30 minutes time delay to Trial 6.
Sum of trials 1 to 5	Sum of recalled words from Trial 1 to Trial 5.
Learning over trials	Sum of recalled words in Trial 1 to Trial 5 minus five times the number of recalled words in Trial 1.
Proactive inhibition	Difference between recalled words in Trial 1 and recalled words in the Interference trial.
Retroactive inhibition (loss after interference)	Difference between recalled words in Trial 5 and Trial 6.
False positive errors	Number of recalled words which had not been part in any of the two lists, summed over all trials.
Interference errors	Number of words that were part of one of the learned lists but recalled in the respective other list recall trial (e.g. “bird” was learned in the first list but it was given as an answer in the interference list recall)
Perseveration errors	Number of repetitions within one trial independent of whether they have been learned in any of the lists or not, summed over all trials.
Percent of perseveration error	Ratio of the perseveration error and the total number of recalled words over all trials.

Statistical analysis

To assess the comparability of the 10 list versions as well as effects of age and sex on the number of recalled words in the different trials (Trial 1 to Trial 7), we used linear mixed-effect models. We included a random intercept for participants in order to account for correlations between repeated measures of AVLT scores. We also added a random slope for Trials 6 and 7 as they are not part of the learning slope of the other trials. As main fixed effects we included: trial (Trial 1 to Trial 7), delay time (centered) for Trial 7, and education (low, middle, and high). Further, we included interaction terms between trial and list version (A to J), trial and age (centered), and trial and sex (men and women). In an additional model, we added a trial by age^2 interaction term to examine non-linear aging effects. These analyses did not include the Interference trial as this was not different between the list versions. As we used mixed-effect regression analyses we calculated Cohen's f^2 based on marginal R^2 estimates to report effect sizes for continuous and categorical fixed effects.[42–44] We chose to calculate Cohen's f^2 as it gives the proportion of variance in the outcome variable that is uniquely accounted for by the independent variable of interest in relation to the proportion of unexplained variance [$f^2 = (R^2_{AB} – R^2_A) / (1 – R^2_{AB})$; R^2_{AB}: proportion of variance accounted for by the variable of interest (B) and the set of other variables (A); R^2_A: proportion of variance accounted for by the set of other variables (A)].[42,45] Here, $f^2 = 0.02$ indicates a small effect, $f^2 = 0.15$ indicates a medium effect and $f^2 = 0.35$ indicates a strong effect.[42(pp410-414)]

To assess the comparability of the 10 list versions as well as effects of age and sex on error variables, for count data (false positive, interference, and perseveration errors) we used hurdle models [46] with a negative binomial distribution for the count part of the models. For percentage of perseveration error we used a zero-inflated beta regression model.[47] All models included list version, age (centered), sex, and education as independent variables.

Age and sex effects in sum of trials 1 to 5, learning over trials, proactive and retroactive inhibition were estimated using linear models. Sum of trials 1 to 5, learning over trials, and proactive or retroactive inhibition variables were derived using list version corrected trial performance. For the list version correction, we used the estimates of the interaction between trial and list version and subtracted them from the original trial performance.

Linear models included age (centered), sex, and education as independent variables. Also for the linear models we calculated Cohen's f^2 to report the effect sizes.[42,43]

All models including individual list versions (linear mixed effect, hurdle and zero-inflated beta regression models) used list version E as the reference category, as it was closest to the sample median. The reference category for the variable sex was women and for education the reference category was the middle educational level.

All analyses were done using R version 3.6.2 and RStudio version 1.2.5033.[48,49] Main functions used for the analyses were the lme function from the nlme package[50] for the linear mixed effects models, the hurdle function from the pscl package[46] for the hurdle models, and the gamlss function from the gamlss package[47] for the zero-inflated beta regression model.

RESULTS

Sample characteristics

Our sample of analysis (n = 3,830) was younger than the excluded sample (n = 170, η^2 = 0.02, small effect) but not considerably different in the proportion of men and women (Cohen's w = 0.03). Included and excluded participants showed no difference in MWT-B performance (η^2 = 0.00). The exclusion of participants did not have a strong influence on the distribution of age (analysis sample: median [IQR] = 54.5 [44.5, 65.7]; complete sample: median [IQR] = 55.0 [45.0, 66.4]), sex (analysis sample: women n = 2,180 (57 %), men n = 1,650 (43 %); complete sample: women n = 2,264 (57 %), men n = 1,736 (43 %)), educational level (analysis sample: low n = 76 (2 %), middle n = 1,710 (45 %), high n = 2,044 (53 %); complete sample: low n = 82 (2 %), middle n = 1,776 (45 %), high n = 2,101 (53 %)) and MWT-B performance (analysis sample: median [IQR] = 31.0 [29.0, 33.0]; complete sample: median [IQR] = 31.0 [29.0, 33.0]) between our total sample and the sample of analysis. Table 2 gives participant characteristics for men and women separately. Men and women differed in educational level with men being more often in the high and women being more often in the middle category.

Table 2. Participant characteristics

Characteristic		Women (n = 2,180)	Men (n = 1,650)	Effect size	*p*
Age, median years [IQR]		54.7 [44.7, 65.2]	54.3 [43.9, 66.0]	0.00†	.874
Age range, years		30-94	30-93		
Age, n (%)	30-39 years	361 (17)	311 (19)	0.05‡	.230
	40-49 years	439 (20)	305 (18)		
	50-59 years	602 (28)	422 (26)		
	60-64 years	223 (10)	157 (10)		
	65-69 years	210 (10)	157 (10)		
	70-74 years	140 (6)	117 (7)		
	75-79 years	127 (6)	107 (6)		
	80+ years	78 (4)	74 (4)		
Education, n (%)	low	58 (3)	18 (1)	0.13‡	<.001
	middle	1081 (50)	629 (38)		
	high	1041 (48)	1003 (61)		
MWT-B, median [IQR][a]		31 [29, 33]	31 [29, 33]	0.00†	.020
Time delay, mean minutes (SD)		22.3 (3.9)	22.3 (4.0)	0.00†	.636

Note. SD, standard deviation; IQR, Interquartile range; ISCED, International Standard Classification of Education; MWT-B, Multiple-choice Vocabulary Intelligence Test; Low education: ISCED-11: 0-2; Middle education: ISCED-11: 3-6; High education: ISCED-11:7-8. Skewed variables are represented by median and IQR. Symmetrically distributed variables are represented by mean and SD. *P*-values were derived for categorical variables by chi-squared tests, for symmetrically distributed continuous variables by a two-group ANOVA, and for skewed continuous variables by a Kruskal-Wallis test.

[a]Missing values in women: n = 172. Missing values in men: n = 121.

†η^2-value. η^2 = 0.01 indicates a small effect, η^2 = 0.06 indicates a medium effect and η^2 = 0.14 indicates a large effect)

‡Cohen's *w*. *w* = 0.10 indicates a small effect, *w* = 0.30 indicates a medium effect and *w* = 0.50 indicates a large effect.[42(chap7.2.3)]

Sample characteristics were comparable across all AVLT list versions (Table 3).

Table 3. Participant characteristics stratified for list version

		List version									
		A	B	C	D	E	F	G	H	I	J
n		370	396	371	370	417	365	374	396	380	391
Age, median years [IQR]		54.5 [45.6, 66.4]	54.1 [43.4, 65.2]	54.4 [45.9, 64.2]	53.9 [45.2, 64.5]	54.8 [43.8, 66.1]	57 [46.5, 66.4]	54.1 [43.6, 63.2]	54.3 [43.1, 65.5]	55.1 [47.8, 68.7]	53.9 [42.6, 65.8]
Sex, n (%)	women	202 (55)	228 (58)	220 (59)	211 (57)	232 (56)	219 (60)	206 (55)	226 (57)	231 (61)	205 (52)
	men	168 (45)	168 (42)	151 (41)	159 (43)	185 (44)	146 (40)	168 (45)	170 (43)	149 (39)	186 (48)
Education, n (%)	low	6 (2)	5 (1)	5 (1)	7 (2)	8 (2)	4 (1)	9 (2)	7 (2)	15 (4)	10 (3)
	middle	160 (43)	178 (45)	173 (47)	173 (47)	184 (44)	160 (44)	166 (44)	190 (48)	165 (43)	161 (41)
	high	204 (55)	213 (54)	193 (52)	190 (51)	225 (54)	201 (55)	199 (53)	199 (50)	200 (53)	220 (56)
MWT-B, median [IQR][a]		31 [29, 33]	31 [28, 33]	31 [28, 33]	31 [29, 33]	31 [29, 33]	31 [29, 33]	31 [29, 33]	31 [29, 33]	31 [29, 33]	31 [29, 33]
Time delay, mean minutes (SD)		22.5 (4.1)	22.3 (3.9)	22.0 (3.6)	22.2 (4.1)	22.4 (4.2)	22.1 (4.2)	22.2 (3.7)	22.1 (3.8)	22.7 (3.9)	22.2 (3.9)

Note. SD, standard deviation; IQR, Interquartile range; ISCED, International Standard Classification of Education; MWT-B, Multiple-choice Vocabulary Intelligence Test. Low education: ISCED-11: 0-2; Middle education: ISCED-11: 3-6; High education: ISCED-11:7-8.
[a]Missing values: n = 293.

AVLT performance

Figure 1 shows the performance of participants in Trials 1 to 7 and the number of errors stratified for list version. On average participants recalled 6.95 (SD = 2.01) words in Trial 1, 9.72 (SD = 2.38) words in Trial 2, 10.98 (SD = 2.42) words in Trial 3, 11.67 (SD = 2.31) words in Trial 4, 12.05 (SD = 2.23) words in Trial 5, 10.30 (SD = 3.08) words in Trial 6, and 10.32 (SD = 3.24) words in Trial 7. Over all list versions the median of false positive errors was 1 (range: 1-33), the median of interference errors was 0 (range: 0-9), and the median of perseveration errors was 3 (range: 0-59). Supplementary Tables S1 to S11 show participant performance in all AVLT outcome variables stratified by age in decades and sex over all list versions and stratified for each list version.

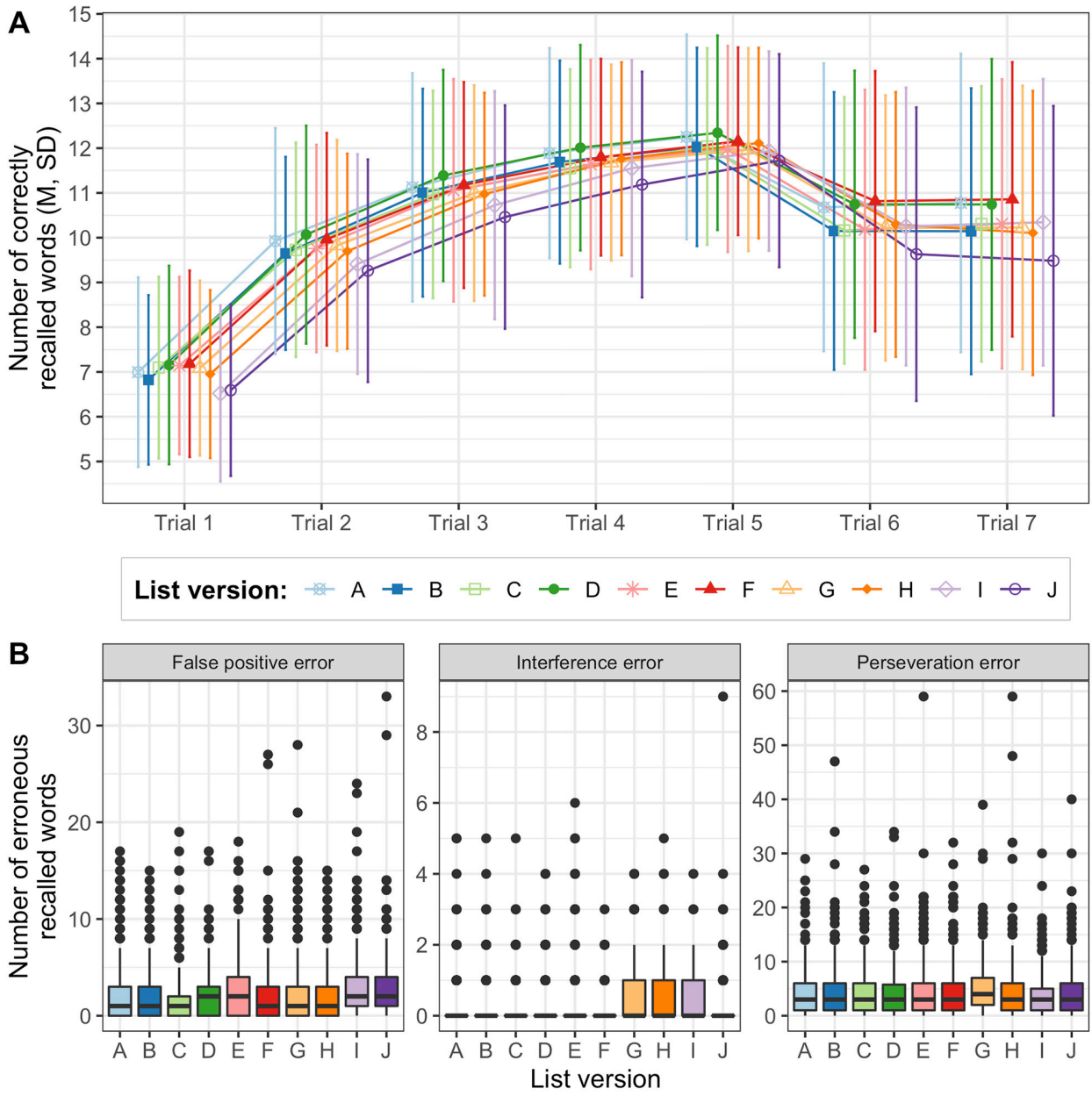

Figure 1. Correctly and erroneously recalled words across 10 list versions. (A) Number of correctly recalled words per trial (mean and standard deviation) stratified for list version, (B) number of erroneously recalled words summarized over all trials stratified for list version. *Abbreviations*: M, Mean; SD, standard deviation.

List version differences

On average our word lists had 25.7 syllables, with a maximum 27 syllables and a minimum 23 syllables. No word had more than three syllables.

Linear mixed models showed, after adjustment for possible confounders, mean list version differences of less than 0.82 words per trial (standardized estimate = 0.27) between reference list version E and the other list versions. List versions A to I showed high comparability. Only list version J was slightly more difficult than the other list versions over all trials (Figure 2 A). However, the overall effect of the list version on the model was very small (f^2 = 0.015).

For women at the mean sample age of 54.5 years, with a middle educational level and performing the reference list version E, the predicted probability of having at least one false positive error was 74 % (95 % CI [69 %, 78 %]). For the same participants, the predicted probability for an interference error was 23 % (95 % CI [19 %, 28 %]) and 86 % (95 % CI [82 %, 89 %]) for a perseveration error. All other list versions showed that the adjusted odds ratio (OR) of making at least one false positive error compared to list version E ranged between 0.619 and 1.208, with list versions A and C showing significantly lower odds ($p < .05$) (Figure 2 B). For interference errors, the adjusted OR of making at least one error compared to list version E ranged between 0.263 and 1.227 with list versions A and F showing significantly lower odds ($p < .05$) (Figure 2 B). For perseveration errors, the adjusted OR of making at least one error compared to list version E ranged between 0.875 and 1.325 with none of the lists showing significantly lower or higher odds ($p > .05$) (Figure 2 B).

The count part of the hurdle model yielded, for participants who made errors, the multiplicative effects on the rate ratio (RR) for the list versions different from version E. For false positive errors, it ranged between 0.652 and 0.990, for the interference error it ranged between 0.360 and 0.912, and for perseveration errors it ranged between 0.867 and 1.092 (Figure 2 C). Significant differences ($p < .05$) in the RR with reference to list version E were observed for false positive errors for list versions B, C, D, G, and H and for interference errors for list versions C, F, H, I, and J. In perseveration errors, we did not find any differences in RRs between list versions (Figure 2 C).

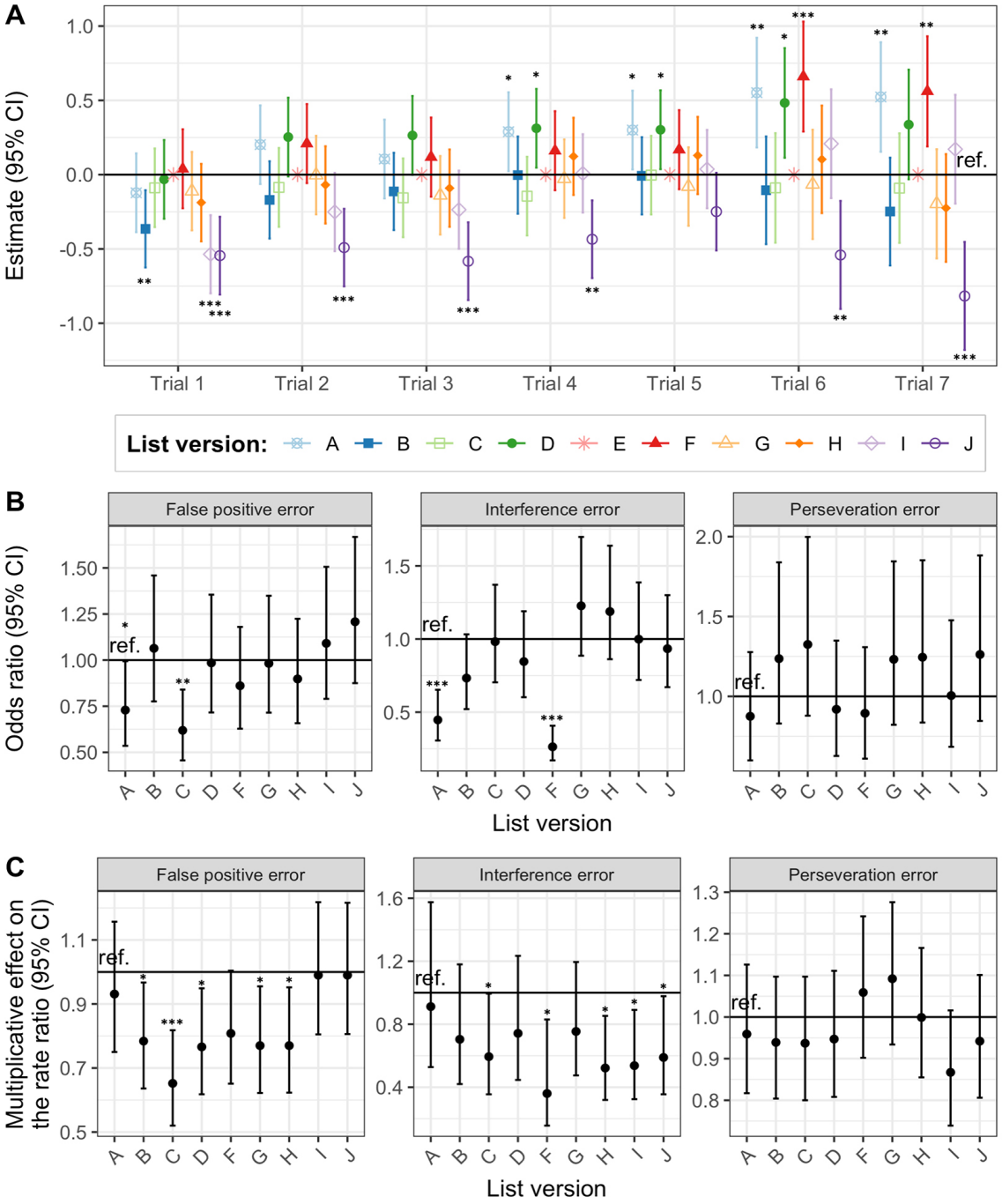

Figure 2. List Version Effects for Number of Correctly and Erroneously Recalled Words. (A) Additive effects of list versions on the number of recalled words in Trial 1 to 7 with reference to list version E, adjusted for age, sex, education, delay time, and age^2. (B) Adjusted odds ratios of making at least one error in the different list versions with reference to list version E (ref.), adjusted for age, sex, and education. Adjusted odds ratios are given by the exponential value of the coefficient estimates of the logistic regression part of the hurdle model. (C) Multiplicative effects on the rate ratio for all participants making at least one error, adjusted for age, sex, and education. Multiplicative effects are given by the exponential value of the coefficient estimates of the negative-binomial part of the hurdle model. *Abbreviations*: * $p < .05$, ** $p < .01$, *** $p < .001$.

Age effects

We observed medium to large age effects on the number of correctly recalled words across trials (f^2 = 0.289) (Figure 3 A). Performance dropped with each decade of age between 0.57 (95 % CI [0.53, 0.61]) words for Trial 1 and 1.07 (95 % CI [1.00, 1.13]) words for Trial 7 (Figure 3 A and Table 4). In addition, we also observed small nonlinear effects of age on the number of correctly recalled words across trials (Table 4).

Sum of trials 1 to 5 decreased on average with increasing age by 3.40 (95 % CI [3.22, 3.58]; standardized estimate = 0.34, 95 % CI [0.32, 0.36]; f^2 = 0.349) words per decade (Figure 3 B). Learning over trials decreased on average with increasing age by 0.52 (95 % CI [0.38, 0.67]; standardized estimate=0.08, 95 % CI [0.06, 0.10]; f^2 = 0.013) words per decade (Figure 3 B). The effect on retroactive inhibition increased on average with increasing age by 0.36 (95 % CI [0.32, 0.40]; standardized estimate = 0.19, 95 % CI [0.17, 0.21]; f^2 = 0.076) words per decade of age (Figure 3 B). No age effects were found on proactive inhibition (estimate = 0.00, 95 % CI [-0.05, 0.04]; standardized estimate = 0.00, 95 % CI [-0.02, 0.02]; f^2 = 0.000) (Figure 3 B).

The odds of making at least one error increased per year of age for false positive as well as interference errors with an adjusted OR of 1.018 (95 % CI [1.012, 1.023]) and 1.035 (95 % CI [1.029, 1.041]), respectively. In contrast, perseveration errors decreased per year of age with an adjusted OR of 0.982 (95 % CI [0.975, 0.988]). Similarly, within those participants making errors (number of participants making at least one error: $n_{\text{false positive errors}}$ = 2,750; $n_{\text{interference errors}}$ = 840; $n_{\text{perseveration errors}}$ = 3,271) with each increasing year of age the RR of errors was 1.014 (95 % CI [1.011, 1.018]) for false positive errors and 1.030 (95 % CI [1.020, 1.040]) for interference errors, but 0.993 (95 % CI [0.990, 0.995]) for perseveration errors. The decrease in the odds of making at least one perseveration error with increasing age remained also after taking the total number of recalled words into account (OR = 0.981, 95 % CI [0.975, 0.988]). However, the decrease in RRs with increasing age for participants making at least one perseveration error was not found for the percentage of perseveration errors.

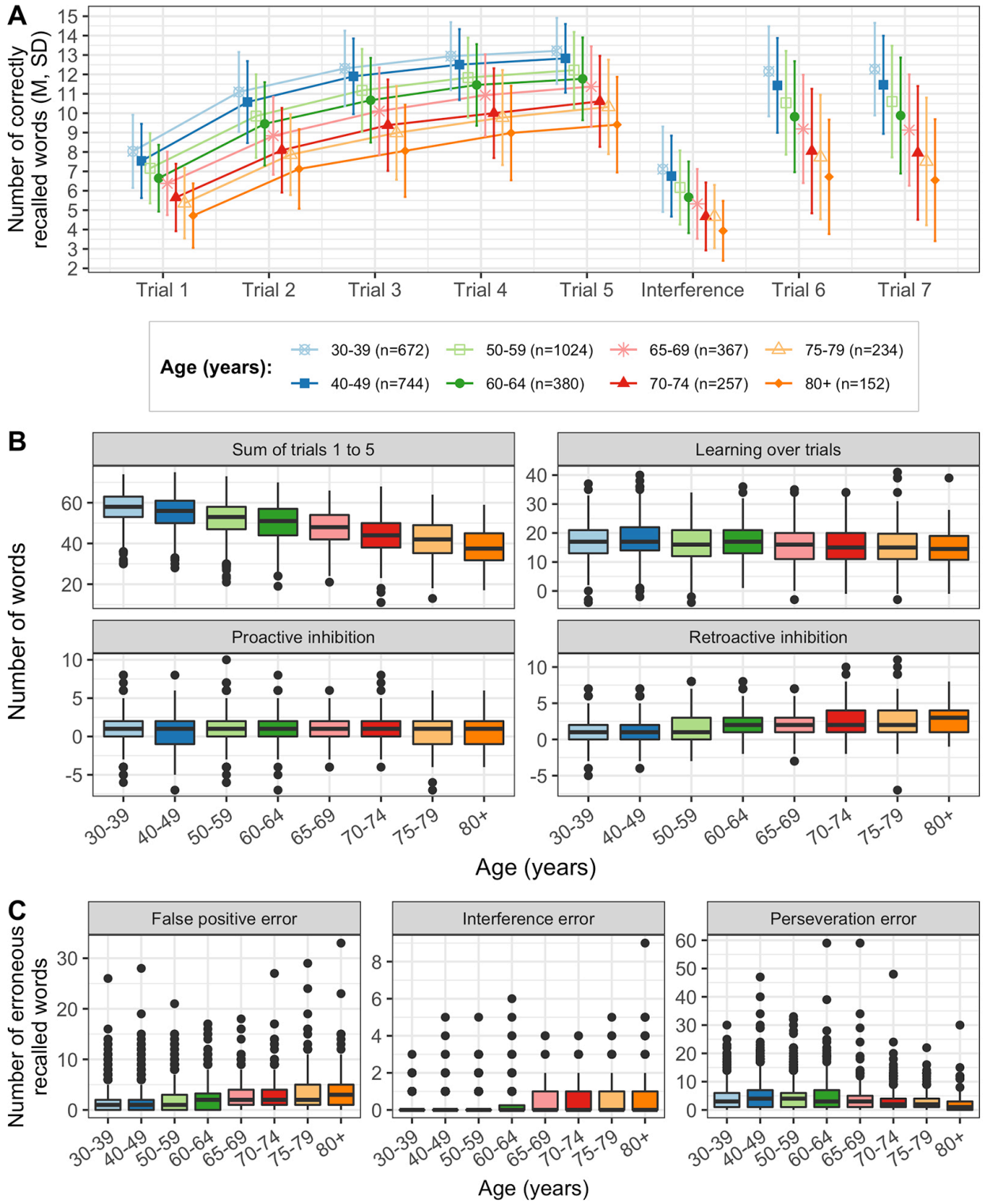

Figure 3. Age Effects on AVLT Outcome Variables. (A) Number of correctly recalled words separated for each age group in decades. (B) Performance in sum of trials 1 to 5, learning over trials, proactive and retroactive inhibition separated for each age group in decades. (C) Number of erroneous recalled words separated for each age group in decades. *Abbreviations*: M, Mean; SD, standard deviation.

Table 4. Effects of Age and Age^2 on the Number of Correctly Recalled Words per Trial

	Model 1			Model 2		
Fixed effects	Estimate (95 % CI)	Std. estimate (95 % CI)	*p*	Estimate (95 % CI)	Std. estimate (95 % CI)	*p*
Trial 1:Age	-0.057 (-0.061, -0.053)	-0.019 (-0.021, -0.018)	<.001	-0.057 (-0.061, -0.052)	-0.019 (-0.0.20, -0.017)	<.001
Trial 2:Age	-0.073 (-0.077, -0.068)	-0.024 (-0.026, -0.023)	<.001	-0.072 (-0.077, -0.068)	-0.024 (-0.026, -0.023	<.001
Trial 3:Age	-0.075 (-0.080, -0.071)	-0.025 (-0.027, -0.024)	<.001	-0.075 (-0.079, -0.070)	-0.025 (-0.026, -0.023)	<.001
Trial 4:Age	-0.071 (-0.075, -0.066)	-0.024 (-0.025, -0.022)	<.001	-0.070 (-0.074, -0.066)	-0.023 (-0.025, -0.022)	<.001
Trial 5:Age	-0.065 (-0.070, -0.061)	-0.022 (-0.023, -0.020)	<.001	-0.065 (-.069, -.060)	-0.022 (-0.023, -0.020)	<.001
Trial 6:Age	-0.102 (-0.108, -0.096)	-0.034 (-0.036, -0.032)	<.001	-0.101 (-0.107, -0.095)	-0.034 (-0.036, -0.032)	<.001
Trial 7:Age	-0.107 (-0.113, -0.100)	-0.036 (-0.038, -0.034)	<.001	-0.106 (-0.112, -0.099)	-0.035 (-0.037, -0.033)	<.001
Trial 1:Age^2				-0.001 (-0.001, 0.000)	0.000 (0.000, 0.000)	<.001
Trial 2:Age^2				-0.001 (-0.001, 0.000)	0.000 (0.000, 0.000)	<.001
Trial 3:Age^2				-0.001 (-0.001, -0.001)	0.000 (0.000, 0.000)	<.001
Trial 4:Age^2				-0.001 (-0.001, -0.001)	0.000 (0.000, 0.000)	<.001
Trial 5:Age^2				-0.001 (-0.001, -0.001)	0.000 (0.000, 0.000)	<.001
Trial 6:Age^2				-0.001 (-0.001, -0.001)	0.000 (0.000, 0.000)	<.001
Trial 7:Age^2				-0.001 (-0.002, -0.001)	0.000 (-0.001, 0.000)	<.001

Note. Std., standardized; CI, confidence interval. Excerpt from linear mixed effect models which are adjusted for the influence of list version, age (in years), sex, education, and delay time for trial 7. Model 1 includes an interaction between trials and age (in years), while Model 2 additionally includes an interaction between trials and age^2. Age reference: mean sample age.

Sex effects

Cohen's f^2 for the effect of sex on the number of correctly recalled words as outcome was 0.070. Women recalled on average about one word more per trial than men (Figure 4 A, Table 5). The inclusion of quadratic age effects only led to small changes in the estimates.

Women performed better than men in the sum of trials 1 to 5, recalling on average 5.333 more words (95 % CI [4.822, 5.845]; standardized estimate = 0.537, 95 % CI [0.485, 0.588]; f^2 = 0.109; Figure 4 B). Similarly, but to a lesser extent, women learned over trials on average about 1.263 more words than men (95 % CI [0.857, 1.670]; standardized estimate = 0.198, 95 % CI [0.134, 0.262]; f^2 = 0.010; Figure 4 B). Retroactive inhibition was stronger in men and resulted in forgetting on average 0.267 (95 % CI [0.151, 0.382]; standardized estimate = 0.143, 95 % CI [0.081, 0.205]; f^2 = 0.005) words more than women (Figure 4 B). We did not observe sex differences in proactive inhibition (estimate = -0.079, 95 % CI [-0.207, 0.049]; standardized estimate = -0.040, 95 % CI [-0.105, 0.025]; f^2 = 0.000) (Figure 4 B).

The odds of making at least one error were higher in men for false positive as well as for interference errors with adjusted ORs of 1.265 (95 % CI [1.093, 1.465]) and 1.285 (95 % CI [1.093, 1.511]), respectively. In contrast, the odds of having at least one perseveration error were not significantly different between men and women (OR = 0.837, 95 % CI [0.696, 1.006]). Within the group of participants making errors (number of participants making at least one error: $n_{\text{false positive errors}}$ = 2,750; $n_{\text{interference errors}}$ = 840; $n_{\text{perseveration errors}}$ = 3,271), on average men also made more false positive errors (multiplicative effect on RR = 1.226, 95 % CI [1.109, 1.356]). For interference errors we found no significant multiplicative effect on the RR (multiplicative effect on RR = 1.091, 95 % CI [0.836, 1.424]). For perseveration errors, within those participants making errors, women made more errors than men (multiplicative effect on RR = 0.879, 95 % CI [0.809, 0.937]). There was no effect of sex on the percentage of perseveration errors.

Visual inspection of trial performance and error count showed parallel trajectories of men and women across age. Therefore, we did not include an interaction term between age and sex in our analysis.

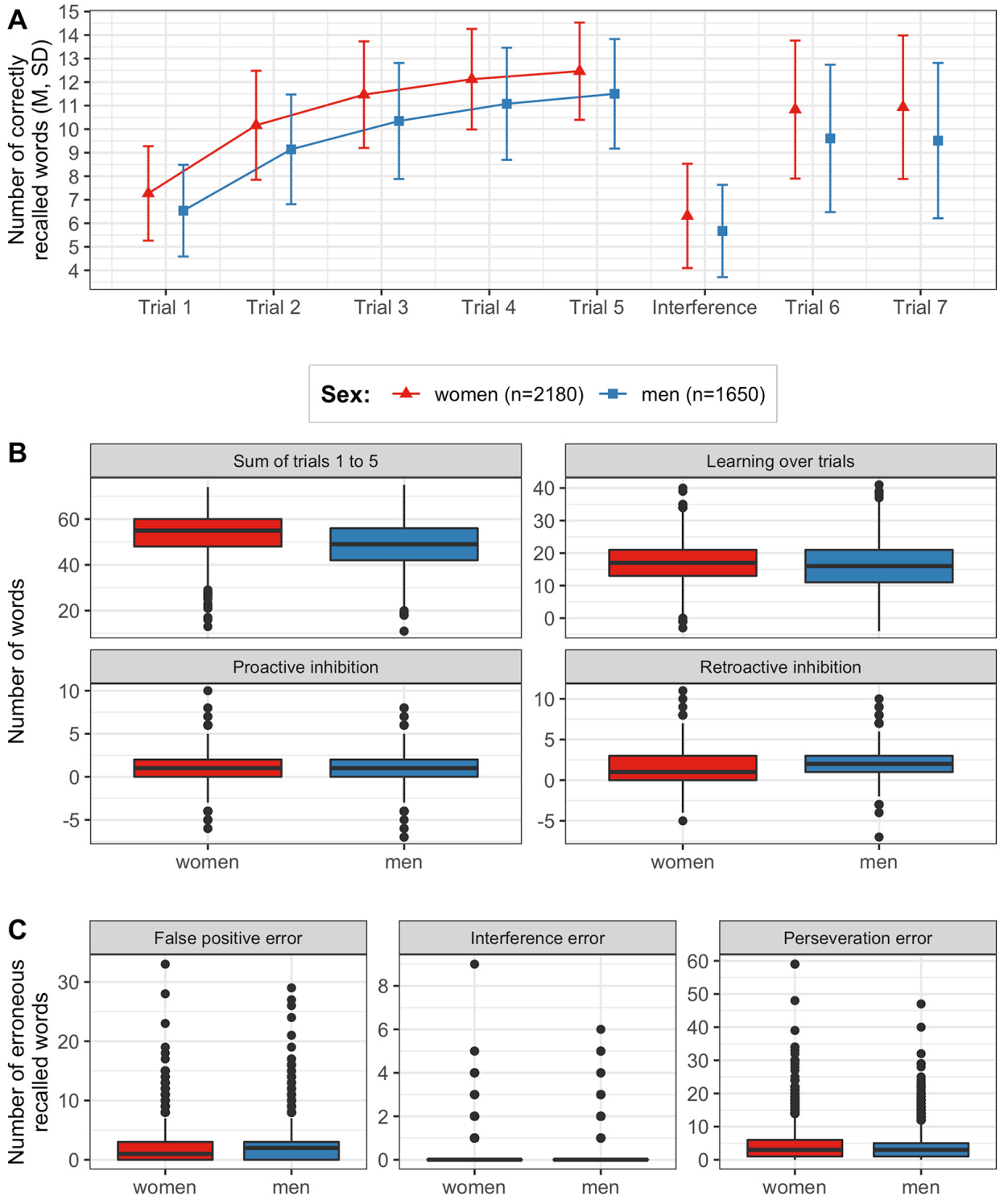

Figure 4. Sex Effects on AVLT Outcome Variables. (A) Number of correctly recalled words separated for men and women. (B) Performance in sum of trials 1 to 5, learning over trials, proactive and retroactive inhibition separated for men and women. (C) Number of erroneous recalled words separated for men and women. *Abbreviations*: M, Mean; SD, standard deviation.

Table 5. Effects of Sex on the Number of Correctly Recalled Words per Trial

	Model 1			Model 2		
Fixed effects	Estimate (95 % CI)	Std. estimate (95 % CI)	*p*	Estimate (95 % CI)	Std. estimate (95 % CI)	*p*
Trial 1:Sex	-0.822 (-0.945, -0.700)	-0.275 (-0.315, -0.234)	<.001	-0.807 (-0.929, -0.685)	-0.270 (-0.310, -0.229)	<.001
Trial 2:Sex	-1.107 (-1.230, -0.985)	-0.370 (-0.411, -0.329)	<.001	-1.093 (-1.215, -0.971)	-0.365 (-0.406, -0.324)	<.001
Trial 3:Sex	-1.200 (-1.322, -1.077)	-0.401 (-0.441, -0.360)	<.001	-1.179 (-1.301, -1.057)	-0.394 (-0.434, -0.353)	<.001
Trial 4:Sex	-1.127 (-1.249, -1.005)	-0.376 (-0.417, -0.335)	<.001	-1.107 (-1.229, -0.985)	-0.370 (-0.410, -0.329)	<.001
Trial 5:Sex	-1.045 (-1.168, -0.923)	-0.349 (-0.390, -0.308)	<.001	-1.026 (-1.148, -0.904)	-0.343 (-0.383, -0.302)	<.001
Trial 6:Sex	-1.293 (-1.463, -1.123)	-0.432 (-0.488, -0.375)	<.001	-1.272 (-1.442, -1.103)	-0.425 (-0.481, -0.368)	<.001
Trial 7:Sex	-1.482 (-1.651, -1.312)	-0.495 (-0.551, -0.438)	<.001	-1.458 (-1.628, -1.289)	-0.487 (-0.543, -0.430)	<.001

Note. Std., standardized; CI, confidence interval. Excerpt from linear mixed effect models which are adjusted for the influence of list version, age, sex, education, and delay time for trial 7. Model 1 includes an interaction between trials and age, while Model 2 additionally includes an interaction between trials and age^2. Sex reference: women.

DISCUSSION

We developed 10 German versions of the AVLT and evaluated their comparability in a large population-based sample, including participants between 30 and 94 years of age without dementia, stroke or severe traumatic brain injury diagnosis. Additionally, we estimated aging as well as sex effects in detail for a wide range of outcome variables and we provide detailed information on performance over age, stratified for men and women. We were able to include 95.75 % of the original sample in our analysis which reduces the likelihood of introducing bias due to data exclusions.

We chose a linear mixed-effects modelling approach, which allows for longitudinal modelling of recalled words over different trials by accounting for the correlation between repeated measurements.[51–53] An advantage of this approach was that the comparability of different list versions over multiple trials could be investigated in a joint statistical model, instead of considering each trial separately. In standard approaches of parallel testing this

is often not included.[54] Overall, results showed that our 10 list versions were highly comparable because list version had only a very small effect on recall performance (f^2 = 0.015), which corresponds to a maximum mean difference of less than one word from the reference list version. Only list version J proved to be slightly more difficult over all trials than the other list versions with the maximum mean difference of 0.82 words (standardized estimate = 0.27) for the delayed recall. Interference and false positive errors were infrequent and differed only slightly between list versions, whereas the more frequent perseveration errors showed no list differences. Therefore, we consider the significant differences in interference and false positive errors to be more a result of small variance and large sample size, then being indicative of clinically relevant differences between test versions.

Age showed to have a medium effect on recall performance (f^2 = 0.289). Recall performance decreased between 0.6 to 1 word (of 15 words in total) per decade of age, which is similar to findings of previous studies with comparable age ranges, but smaller sample sizes.[11,55,56] This supports the validity of our list versions in comparison to other test versions and shows the strength of effects. As a measure of learning ability, we chose both learning over trials and the frequently used sum of trials 1 to 5. In line with previous studies, the sum of trials 1 to 5 decreased strongly with higher age, as shown by a high effect size (f^2 = 0.349).[11,57,58] Learning over trials is not influenced by baseline performance (Trial 1) which makes it a better estimate of pure learning ability.[10] We found negative, but only very small age effects (f^2 = 0.013) of about half a word per decade for learning over trials, which has been also reported in some,[10,55] but not all studies.[13] These differences might be due to sample size differences. This suggests that while learning ability decreases slightly with higher age in non-demented, largely well-educated participants, the main effects of aging are noticeable at the first recall, which is thought to mainly reflect attentional ability.[58]

Adding the interference trial to the AVLT enabled us to also measure inhibition.[55] We observed small aging effects (f^2 = 0.076) on retroactive inhibition, but not proactive inhibition (f^2 = 0.000). Proactive inhibition measures the negative effect of previously learned material on new information. It has been shown to be affected in patients with frontal lobe lesions, but has not shown systematic age effects.[12] In contrast, retroactive

inhibition, which measures the adverse effect of new information on previously learned information, has been found to be increased in older participants, but results are inconsistent.[8,12] One reason for the inconsistency could be that retroactive inhibition is stronger in older age groups, which are not represented in most studies testing retroactive inhibition.[12]

Whilst error variables are often calculated, age effects are examined in depth in only very few studies. This might be due to their low frequency, because of which our results also need to be interpreted with precaution. Nevertheless, they are interesting to examine as very high values have been associated with pathological changes.[7,28] We found both false positive and interference errors to slightly increase with age whereas perseveration errors decreased. Previous studies found no increase in false positive and interference errors. Those studies however had much smaller sample sizes, and we likely had more power to detect the small effect sizes (OR between 0.982 and 1.035) given our very large sample size.[8,11] The decrease in perseveration errors with age is surprising, as this type of error, mostly examined in the Wisconsin Card Sorting Test, has been reported to increase with age.[59] However, for the AVLT, some previous studies also reported decreases with age.[8,11] These differences might be due to different demands that are measured by both types of perseveration errors.[60] To rule out the effect of the total number of words recalled, we corrected the perseveration error accordingly. This only marginally affected the estimates which remained significant. Whether and why perseveration errors would decrease with age warrants replication and further examination in future studies.

Sex differences in the number of recalled words per trial were not consistently found in previous studies. The often-used German test version by Helmstaedter[7] for example does not provide sex-specific norms. However, a number of studies across countries have reliably shown that, on average, women outperform men in most of the outcome variables.[9,11,24,25] We observed small to medium effects sizes (f^2 = 0.072) of sex on recall performance. Over all trials and age ranges, women recalled, on average, about one word (of 15 words in total) more than men, in accordance with previous studies.[25,61] We also observed small to medium effect sizes (f^2 = 0.109) for sex differences in learning performance using the sum of trials 1 to 5. However, similarly to age, this is largely driven by the sex differences in the first recall as learning over trials shows way smaller

(f^2 = 0.010), though still significant, sex differences. These have not previously been found by others.[55] However, in comparison to other studies, the sex differences in the number of correctly recalled words over trials reported here were stronger than in previous studies; accordingly, we were able to detect very small effects of sex in learning over trials. Additionally, as mentioned in the introduction, another reason for inconsistent findings regarding sex differences may be sample differences in socio-economic parameters such as education and employment status.[27]

The general level of difficulty of our list versions appears to be comparable to other studies,[56,62] although our participants performed slightly better than those in Van der Elst et al.[11] and Cavaco et al..[55] In relation to the most widely used German AVLT version, participants in our study, especially women, showed better performance over most outcomes than those in Helmstaedter et al..[7] However, our participants showed worse performance than participants in Speer et al.[9] who used the same test version as Helmstaedter et al..[7] This suggests that the observed differences are most likely due to educational and health status differences between samples, rather than to differences between AVLT versions.

LIMITATIONS

Limitations of our study include its cross-sectional design, as a consequence of which we could not estimate the test-retest reliability of our test versions. We did not introduce different interference lists, as we wanted to keep the influence of the interference list as similar as possible between the learning list versions. Due to testing time restrictions we also did not include a recognition trial. Our sample was highly educated and included only a very limited number of participants in ISCED levels one and two. Therefore, we grouped them together in a heterogeneous low education category. Our sample was not representative of the education levels in the German population.[63] Therefore, we could not analyse effects of education in detail and only corrected for educational level in our models for it. Our sample also included only a small number of very old participants (80+ years, n = 152). In terms of episodic memory, this can be a very heterogeneous group, even within a non-dementia sample. Further studies that specifically examine this age group would be helpful to support the generalizability of our findings to the very old. We

did not test these list versions for discriminative validity between impaired and healthy participants. However, since we kept the task design very comparable to previous task versions and we were able to show relatively comparable difficulty levels to previously published tasks, we expect a similar discriminative validity. Another limitation is that this paper focused on the presentation of comparable AVLT list versions and data was therefore not analysed at the item level. We would like to point out, however, that especially in the item analysis, the difficulty of a word most likely applies only to the German language, because when translated into another language the word length, the number of syllables and the word frequency may be different, which may affect memory performance.[64] Nevertheless, it would be interesting to analyse this in future studies, especially as an alternative analysis for error values. Using the different list versions within another sample would further help to strengthen the observed results. Finally, it is worth mentioning that the retroactive inhibition outcome measure, in other studies also called retroactive interference,[12,65] is not a pure inhibition effect, as Trial 6 includes no presentation of the words before the recall as Trial 5 does.

CONCLUSIONS

We created ten list versions for the AVLT, of which nine were shown to be highly comparable. These test versions could be used in longitudinal studies to avoid bias resulting from test repetition. Because of the simplicity of the test design and the translatability of this test into other languages, we encourage other scientists to use these versions and test them further also in different languages. Additionally, we estimated age and sex effects in a large sample and provide detailed reference values of AVLT performance stratified for age in decades and sex especially for participant groups with middle and high educational levels. Learning performance, inhibition effects, and errors should be examined in more depth in other large samples as those outcomes could provide additional valuable information about cognitive status.

ACKNOWLEDGEMENTS

We thank all participants of the Rhineland Study and the study personnel which were involved in the extensive data collection for this study. Special thanks go to Dr Natascha Merten and Dr David D. Ward for their helpful discussions during our “cognition meetings”. Finally, we would like to thank our IT and data management team, especially Christof Meigen and André Medek for their contribution on data processing and management.

DECLARATION OF INTEREST

The Rhineland Study at the DZNE is funded by the Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the German State of North Rhine-Westphalia through the German Center for Neurodegenerative Diseases (DZNE). We have no conflicts of interest to disclose.

DATA AND INSTRUMENT AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because of data protection regulations. Access to data can, however, be provided to scientists in accordance with the Rhineland Study’s Data Use and Access Policy. Test instrument information including the word list versions, are not publicly available to avoid pre-knowledge effects in participants who will continuously attend the study. Requests to access datasets or test instrument information should be directed to Dr Monique Breteler, RS-DUAC@dzne.de.

REFERENCES

1. Rey A. Mémorisation d’une série de 15 mots en 5 répetitions. In: Rey A, ed. *L’examen Clinique En Psychologie*. 3rd ed. Presses Universitaries de France; 1970:141-193.
2. Taylor EM. 8. Learning and Memory. In: Taylor EM, ed. *Psychological Appraisal of Children with Cerebral Defects*. 3rd ed. Harvard University Press; 1968:417-437.

3. Estévez-González A, Kulisevsky J, Boltes A, Otermín P, García-Sánchez C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *Int J Geriatr Psychiatry*. 2003;18(11):1021-1028. doi:10.1002/gps.1010

4. Boraxbekk C-J, Lundquist A, Nordin A, Nyberg L, Nilsson L-G, Adolfsson R. Free recall episodic memory performance predicts dementia ten years prior to clinical diagnosis: Findings from the Betula Longitudinal Study. *Dement Geriatr Cogn Dis Extra*. 2015;5(2):191-202. doi:10.1159/000381535

5. Boake C. Edouard Claparede and the auditory verbal learning test. *J Clin Exp Neuropsychol*. 2000;22(2):286-292. doi:10.1076/1380-3395(200004)22:2;1-1;FT286

6. Helmstaedter C, Durwen HF. VLMT: Verbaler Lern- und Merkfähigkeitstest - Ein praktikables und differenziertes Instrumentarium zur Prüfung der verbalen Gedächtnisleistungen. *Schweizer Arch für Neurol und Psychiatr*. 1990;Band 141(Heft 1):21-30.

7. Helmstaedter C, Lendt M, Lux S. *VLMT - Verbaler Lern- Und Merkfähigkeitstest*. Beltz Test GmbH; 2001.

8. Volz-Sidiropoulou E, Poll E, Forkmann T, Gauggel S. Erweiterte Altersnormen zum Verbalen Lern- und Merkfähigkeitstest (VLMT). *Klin Diagnostik und Eval*. 2010;3(August):226-243.

9. Speer P, Wersching H, Bruchmann S, et al. Age- and gender-adjusted normative data for the German version of Rey's Auditory Verbal Learning Test from healthy subjects aged between 50 and 70 years. *J Clin Exp Neuropsychol*. 2014;36(1):32-42. doi:10.1080/13803395.2013.863834

10. Ivnik RJ, Malec JF, Tangalos EG, Petersen RC, Kokmen E, Kurland LT. The Auditory-Verbal Learning Test (AVLT): Norms for ages 55 years and older. *Psychol Assess*. 1990;2(3):304-312. doi:10.1037/1040-3590.2.3.304

11. Van der Elst W, van Boxtel MPJ, van Breukelen GJP, 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*. 2005;11(3):290-302. doi:10.1017/S1355617705050344

12. Vakil E, Greenstein Y, Blachstein H. Normative data for composite scores for children and adults derived from the rey auditory verbal learning test. *Clin Neuropsychol*. 2010;24(4):662-677. doi:10.1080/13854040903493522

13. Bezdicek O, Stepankova H, Moták L, et al. Czech version of Rey Auditory Verbal Learning test: Normative data. *Aging, Neuropsychol Cogn*. 2014;21(6):693-721. doi:10.1080/13825585.2013.865699

14. Messinis L, Nasios G, Mougias A, et al. Age and education adjusted normative data and discriminative validity for Reys Auditory Verbal Learning Test in the elderly Greek population. *J Clin Exp Neuropsychol*. 2016;38(1):23-39. doi:10.1080/13803395.2015.1085496

15. O'Bryant SE, Edwards M, Johnson L, Hall J, Gamboa A, O'jile J. Texas Mexican American adult normative studies: Normative data for commonly used clinical neuropsychological measures for English- and Spanish-speakers. *Dev Neuropsychol*. 2018;43(1):1-26. doi:10.1080/87565641.2017.1401628

16. Vogel A, Stokholm J, Jørgensen K. Performances on Rey Auditory Verbal Learning Test and Rey Complex Figure Test in a healthy, elderly Danish sample - reference data and validity issues. *Scand J Psychol*. 2012;53(1):26-31. doi:10.1111/j.1467-9450.2011.00909.x

17. Lavoie M, Bherer L, Joubert S, et al. Normative data for the Rey Auditory Verbal Learning Test in the older French-Quebec population. *Clin Neuropsychol*. 2018;32(sup1):15-28. doi:10.1080/13854046.2018.1429670

18. Schoenberg MR, Dawson KA, Duff K, Patton D, Scott JG, Adams RL. Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch Clin Neuropsychol*. 2006;21(7):693-703. doi:10.1016/j.acn.2006.06.010

19. Fernaeus SE, Östberg P, Wahlund LO, Hellström Å. Memory factors in Rey AVLT: Implications for early staging of cognitive decline. *Scand J Psychol*. 2014;55(6):546-553. doi:10.1111/sjop.12157

20. Ding X, Charnigo RJ, Schmitt FA, Kryscio RJ, Abner EL. Evaluating trajectories of episodic memory in normal cognition and mild cognitive impairment: Results from ADNI. *PLoS One*. 2019;14(2):e0212435. doi:10.1371/journal.pone.0212435

21. Landau SM, Harvey D, Madison CM, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*. 2010;75(3):230-238. doi:10.1212/WNL.0b013e3181e8e8b8

22. Alenius M, Koskinen S, Hallikainen I, et al. Cognitive performance among cognitively healthy adults aged 30-100 years. *Dement Geriatr Cogn Dis Extra*. 2019;9(1):11-23. doi:10.1159/000495657

23. Skirbekk V, Loichinger E, Weber D. Variation in cognitive functioning as a refined approach to comparing aging across countries. *Proc Natl Acad Sci U S A*. 2012;109(3):770-774. doi:10.1073/pnas.1112173109

24. McCarrey A, An Y, Kitner-Triolo M, Ferrucci L, Resnick S. Sex differences in cognitive trajectories in clinically normal older adults. *Psychol Aging*. 2016;31(2):166-175. doi:10.1037/pag0000070.Sex

25. Krueger LE, Salthouse TA. Differences in acquisition, not retention, largely contribute to sex differences in multitrial word recall performance. *Pers Individ Dif*. 2010;49(7):768-772. doi:10.1016/j.paid.2010.06.024

26. Ferreira Correia A, Campagna Osorio I. The Rey Auditory Verbal Learning Test: Normative data developed for the Venezuelan population. *Arch Clin Neuropsychol*. 2014;29(2):206-215. doi:10.1093/arclin/act070

27. Asperholm M, Nagar S, Dekhtyar S, Herlitz A. The magnitude of sex differences in verbal episodic memory increases with social progress: Data from 54 countries across 40 years. *PLoS One*. 2019;14(4):10-14. doi:10.1371/journal.pone.0214945

28. Lezak MD. Memory I: Tests. In: Lezak MD, ed. *Neuropsychological Assessment*. Oxford University Press; 1995:429-489.

29. Crawford JR, Stewart LE, Moore JW. Demonstration of savings on the AVLT and development of a parallel form. *J Clin Exp Neuropsychol*. 1989;11(6):975-981. doi:10.1080/01688638908400950

30. Uchiyama CL, D'Elia LF, Dellinger AM, et al. Alternate forms of the Auditory-Verbal Learning Test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Arch Clin Neuropsychol*. 1995;10(2):133-145. doi:10.1093/arclin/10.2.133

31. Hoogendam YY, Hofman A, Van Der Geest JN, Van Der Lugt A, Ikram MA. Patterns of cognitive function in aging: the Rotterdam Study. *Eur J Epidemiol*. 2014;29(2):133-140. doi:10.1007/s10654-014-9885-4

32. Shapiro DM, Harrison DW. Alternate forms of the AVLT: A procedure and test of form equivalency. *Arch Clin Neuropsychol*. 1990;5(4):405-410. doi:10.1093/arclin/5.4.405

33. Majdan A, Sziklas V, Jones-Gotman M. Performance of healthy subjects and patients with resection from the anterior temporal lobe on matched tests of verbal and visuoperceptual learning. *J Clin Exp Neuropsychol*. 1996;18(3):416-430. doi:10.1080/01688639608408998

34. Van Den Burg W, Kingma A. Performance of 225 Dutch school children on Rey's Auditory Verbal Learning Test (AVLT): Parallel test-retest reliabilities with an interval of 3 months and normative data. *Arch Clin Neuropsychol*. 1999;14(6):545-559. doi:10.1016/S0887-6177(98)00042-0

35. Strauss E, Sherman E, Spreen O. *A Compendium of Neuropsychological Tests - Adminstration, Norms, and Commentary*. 3rd ed. (Strauss E, Sherman E, Spreen O, eds.). Oxford University Press; 2006.

36. Brysbaert M, Buchmeier M, Conrad M, Jacobs AM, Bölte J, Böhl A. The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Exp Psychol*. 2011;58(5):412-424. doi:10.1027/1618-3169/a000123

37. Suresh K. An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4(1):8-11. doi:10.4103/0974-1208.82352

38. Coors A, Merten N, Ward DD, Schmid M, Breteler MMB, Ettinger U. Strong age but weak sex effects in eye movement performance in the general adult population: Evidence from the Rhineland Study. *Vision Res*. 2021;178:124-133. doi:10.1016/j.visres.2020.10.004

39. Corsi PM. Human memory and the medial temporal region of the brain. Published online 1972.

40. Reitan RM. *Trail Making Test Manual for Administration and Scoring*. Reitan Neuropsychological Laboratory; 1992.

41. Lehrl S. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. Spitta Verlag GmbH & Co. KG; 2005.

42. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum Association; 1988.

43. Lorah J. Effect size measures for multilevel models: definition, interpretation, and TIMSS example. *Large-Scale Assessments Educ*. 2018;6:8. doi:10.1186/s40536-018-0061-2

44. Nakagawa S, Johnson PCD, Schielzeth H. The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J R Soc Interface*. 2017;14(134):20170213. doi:10.1098/rsif.2017.0213

45. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Front Psychol*. 2012;3(111):1-6. doi:10.3389/fpsyg.2012.00111

46. Zeileis A, Kleiber C, Jackman S. Regression models for count data in R. *J Stat Softw*. 2008;27(8):1-25. doi:10.18637/jss.v027.i08

47. Rigby R., Stasinopoulos DM. Generalized additive models for location, scale and shape. *Appl Stat*. 2005;54(3):507-554. doi:10.1111/j.1467-9876.2005.00510.x

48. R Core Team. R: A language and environment for statistical computing. Published online 2019. https://www.r-project.org/

49. RStudio Team. RStudio: Integrated development for R. Published online 2019. http://www.rstudio.com/

50. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and nonlinear mixed effects models (version 3.1-142). Published online 2019. https://cran.r-project.org/package=nlme

51. Van der Elst W, Molenberghs G, van Tetering M, Jolles J. Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. *Clin Neuropsychol*. 2017;31(6-7):1173-1187. doi:10.1080/13854046.2017.1294202

52. Van der Elst W, Molenberghs G, Van Boxtel MPJ, Jolles J. Establishing normative data for repeated cognitive assessment: A comparison of different statistical methods. *Behav Res Methods*. 2013;45(4):1073-1086. doi:10.3758/s13428-012-0305-y

53. Quené H, van den Bergh H. Examples of mixed-effects modeling with crossed random effects and with binomial data. *J Mem Lang*. 2008;59(4):413-425. doi:10.1016/j.jml.2008.02.002

54. García-Pérez MA. Statistical criteria for parallel tests: A comparison of accuracy and power. *Behav Res Methods*. 2013;45(4):999-1010. doi:10.3758/s13428-013-0328-z

55. Cavaco S, Gonçalves A, Pinto C, et al. Auditory verbal learning test in a large nonclinical Portuguese population. *Appl Neuropsychol*. 2015;22(5):321-331. doi:10.1080/23279095.2014.927767

56. Blachstein H, Vakil E. Verbal learning across the lifespan: an analysis of the components of the learning curve. *Aging, Neuropsychol Cogn*. 2015;23(2):133-153. doi:10.1080/13825585.2015.1063579

57. Geffen G, Moar KJ, O'Hanlon AP, Clark CR, Geffen LB. Performance measures of 16- to 86-year-old males and females on the auditory verbal learning test. *Clin Neuropsychol*. 1990;4(1):45-63. doi:10.1080/13854049008401496

58. Woodard JL. Memory performance indexes for the Rey Auditory Verbal Learning Test. In: Poreh AM, ed. *The Quantified Process Approach to Neuropsychological Assessment*. 1st ed. Psychology Press Taylor & Fracis Group; 2016:105-142. doi:10.4324/9780203720899

59. Head D, Kennedy KM, Rodrigue KM, Raz N. Age-differences in perseveration: Cognitive and neuroanatomical mediators of performance on the Wisconsin Card Sorting Test. *Neuropsychologia*. 2009;47(4):1200-1203. doi:10.1016/j.neuropsychologia.2009.01.003

60. Foldi NS, Helm-Estabrooks N, Redfield J, Nickel DG. Perseveration in normal aging: A comparison of perseveration rates on design fluency and verbal generative tasks. *Aging, Neuropsychol Cogn*. 2003;10(4):268-280. doi:10.1076/anec.10.4.268.28970

61. Golchert J, Roehr S, Luck T, et al. Women outperform men in verbal episodic memory even in oldest-old age: 13-year longitudinal results of the AgeCoDe/AgeQualiDe study. *J Alzheimer's Dis*. 2019;69(3):857-869. doi:10.3233/JAD-180949

62. Tuokko H, Griffith LE, Simard M, Taler V. Cognitive measures in the Canadian Longitudinal Study on Aging. *Clin Neuropsychol*. 2016;31(1):233-250. doi:10.1080/13854046.2016.1254279

63. Statistisches Bundesamt. Internationale Bildungsindikatoren im Ländervergleich. Published online 2020:1-88. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Publikationen/Downloads-Bildungsstand/bildungsindikatoren-1023017207004.pdf?__blob=publicationFile

64. Baddeley AD, Thomson N, Buchanan M. Word length and the structure of short-term memory. *J Verbal Learning Verbal Behav*. 1975;14(6):575-589. doi:10.1016/S0022-5371(75)80045-4

65. Braver TS, West R. Working memory, executive control and aging. In: Craik FIM, Salthouse TA, eds. *The Handbook of Aging and Cognition*. 3rd Edition. Psychology Press Taylor & Fracis Group; 2008:311-372. doi:10.1002/acp.1505

3.2 A functional MRI paradigm for efficient mapping of memory encoding across sensory conditions

Meta M. Boenniger,[1] Kersten Diers,[2] Sibylle C. Herholz,[1] Mohammad Shahid,[1] Tony Stöcker,[3] Monique M. B. Breteler,[1,4] and Willem Huijbers[1,5]

[1]Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[2]Image Analysis Group, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[3]MR Physics, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[4]Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Bonn, Germany

[5]Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Journal: Frontiers in Human Neuroscience 2021; 14:591721,
doi: 10.3389/fnhum.2020.591721

ABSTRACT

We introduce a new and time-efficient memory-encoding paradigm for functional magnetic resonance imaging (fMRI). This paradigm is optimized for mapping multiple contrasts using a mixed design, using auditory (environmental/vocal) and visual (scene/face) stimuli. We demonstrate that the paradigm evokes robust neuronal activity in typical sensory and memory networks. We were able to detect auditory and visual sensory-specific encoding activities in auditory and visual cortices. Also, we detected stimulus-selective activation in environmental-, voice-, scene-, and face-selective brain regions (parahippocampal place and fusiform face area). A subsequent recognition task allowed the detection of sensory-specific encoding success activity (ESA) in both auditory and visual cortices, as well as sensory-unspecific positive ESA in the hippocampus. Further, sensory-unspecific negative ESA was observed in the precuneus. Among others, the parallel mixed design enabled sustained and transient activity comparison in contrast to rest blocks. Sustained and transient activations showed great overlap in most sensory brain regions, whereas several regions, typically associated with the default-mode network, showed transient rather than sustained deactivation. We also show that the use of a parallel mixed model had relatively little influence on positive or negative ESA. Together, these results demonstrate a feasible, versatile, and brief memory-encoding task, which includes multiple sensory stimuli to guarantee a comprehensive measurement. This task is especially suitable for large-scale clinical or population studies, which aim to test task-evoked sensory-specific and sensory-unspecific memory-encoding performance as well as broad sensory activity across the lifespan within a very limited time frame.

INTRODUCTION

With neurodegenerative diseases as one of the main challenges in aging populations, the precise, comprehensive, and robust measurement of cognitive functions is of great importance. Functional magnetic resonance imaging (fMRI) is one measurement that helps us to bridge the space between biology and behavioral outcomes. Several large-scale studies have employed fMRI to map brain activity in the general population, including the Rotterdam Study,[1] UK Biobank,[2] and the Rhineland Study.[3] These large-scale population studies are usually not designed to answer one specific hypothesis. Rather, they aim to perform an extensive and deep phenotyping that allows addressing multiple questions. As they are mostly prospective studies, they also need to anticipate future questions. Therefore, tasks and paradigms should ideally be as versatile as possible. In the absence of a specific hypothesis, resting-state fMRI is often employed, mostly for practical considerations, as it is rather easy to apply and can also inform about neural dysfunction.[4] Task-evoked fMRI provides complementary information, that is, the brain's response to specific demands,[5,6] and evokes activity in cortical networks under more restrained conditions.[7,8] Therefore, task fMRI is often considered. However, most conventional task paradigms are not easily applied in clinical or large-scale population studies[9] for the following reasons.

First, conventional task paradigms from cognitive neuroscience are typically developed and applied in experimental studies that pose less time constraints than population studies. However, in clinical or large-scale population studies, acquisition time is often more restricted, as the burden to participants, or patients, should be limited and costs add up easily. Additionally, fMRI acquisition time typically competes with anatomical or clinically motivated MRI sequences, including T1, T2, fluid-attenuated inversion recovery (FLAIR), susceptibility-weighted imaging (SWI), perfusion, and diffusion.[10,11] Thus, to be feasible for clinical or population-based imaging, a task paradigm should be as time-efficient as possible.

Second, conventional fMRI task paradigms have often been developed in homogenous cohorts of young adults. In a large-scale population or clinical studies, the cohort of participants is typically more heterogeneous with respect to age, education, lifestyle, and health factors. This heterogeneity can result in problems when task instructions are

tailored to a specific age group (such as young adults). As a consequence, paradigms might show ceiling and/or floor effects for subgroups. Thus, an ideal paradigm should have very simple or no instructions and yet remain informative across the entire cohort.

Finally, conventional task paradigms are typically designed to answer a specific hypothesis, often from the field of cognitive neuroscience. As mentioned above, large-scale, population-based studies mostly aim to employ fMRI to estimate neuronal activity related to multiple research questions or outcomes at the same time. An ideal task paradigm for population-based studies should permit the analysis of multiple contrasts that span a wide range of cognitive functions.

To address these various requirements, we designed a novel task paradigm that we consider especially suited for large-scale studies. It measures predominantly memory encoding, but also perception and attention in both the auditory and visual domains within ten minutes of fMRI acquisition time using simple instructions. To our knowledge, memory-encoding paradigms so far presented stimuli of one sensory condition or did face-name associative memory tasks[12–16] within a similar time frame. We optimized our task to allow mapping of a versatile number of contrasts that are relatively straightforward to interpret. To enable the separation of sensory-specific and sensory-unspecific activities,[17–19] we used two sensory modalities, auditory and visual. Twenty-five percent of the total time consisted of passive rest blocks as baseline/rest condition.[20] Each sensory condition contained two distinct sub-conditions to cover a wide range of information on visual and auditory system activations as well as joined activation for sensory-unspecific functions like overall memory. Within the visual condition, we chose to present faces and spatial scenes, motivated by work on face-selective and scene-selective brain regions.[21–24] Further, those stimuli seemed to show differences in age-related reductions in neural dedifferentiation, which makes them interesting for longitudinal studies.[25] To select auditory stimuli on a similar level of specificity, we chose voice and environmental stimuli motivated by previous work on voice-selective brain regions.[26–31] This decision was further supported by studies showing that similarities as well as differences exist between the regional activation of voice and face perception.[32] Due to the simplicity of the design and to keep the paradigm language free, we did not include language stimuli. A post-fMRI recognition test, with previously seen/heard and novel items, enables the computation of

contrasts between subsequently remembered (hit) and forgotten (miss) items.[24,33–35] In the following, we will refer to these contrasts as encoding success activity (ESA). We used a parallel mixed block/event design to include a large number of stimuli within a limited time and to enable the already versatile number of contrasts also for the separation of sustained (block) and transient (event) activities.[36–38] Differentiating both can help to get a more complex understanding of the functional processes underlying a task. Sustained effects give more information about the maintenance of activity throughout a set of stimuli, for example, representing also overall attentional performance or arousal, whereas transient effects are specific for each trial of a task.[37]

Thereby, our task allows a large degree of flexibility to analyze the data in multiple ways with regard to other outcomes of interest. This is important in studies spanning years to decades, as research questions and analysis techniques change over time. In this study, we introduce our task paradigm and demonstrate several possible analyses to generate a range of different behavioral and neuronal measures relating to perception and memory encoding. These outcome measures are then available for further analyses in the context of the overall population study.

MATERIALS AND METHODS

Participants

We recruited 60 young adults between the ages of 19 and 30 years (M = 24.18, SD = 2.90; 36 females), from the University of Bonn community in the context of the pilot studies for the Rhineland Study, a prospective cohort study. The study was carried out in accordance with the recommendations of the International Council for Harmonization (ICH) Good Clinical Practice (GCP) standards (ICH-GCP). We obtained written informed consent from all participants in accordance with the Declaration of Helsinki. No incentives were offered to the participants. The medical ethics committee of the Medical Faculty of the University of Bonn approved the study. All participants had normal or corrected-to-normal vision. Hearing levels were calibrated individually before the experiment, for the sounds to be easily audible above the scanner noise. For one participant, visual retrieval data were not available in the fMRI analysis. ESA contrasts for this participant were therefore analyzed only on the basis of the auditory retrieval information. To detect possible floor effects of

the task, we obtained behavioral task data of 21 persons older than 30 years (M = 52.71, SD = 15.55; age range = 31-77; 12 females) (see Supplementary Material “Behavioral results in older adults”).

Stimuli

A total of 160 auditory and 160 visual items were presented during the encoding task. Auditory stimuli had durations between 538 and 2,771 ms (M = 1,630 ms, SD = 488 ms) and consisted of 80 environmental and 80 human vocal sounds. The environmental sounds included a mix of sounds from animals, traffic, tools, and musical instruments, selected from previous auditory experiments.[18,26,39] The vocal sounds consisted of vocal utterances, void of semantic content, such as laughing, crying, or coughing, selected from previous experiments, from the Oxford Vocal (OxVoc) Sounds database[26,40] or were recorded for the purpose of this study. The recordings were from various male and female voices. Duration of the auditory stimuli was not equalized, because some are by nature rather short but nevertheless distinct, whereas others need a longer duration to be distinct (e.g. cockcrow/doorbell vs. laughter/wind). To match stimuli for low-level physical properties, we normalized all auditory stimuli to the same amplitude using version 2.0.6 of Audacity® recording and editing software. The visual items consisted of color photographs of 80 faces (size 570 × 360 pixels) and 80 scenes (size 500 × 375 pixels) on a black background. Face stimuli contained faces from individuals with various ethnicities, between 18 and 90 years of age, with an equal number of male and female faces selected from previous experiments.[41–43] Scenes were pictures from nature or urban outdoor environments selected from Huijbers et al..[44] Colors from the original scenic images were slightly de-saturated to match the color contrast in the facial images.

From all available stimuli, we selected the final set of 160 stimuli with the aim to reach a hit-rate of ~50 % (for an explanation, see section Behavioral analysis) and a false alarm (FA) rate as low as possible.

Auditory stimuli were presented via S14 Insert Earphones (Sensimetrics, Malden, USA). Visual stimuli were presented on a screen located at the head of the magnet bore and

seen via a mirror mounted on the head coil. All stimuli were presented using PsychoPy software v1.82,[45] running on a Windows PC.

MRI acquisition

fMRI data were acquired with a 3-Tesla Siemens MAGNETOM Prisma system (Siemens Medical Systems, Erlangen, Germany). The scanner was equipped with a 64-channel phased-array head/neck coil. We used inflatable air pads to restrict head movement, and participants were instructed to lie still for the duration of the scan. For the applicability in large-scale testing, we decided on a standard fMRI scanning protocol: we acquired two task fMRI sessions of 140 volumes using echo-planar imaging (EPI), including four dummies. Each volume consisted of 32 axial slices of 3 mm thickness with a 0.75 mm gap. The repetition time (TR) was 2,000 ms, echo time (TE) was 30 ms, flip angle was 84°, readout bandwidth was 2,300 Hz/pixel, the slice orientation was anterior commissure-posterior commissure (AC-PC), and field of view (FOV) was 192 × 192 mm, resulting in an effective voxel size of 3.0 × 3.0 × 3.75 mm.

Task design and implementation

The task was designed as a mixed model[37] and included 180 events (trials), grouped into 32 blocks (Figure 1). Out of these 32 blocks, eight were rest blocks (fixation), eight isolated auditory blocks, eight isolated visual blocks, and eight parallel auditory/visual blocks. Sixteen blocks contained auditory stimuli, half isolated auditory and half parallel with visual images. Of these 16 blocks, eight blocks contained environmental sounds (four isolated and four parallel), and eight blocks contained vocal sounds (four isolated and four parallel). Similarly, for the visual blocks, half were presented in isolation and half in parallel with sounds (scene and face images equally distributed). In each block, five items —sounds, images, or both— were presented for a total of 16 s per block (Figure 1). Within the auditory blocks, the inter-trial interval between items was 200-2,700 ms. Within the visual stimulus blocks, the inter-trial interval between items was 200-2,200 ms. The difference in the inter-trial intervals between auditory and visual blocks is due to the variable duration of sounds. Each image was presented for exactly 2,000 ms. Inter-trial

intervals as well as the order of blocks and the order of stimuli within the blocks were once randomly assigned and remained the same for all participants. A white fixation cross on black background was shown during the rest blocks, the inter-trial intervals, the isolated auditory blocks, and the initial and final 8 s of each run. By design, we tried to ensure that the different blocks and items would result in separate, uncorrelated regressors (see section Parallel mixed model analysis). As we cannot predetermine which items will be remembered or forgotten, we also evaluated the collinearity of the regressors after data collection (see section Parallel mixed model analysis).

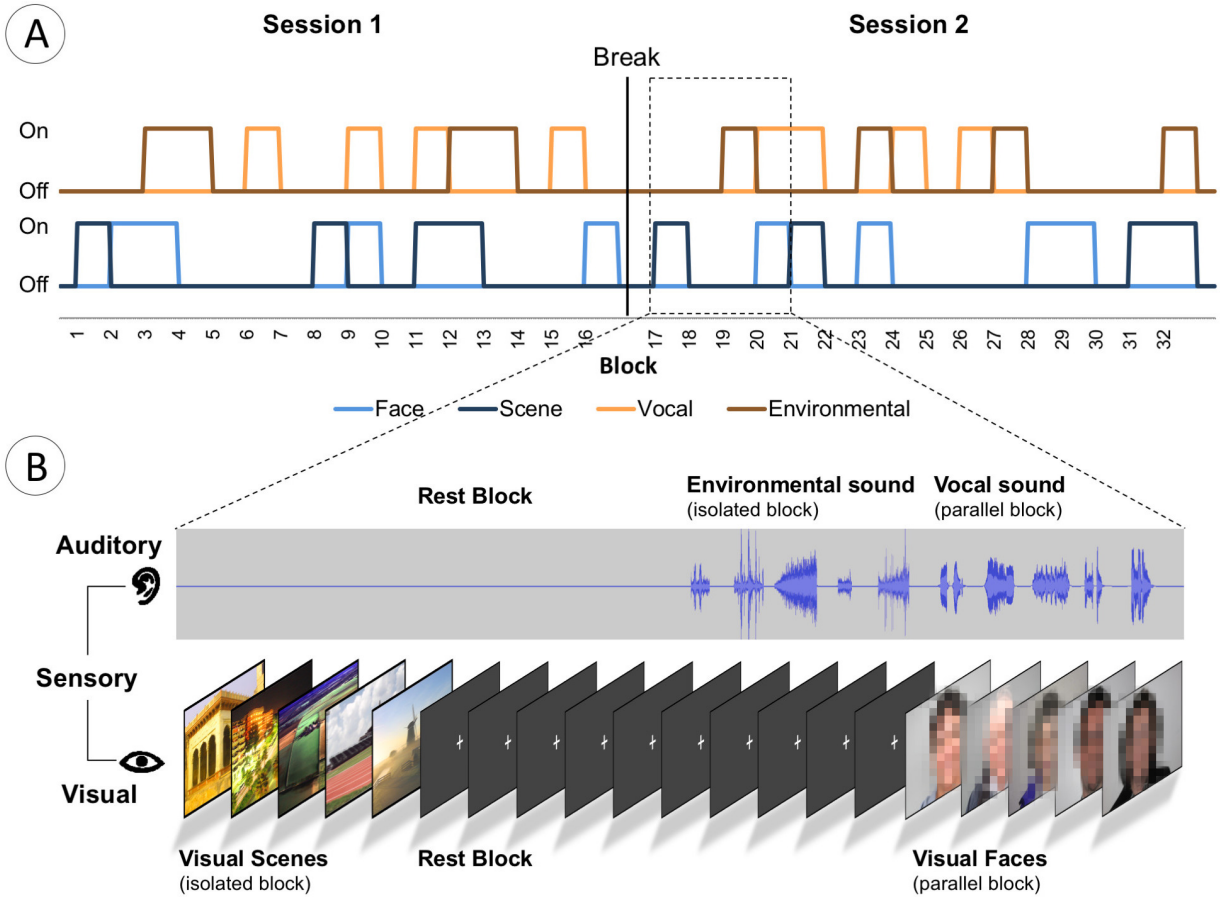

Figure 1. Scheme of the task paradigm. (A) The order of presented stimuli and sensory conditions. “On” represents presentation of the stimuli, and “Off” represents no presentation of the stimuli. Each block contains five stimuli of one category and lasted 16 s. In parallel conditions, five auditory and five visual stimuli are presented in one block. The auditory events consisted of environmental or vocal stimuli (see auditory timeline). The visual events consisted of face or scene stimuli (see visual timeline; faced were not pixelated in the original task). (B) Four exemplary blocks in detail.

The task presentation was distributed over two sessions (containing the same number of blocks for each stimulus and presentation condition) of 4:54 minutes each separated by a short question about the participants' well-being. At the beginning of these sessions, participants were given written instructions via the screen to pay attention to the sounds and images ("please watch and listen carefully"). No motor responses were required in our task, which had several advantages. Apart from keeping the task simple, movement artifacts during scanning were minimized. Also, the lack of motor activity facilitates the interpretation of sensory and memory-related fMRI data across the life span.[46,47]

Before the fMRI sessions, vision and hearing abilities were corrected to normal, by using MRI-compatible glasses and a volume adjustment during the initial scout scan, respectively. Following the scout scan, participants did a very short training session of eight visual and eight auditory items including encoding and retrieval, to get acquainted with the task procedures and to ensure they understood the instructions. Our following encoding task was therefore explicit. After completing the two memory-encoding sessions, participants' memory retrieval was tested by two separate subsequent memory tests. Recognition of auditory stimuli was tested first, followed by a visual recognition test. Across the auditory and visual recognition tests, 160 previously encoded (old) and 160 novel (new) items were presented (80 environmental sounds/80 vocal sounds/80 face images/80 scene images). The participants responded with two buttons ("Yes" and "No") to a forced-choice question (in German): "Did you hear/see this item previously?" ("Haben Sie das Geräusch bereits gehört?" or "Haben Sie das Bild bereits gesehen?"). The recognition tests were self-paced, and items were presented in blocks. In each block, five old items (previously encoded) and five new items were presented in a random order. In each block, items were of the same type. Across the blocks, the presentation order of the encoding intervals was maintained to ensure an approximately equal time distance between encoding and retrieval. Both recognition tests were done inside the MRI bore immediately after the encoding runs. The auditory recognition test was done during a diffusion MRI scan, and the visual recognition test was performed during an anatomical T1-MPRAGE scan. Diffusion and anatomical MRI data are not included in this manuscript, yet some of that data have been examined in relation to head motion.[8] The fMRI and behavioral data in this paper have not been published previously. All task scripts are uploaded under https://www.rheinland-studie.de/data-code/boenniger2020.

Behavioral analysis

Behavioral analyses were implemented in R v3.3.2 (http://www.r-project.org/). To quantify memory performance, we examined the percentage of correct responses for previously presented items (labeled as hit-rate) and the percentage of incorrect responses for new items (labeled as FA-rate). We also calculated the discriminability index d-prime (*d'*), by taking the z-standardized hit-rate minus the z-standardized FA-rate. Additionally, we calculated the response bias (*c*) by taking the sum of the z-standardized hit- and FA-rates multiplied by -0.5. Differences between hit- and FA-rates were calculated using paired and two-sided t-tests. A one-sample t-test was used to examine the response bias. To assess the main effects and the possible interaction between sensory modality (auditory/visual) and presentation condition (isolated/parallel) on memory performance, we used an ANOVA. Correlation analyses described in the supplements employ Pearson's method, unless otherwise indicated. Reliability analysis was done by splitting up the task into its two sessions (for details, see Supplementary Material "Analysis of reliability") and calculating the intraclass correlation coefficients (ICC)[48,49] with a two-way model using single units for each participant, estimating the consistency between the two sessions.

Functional MRI preprocessing

fMRI data were preprocessed using MATLAB (MathWorks, Natick, MA, USA), the Statistical Parametric Mapping Toolbox (SPM8, UCL, London, UK), and GLM Flex (MGH, http://mrtools.mgh.harvard.edu/index.php/GLM_Flex, MA, USA). First, we dropped the four dummy volumes. Second, we realigned the time series to the first volume. Third, we normalized the data to a standard EPI template in Montreal Neurological Institute (MNI) 152 space. Fourth, we smoothed the data with a full-width-half-maximum (FWHM) kernel of 8mm. For assessing the reliability, we split the task into two sessions (for details see Supplementary Material "Analysis of reliability") calculated on the basis of the slice time-corrected and normalized data. We calculated ICC values before smoothing the data, using a publicly available online script for MATLAB by C. Pernet (https://github.com/CPernet/spmrt/blob/master/spmrt_fMRI_ICC.m, downloaded 8th February 2018). After calculating the ICCs on the voxel level, we smoothed the group-

level ICC maps with a FWHM kernel of 8 mm for visualization purposes, as this makes it easier to appreciate the spatial overlap between the contrast and the ICC map.

Parallel mixed model analysis

The subject-level analyses were conducted in SPM8. For the main analyses, the SPM regressors were modeled according to the parallel mixed block/event design.[37] In the Supplementary Material, we also added a model comparison where we modeled the task data according to a block-only design and an event-only design using the respective regressors separately (see Supplementary Material “Model comparison between mixed, block- and event-only modeling”).

In the mixed design, we included two block regressors: one for the auditory blocks and one for the visual blocks (Figure 2). The block onsets were convolved with the canonical hemodynamic response function using the durations. Passive rest blocks (fixation) were not modeled explicitly. The block regressors were solely determined by the task design and therefore fixed across subjects. In addition, we also added eight unique event regressors, which were subject specific, as they were determined by the combination of the task design (stimulus type: environmental sounds/vocal sounds/face images/scene images) and the participants’ performance on the subsequent memory tasks (memory performance: hit/miss) (Figure 2). To avoid collinearity,[50] we included the parallel (auditory and visual) presentation condition in an equal amount to the isolated (auditory or visual) and rest (fixation) blocks to the task. In addition, we modeled blocks for both sensory conditions (auditory/visual), whereas events were modeled for each stimulus condition (environmental sounds/vocal sounds/scene images/face images) separately. To prevent correlations between regressors due to participants who remembered or forgot too many items (events) presented within one block, we aimed for a conservative response bias (see section Behavioral results). This also ensured roughly equal hits/misses, so hits or misses did not dominate single blocks. After data collection, before the analysis, we checked the hemodynamic regressors for collinearity using a correlation analysis. All regressors have shown to be largely independent with $r = 0.3$ within sensory conditions and r around zero between sensory conditions. On average, the event-related regressors were modeled based on 20.65 environmental hits (SD = 5.87), 19.25 environmental

misses (SD = 5.77), 24.63 vocal hits (SD = 6.35), 15.30 vocal misses (SD = 6.30), 26.42 scene hits (SD = 6.16), 13.58 scene misses (SD = 6.16), 23.73 face hits (SD = 6.64), and 16.27 face misses (SD = 6.64). The event onsets were convolved with the canonical hemodynamic response function using no duration. Further, the subject-level models also included regressors for motion parameters, bad-volume regressors, and a high-pass filter (1/128 Hz). The bad volumes were defined by the amount of absolute movement in relation to the previous scans, using a threshold >0.75 mm or 1.5° in one or more directions.

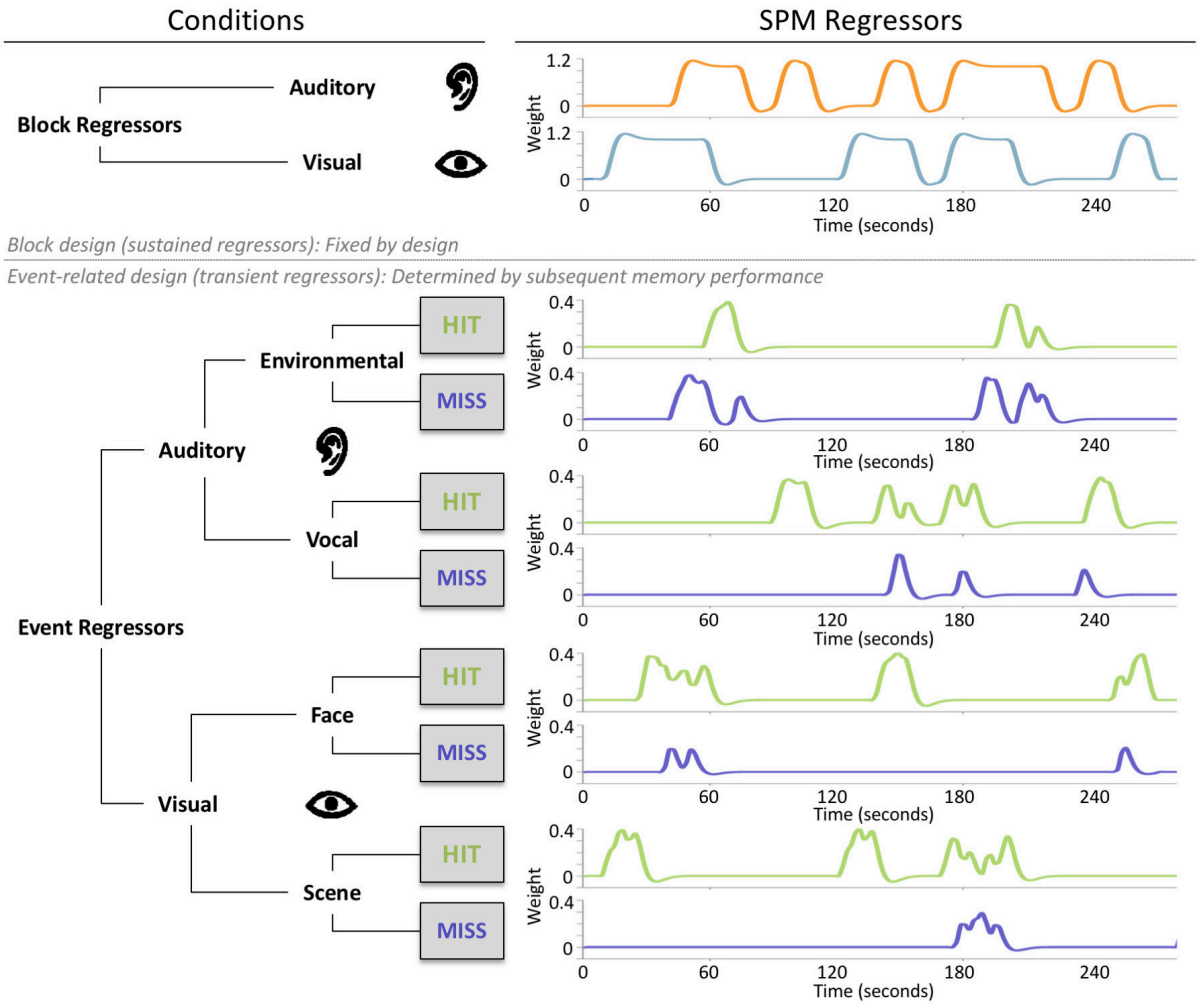

Figure 2. The parallel mixed design consisted of two fixed block regressors, one for the auditory and one for the visual blocks, and eight subject-specific event regressors. These event-related regressors were determined by the combination of stimulus types, auditory (environmental/vocal) or visual (face/scene) and subsequent memory performance (hits/misses). Event regressors represent the SPM regressors of a single exemplary subject.

We defined ten contrasts. First, the block-based contrast was performed between auditory and visual blocks (c1) to assess activity caused by the different sensory conditions. Second, two event-related contrasts based on stimulus type was performed to assess within the sensory conditions differences between stimuli types: (c2) environmental sounds vs. vocal sounds masked by auditory activity greater than visual activity and (c3) face images vs. scene images masked by visual activity greater than auditory activity. Third, we defined ESA based on the events during isolated blocks: (c4) all (visual and auditory) hits vs. all misses (sensory-unspecific ESA), (c5) auditory hits vs. auditory misses (auditory ESA), and (c6) visual hits vs. visual misses (visual ESA). We included ESA only for each sensory type (auditory and visual stimuli), as we considered the type-specific ESA maps (environmental sounds, vocal sounds, scene images, and face images) to be too detailed and to have not enough trials. Fourth, we defined contrasts relative to the rest condition (fixation), to examine differences and similarities between sustained and transient activities: (c7) auditory blocks vs. rest, (c8) isolated auditory events vs. rest, (c9) visual blocks vs. rest, and (c10) isolated visual events vs. rest.

For all group maps, we used a global threshold of $p < .05$ [false discovery rate (FDR) corrected] with a minimum cluster size of five voxels (no cluster-size correction). The same threshold was used to define the masks for the conjunction analyses (c2/c3). Note that the remaining activities within the mask also had to survive the global threshold ($p < .05$, FDR corrected). Statistical group maps were projected to the cortical surface using FreeSurfer (v5.1) via a standard MNI to the FreeSurfer average template transformation or were resliced to 2.0 × 2.0 × 2.0 mm voxels and overlaid on the standard SPM8 individual T1-weighted volume.

RESULTS

Behavioral results

The subsequent memory performance is listed in Table 1. For each sensory and stimulus condition, the hit-rate was significantly greater than the FA-rate. These differences between the hit- and FA-rates indicate that participants were able to successfully encode items in each category. The duration of subsequent memory test was between 10 and 19 min. The auditory retrieval took on average 7.32 minutes (SD = 0.86, range 6.40-

11.16) and visual retrieval 3.86 minutes (SD = 0.77, range 3.03-8.18). Within the auditory blocks, the inter-trial intervals for hits and misses were on average 1,556 ms (range 194-2,713 ms) and 1,509 ms (range 194-2,713 ms), respectively. Within visual blocks, the inter-trial intervals for hits and misses were on average 1,136 (range 200-2,200 ms) and 1,171 ms (range 200-2,200 ms), respectively.

Across all conditions, we found a d' of 1.19 (SD = 0.40) and a c of 0.34 (SD = 0.27) [auditory: c = 0.32 (SD = 0.33); visual: c = 0.39 (SD = 0.33)]. The response bias indicated that participants were relatively conservative [$t_{(59)} = 9.59$, $p < .001$] and thus more likely to rate items as "new". Paired t-tests indicated that memory performance was better for visual items compared with auditory items and for scene images better than for face images, but there was no difference in memory performance between environmental and vocal auditory stimuli (Table 1). Similar results, with slightly lower d' but a comparable response bias, have been observed for the small sample of older participants (see Supplementary Material "Behavioral results in older adults").

In our paradigm, stimuli were presented either in isolation or in parallel with stimuli of the other sensory modality. We computed separate d' values for each of the presentation conditions (parallel/isolated) and sensory condition (auditory/visual). d' values for subsequent memory of auditory stimuli were $M_{isolated} = 1.03$ ($SD_{isolated} = 0.51$) and $M_{parallel} = 0.95$ ($SD_{parallel} = 0.45$) and for visual stimuli d' values of $M_{isolated} = 1.69$ ($SD_{isolated} = 0.70$) and $M_{parallel} = 1.31$ ($SD_{parallel} = 0.60$). Results of the ANOVA supported a better subsequent memory performance for visual than for auditory stimuli independent of the presentation condition (isolated/parallel) [$F_{(1,59)} = 37.29$, $p < .001$]. Also the presentation condition showed a main effect indicating a better subsequent memory performance for items presented in isolation independent of the sensory modality [$F_{(1,59)} = 40.60$, $p < .001$]. Further, we found an interaction effect between sensory modality and presentation condition [$F_{(1,59)} = 25.62$, $p < .001$], which suggested that the parallel presentation of auditory and visual items was more detrimental to learning of visual information than of auditory information.

Table 1. Memory performance

	Hit-rate		FA-rate		Hit-rate vs. FA-rate			*d'*				
	M	SD	M	SD	*t*	df	*p*	M	SD	*t*	df	*p*
Auditory	0.57	0.14	0.23	0.13	17.13	59	< 0.001	0.99	0.46			
Visual	0.63	0.14	0.15	0.12	21.81	59	< 0.001	1.48	0.58			
Auditory vs. Visual										6.09	59	<0.001
Environmental	0.52	0.15	0.19	0.11	15.73	59	< 0.001	1.02	0.54			
Vocal	0.62	0.16	0.27	0.16	15.8	59	< 0.001	1.02	0.50			
Environmental vs. Vocal										0.00	59	1.000
Face	0.59	0.17	0.19	0.15	17.35	59	< 0.001	1.21	0.61			
Scene	0.66	0.15	0.11	0.11	22.88	59	< 0.001	1.87	0.75			
Face vs. Scene										9.47	59	<0.001

Mean (M) and standard deviation (SD) for Hit-rate, False Alarm (FA)-rate and d-prime (*d'*) (n = 60). Paired t-tests are used to depict differences between Hit- and FA-rate and between *d'* of sensory and stimuli conditions.
Note: df indicates the degrees of freedom, *t* indicated the *t*-value and *p* indicates the *p*-value representing the significance level.

Sensory-specific activity

For the auditory blocks, we found the global maxima in the right auditory cortex and for the visual blocks in the left visual cortex (Figure 3 c1). For the environmental sounds, the maxima were in the right temporoparietal junction, and for the vocal sounds, in the right superior temporal gyrus (Figure 3 c2). For scenes, we found maxima in the left parahippocampal gyrus, and for faces, in the left fusiform gyrus (Figure 3 c3). See sensory-specific activity in Table 2 for the MNI coordinates and values of the global maxima (activation) and minima (deactivations). Supplementary Table A.2 in the supplement provides cluster specific peaks for all contrasts.

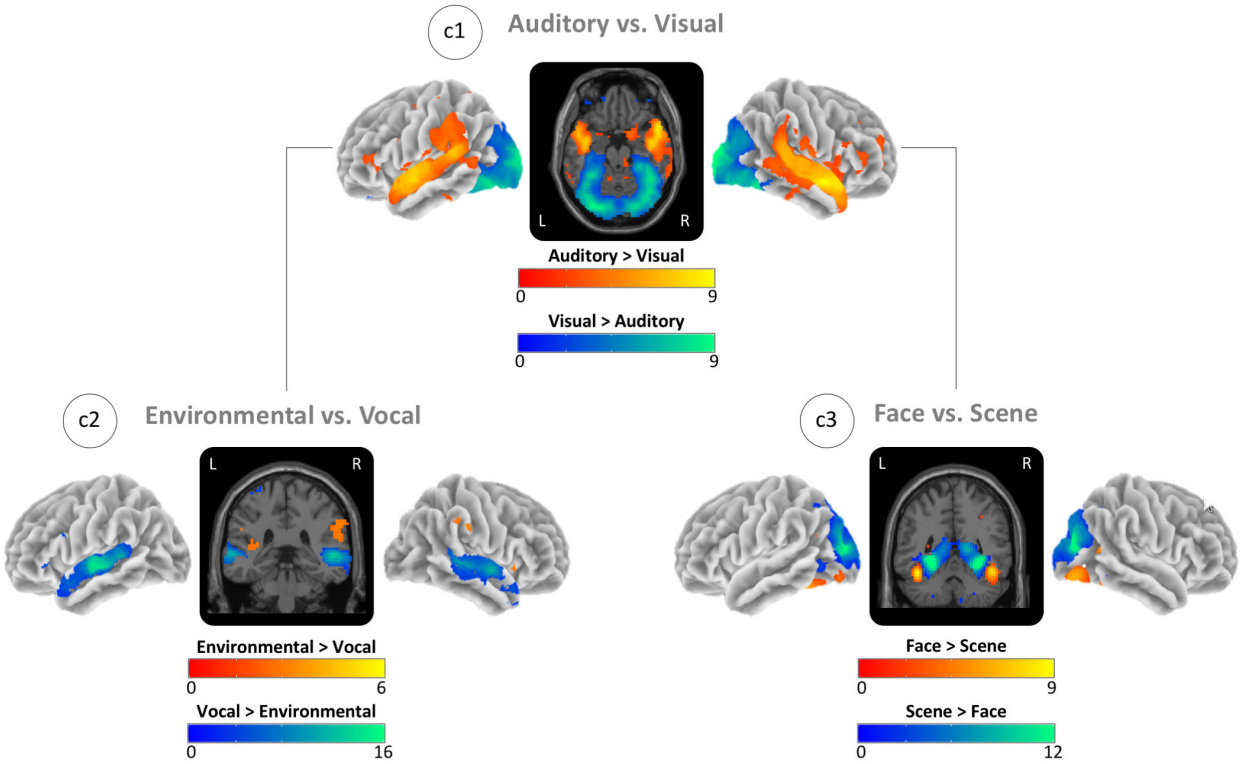

Figure 3. Task-based activity contrasts between (Auditory vs. Visual) and within (Environmental vs. Vocal and Face vs. Scene) sensory conditions. (c1) Block-based contrast between auditory and visual stimulus blocks. (c2) Event-related contrast between environmental and vocal sounds masked by auditory greater visual activity (see c1). (c3) Event-related contrast between face and scene images masked by visual greater auditory activity (see c1). Brain activity is shown at a threshold of $p < .05$ [false discovery rate (FDR) corrected], and the color intensity shows the *t*-value. Contrast maps are uploaded under https://neurovault.org/collections/IABCOPVN/.

Table 2. Maximally activated and deactivated brain regions for each contrast

Contrast		Region	$MNI_{(x,y,z)}$	*t*-value	BA
Sensory–Specific Activity					
c1	Auditory > Visual	Auditory cortex	54,-1, -13	9.90	22/41
	Visual > Auditory	Visual cortex	-3, -91, -4	13.13	17/18
c2	Environmental > Vocal	Temporoparietal junction	57, -28, 32	3.78	39/40
	Vocal > Environmental	Superior temporal gyrus	63, -1, -10	11.65	22
c3	Face > Scene	Fusiform gyrus	-42, -49, -22	8.17	37
	Scene > Face	Parahippocampal gyrus	-27, -49, -7	24.08	19/36
Encoding Success Activity (ESA) for isolated blocks					
c4	Positive ESA	Hippocampus	21, -7, -25	6.09	36/54
	Negative ESA	Precuneus	9, -70, 44	-6.37	7
c5	Positive auditory ESA	Auditory cortex	-60, -13, -4	5.72	22
	Negative auditory ESA	Precuneus	21, -55, 23	-5.06	23
c6	Positive visual ESA	Visual cortex	27, -91, -4	5.31	18
	Negative visual ESA	Precuneus	12, -67, 32	-5.53	7/31

Continuation Table 2.

Contrast		Region	$MNI_{(x,y,z)}$	*t*-value	BA
Sustained (blocks)					
c7	Auditory > rest (activation)	Superior temporal lobe	54, 2, -13	9.71	22
	Auditory < rest (deactivation)	Visual cortex	18, -100, 8	5.39	18
c9	Visual > rest (activation)	Primary visual cortex	-6, -88, -1	14.81	17
	Visual < rest (deactivation)	Temporoparietal junction	-63, -25, 26	6.01	40
Transient (events)					
c8	Auditory > rest (activation)	Primary auditory cortex	-42, -28, 8	7.89	41
	Auditory < rest (deactivation)	Putamen	21, 8, -13	7.10	49/52
c10	Visual > rest (activation)	Fusiform gyrus	30, -55, -13	8.79	37
	Visual < rest (deactivation)	Inferior temporal gyrus	-30, -49, 5	7.15	19/37

Contrasts represent: c1: block-based contrast between auditory versus visual stimuli blocks, c2: event-related contrast between environmental versus vocal sounds, c3: event-related contrast between face versus scene images, c4: ESA for all (visual and auditory) hits versus all misses, c5: ESA for auditory hits versus auditory misses, c6: ESA for visual hits versus visual misses, c7: auditory versus rest blocks, c9: visual versus rest blocks, c8: auditory events versus rest and c10: visual events versus rest. All brain regions are described with MNI coordinates (MNI(x,y,z)), *t*-values of the beta coefficients and the relating Brodmann-Area (BA). Contrast maps are uploaded under https://neurovault.org/collections/IABCOPVN/.
Note: positive ESA: hits > misses contrast, negative ESA: misses > hits contrast

Encoding success activity

The ESA contrast showed the greatest positive ESA (hits > misses) (global maximum) in the right hippocampus and the greatest negative ESA (misses > hits) (global minimum) in the right precuneus (Figure 4 c4). For auditory items, we found the maximum positive ESA in left auditory cortex and the maximum negative ESA in the right precuneus (Figure 4 c5). For visual items, we found the maximum positive ESA in the right visual cortex and the maximum negative ESA in the right precuneus (Figure 4 c6). See ESA in Table 2 for the MNI coordinates and values and Supplementary Table A.2 for all cluster specific peaks.

Together, these maps demonstrate that positive ESA in the auditory and visual cortices is sensory-specific while the positive ESA in hippocampus and the negative ESA in the precuneus are sensory-unspecific.

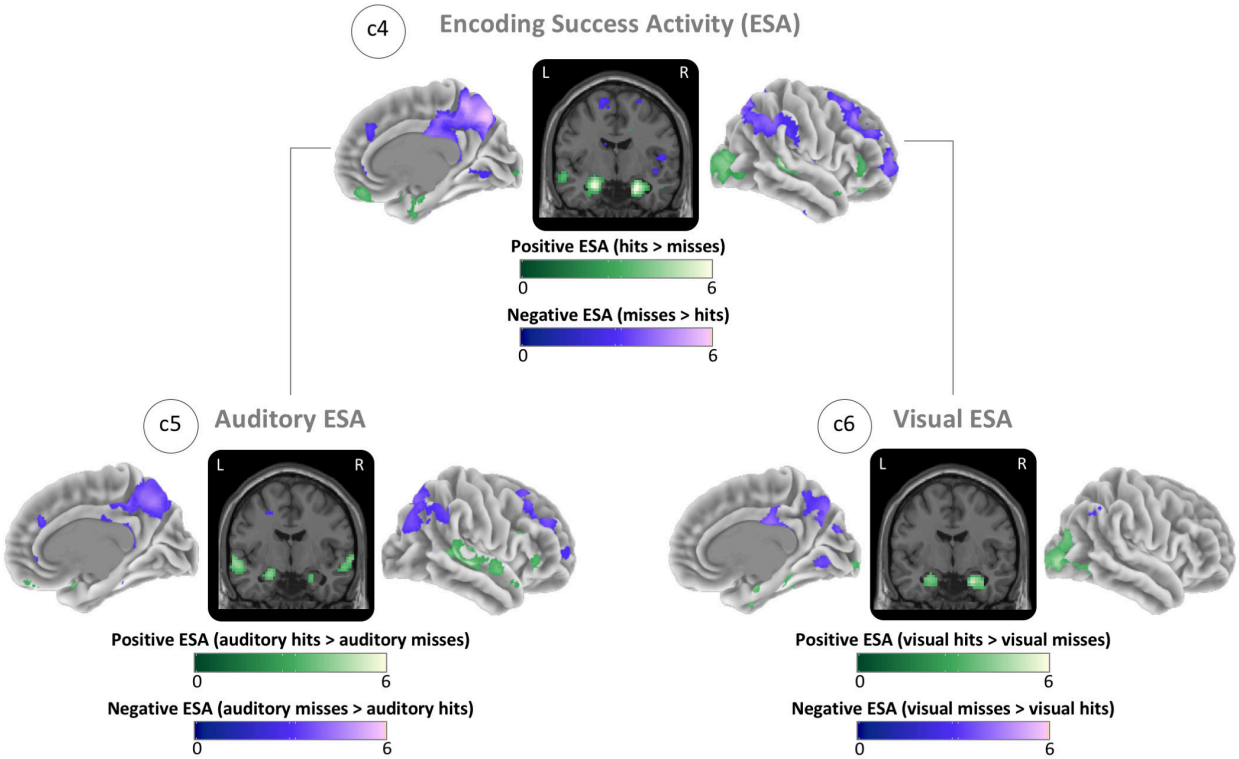

Figure 4. Brain maps of sensory-unspecific encoding success activity (ESA), assessed through the contrasts between activity of subsequently remembered (hit) and subsequently forgotten (miss) stimuli of the isolated encoding condition. (c4) ESA across all conditions (auditory and visual). (c5) ESA of auditory stimuli. (c6) ESA of visual stimuli. Brain activity is shown at a threshold of $p < .05$ [false discovery rate (FDR) corrected], and the color intensity shows the *t*-value. Contrast maps are uploaded under https://neurovault.org/collections/IABCOPVN/.

Sustained and transient activations

To clarify the patterns of sustained and transient activations, we mapped the block- and event-related activity vs. the rest condition for each sensory condition. For sustained (block-based) auditory activity, we found the global maxima in the right superior temporal lobe (Figure 5 c7). For the transient (event-based) auditory activity, we found the maxima in the left primary auditory cortex (Figure 5 c8). For sustained visual activity, we found the maxima in the primary visual cortex (Figure 5 c9). Finally, for transient visual activity, we found the maxima in the right fusiform gyrus (Figure 5 c10). For MNI coordinates and values, see sustained and transient section in Table 2; and for cluster specific peak activation, see Supplementary Table A.2. We also examined the local minima (deactivations) of the same contrasts for which the results and images can be found in the Supplementary Material "Sustained and transient deactivation".

In comparison with the results from the block-only or event-only models, mixed models show slightly different levels of activity in the regions of interest. The directionality and the appearance of the main effects stayed the same (see Supplementary Material "Model comparison between mixed, block- and event-only modeling").

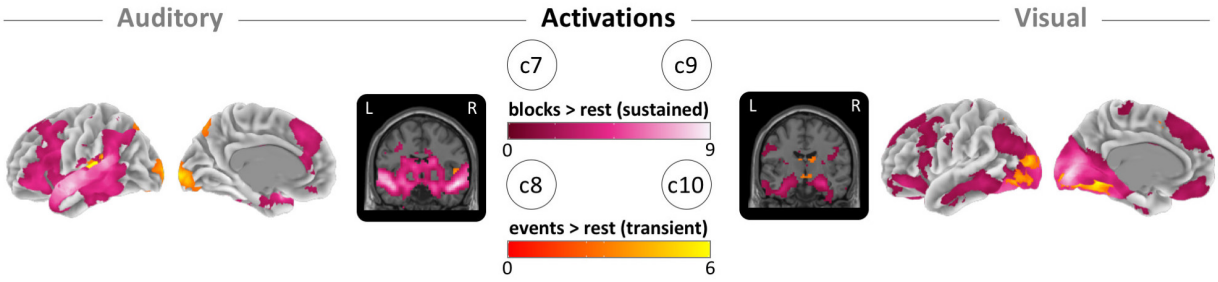

Figure 5. Brain maps of block-related (sustained) and event-related (transient) activations. (c7) Auditory block vs. rest activation (pink). (c8) Auditory event vs. rest activation (orange). (c9) Visual block vs. rest activation (pink). (c10) Visual block vs. rest activation (orange). Brain activity is shown at a threshold of $p < .05$ [false discovery rate (FDR) corrected], and the color intensity shows the *t*-value. Contrast maps are uploaded under https://neurovault.org/collections/IABCOPVN/.

Reliability analysis

ICCs for all calculated behavioral outcomes in total (hit-rate, FA-rate, and d-prime) and separated for sensory and stimulus conditions ranged between 0.400 and 0.812, with the highest ICC in FA-rates and the lowest in stimulus-specific *d′* and hit-rates. Overall, *d′* showed an ICC of 0.675. Sensory-specific *d′* showed ICCs of 0.622 for auditory and 0.649 for visual stimuli. More detailed results are described in Supplementary Material “Analysis of reliability”.

Smoothened voxel-wise ICC analysis for the fMRI data revealed for all contrasts stronger reliability for regions that showed also high activation. In c1, the visual and auditory cortices showed the highest reliability with a global peak of ICC = 0.663 in the left middle occipital gyrus ($MNI_{(x,y,z)}$: -12, -100, -1). In c2, ICCs are lower but still showed a global peak of ICC = 0.206 in the right auditory cortex (middle temporal gyrus, $MNI_{(x,y,z)}$: 57, -37, 5). In c3, ICC values had a similar range as in c1 with a global peak of ICC = 0.608 in the right fusiform gyrus ($MNI_{(x,y,z)}$: 33, -49, -10). C4 showed a maximum ICC of 0.306 in the right fusiform gyrus ($MNI_{(x,y,z)}$: 27, -82, -13). Other local peaks of ICC in c4 are found, for example, in the right hippocampus (ICC = 0.132, $MNI_{(x,y,z)}$: 15, -4, -16) and the left parahippocampal region (ICC = 0.230, $MNI_{(x,y,z)}$: -15, -4, -19). More detailed results are shown in the Supplementary Material “Reliability of sensory-specific and encoding success activity”.

DISCUSSION

We demonstrate the feasibility of a parallel mixed design as an efficient strategy for acquisition of rich fMRI data in limited time. The acquired data can give information about sensory-specific brain activation as well as sensory-specific and sensory-unspecific memory performance (taking the behavioral retrieval task data into account) using the key contrasts c1 to c6. The additional contrasts c7 to c10 show that also information on the difference of sustained (block-based) and transient (event-based) models and resulting activation can be obtained.

Behavior

In the retrieval task, participants showed a hit-rate close to 50 %, which is optimal for ESA modeling, as it ensures a balanced number of observations on each side of the contrast (hits vs. misses). Furthermore, FA-rates were very low, which was reflected in a conservative response bias and resulted in *d′* values far above chance for each stimulus category (environmental and vocal sound, and scene and face images). The *d′* values far above chance suggested that the ESA contrast is driven by memory encoding and not guessing. Although the hit-rates were all close to 50 %, we see slight differences between stimulus conditions (Table 1). Therefore, we cannot rule out completely that some stimulus conditions influenced the weighting and the ESA contrasts and caused small differences between the conditions. Due to the low number of stimuli and the between-subject variance, we did not have sufficient observations for reliable ESA in each separate stimulus category in this sample. If this task is applied in larger studies, there would be also interesting contrasts to examine. For now, we focused on examining auditory ESA, visual ESA, and (overall) ESA.

Already during task construction, we found visual memory performance to be superior for visual scenes compared with visual faces or sounds. This was despite our initial (design) objective to achieve balanced memory scores for each stimulus type. Given this objective, auditory retrieval was tested before visual retrieval, so the time delay between encoding and retrieval for visual stimuli was longer than for auditory stimuli. We enriched the auditory experience by using various speakers for the vocal sounds and the environmental sounds by presenting a large range of stimuli from animals to vehicles. Finally, we degraded the visual scenes slightly by desaturating the originally bright colors of the images.[44] Nevertheless, memory performance for the visual items remained superior, especially for the scenes (Table 1). This finding replicated behavioral work that indicated better memory performance for visual scenes than for any kind of auditory stimuli.[51] In general, visual stimuli are more often remembered and with more detail in comparison with auditory memory, if recalled immediately.[52,53] However, after a time delay, auditory memories are more stable than visual memory.[53] This may imply that it may be more difficult to encode auditory than visual stimuli. In addition, auditory stimuli were presented above the rhythmic scanner noise, whereas visual stimuli were presented in a dark and

visually “quiet” environment. Although we took care to select stimuli that were distinctly different from the scanning sounds with regard to spectral and temporal structure, we must consider the possibility that some acoustic masking occurred. The noisy scanning environment creates additional challenges for the auditory system on several levels of processing, in particular for the detection of signal in noise as well as asking for greater attentional demands and efforts. With regard to applying this paradigm in studies with a wider age range, it must be considered that complex listening skills (such as processing speech in noise) decline with age even when controlling for overall hearing thresholds. Also, considering the blood oxygenation level-dependent (BOLD) effect, it is likely that the continuous scanner noise resulted in continuously high activity in the auditory areas, rendering it more difficult to detect more subtle effects of condition on top of this saturation effect.[54] However, we decided against presenting auditory stimuli in quiet(er) pauses (between volumes), as such a sparse sampling paradigm would have significantly extended the scanning time. In future developments of such paradigms, and especially in studies including older adults, novel methodological approaches such as interleaved silent steady state or special scanning sequences that minimize acoustic impact could be considered [see methodological review by Peelle[55]].

Looking into differences between the two presentation conditions (isolation and parallel presentation), we also observed differences in memory performance. Isolated presentation resulted in better memory performance for both visual and auditory items. This is consistent with the model that memory encoding is limited by a working memory capacity[56] and that it is impaired if semantically incongruent information is presented in parallel.[52] One interpretation is that divided attention between auditory and visual information is detrimental to encoding. Note that we tested the auditory and visual retrieval separately, and the information of the two parallel presented stimulus classes was not congruent. There is a large body of evidence that multisensory encoding of congruent information is beneficial for memory performance[52,57] [for review, see Quak et al.[58]]. Therefore, depending on the scientific aim, one could adapt our paradigm and match voices with faces and scenes with environmental sounds. By doing so, memory performance is likely to improve at the expense of either the factorial design or a longer acquisition period.

Interestingly, the relative difference between isolated and parallel encoding was not the same for auditory and visual stimuli. We found that visual memory performance declined more under parallel conditions, while auditory encoding was less hindered. This has also been found in a working memory study on isolated and parallel retention of auditory (vocal) and visual (abstract objects) information.[59] Together with the finding from Gloede and Gregg[53] that visual memory is more hindered by a delayed recall than auditory memory, this suggests that auditory encoding might be more difficult but relatively robust. One explanation for the robustness of auditory encoding over the presentation conditions might also be related to the noisy scanner environment, creating continuously higher demands on auditory processing during both presentation conditions, as discussed above. This might have reduced the size of the effect of additional between-modality parallel processing for the auditory stimuli. A second explanation could be that learning auditory stimuli in similar detail as visual memory is more difficult and takes more attentional effort.[53] Therefore, parallel conditions that demand more attention influence auditory information less than visual memory.

As discussed above, *d′* values showed that the participant's memory scores were far above chance. Within the young adults, we did not find ceiling effects, and we did not find floor effects in the older adults (see Supplementary Material "Association between age and memory performance"). Together, this makes the task suitable for a life span study. We also assume that our task can be performed by participants with cognitive impairment and dementia. Sperling et al.[41] showed that mild Alzheimer's disease patients were able to do a face-name association task in which participants had to remember which name was associated with which face. In comparison, we had similar to even less instructions in our encoding task, and our retrieval task was easier, as we probed recognition memory, via old/new judgment, and not associative memory with previously seen lures. Further, for the recognition task, the questions and the answering options were shown on the screen for each trial. In conclusion, as long as participants are willing to be scanned for at least 10 minutes, the task should be applicable for people across all age ranges as well as people affected by neurodegenerative diseases.

Sensory-specific activity

Mapping of perceptive auditory and visual brain activity (Figure 3 c1) showed quite consistent results with previous work with activity for auditory conditions in environmental and vocal selective brain regions[26,28,29,32] and with activity for visual conditions in face- and scene-selective brain regions, i.e., the fusiform face and the parahippocampal place area.[21–23,32,60] Response strength differences for vocal vs. environmental sound stimuli are also consistent with previous work.[26,27,61] This imbalance could reflect either the properties of the auditory stimuli (spectral frequencies and temporal structure) or the organization of the auditory system. Although we found slightly stronger responses to scene stimuli, the contrast between face and scene stimuli was more balanced. These more similar levels of activity might also reflect either some property of the visual stimuli (i.e., similar discriminability) or the organization of the visual system.

From a design perspective, the relative imbalance in evoked auditory activity is suboptimal. We mostly used stimuli from previous experiments to replicate known activity patterns by using the parallel mixed design.[26,41,44] The application of two different stimulus conditions for visual and auditory senses allowed a detailed examination of the sensory cortices, and it enabled us to use the task also in people with possible or known problems in parts of the sensory cortices. If, for example, a person has face recognition dysfunction, data from the visual scene stimuli can still be used. As we aim to make this task suitable for large-scale population-based studies, which mostly examine people of different ages and health states, it can give usable and comparable information on a wide range of sensory and memory functions.

Encoding success activity

The ESA maps demonstrated that the auditory cortex and visual cortex showed sensory-specific ESA (Figure 4). In contrast, the hippocampus and a subset of default network structures —including the precuneus and angular gyrus— showed ESA across both sensory domains. These results suggest that sensory-independent, or multimodal, brain regions form a core memory network.[62–65] The precuneus showed negative ESA, consistent with previous findings on task-induced deactivation in the default network.[66–69] Especially negative ESA seems to be altered under the influence of early Alzheimer's

disease pathology.[70–73] Although hit-rates for visual stimuli and especially for scene stimuli were slightly higher and might have resulted in an unbalanced weighting (compare Discussion - Behavior section), the consistency with previous results demonstrates that small differences in the weighting did not influence results strongly. Therefore, our paradigm might be an efficient alternative for clinical and population studies that are interested in the functional responses of the memory system. Further, the encoding of both auditory and visual information allows investigators to disentangle factors that influence sensory-specific responses vs. alterations to the core memory system. One idea could be that age-related hearing loss is likely to affect auditory ESA, glaucoma in the retina is likely to affect visual ESA, and Alzheimer's pathology might target ESA in the core memory system. Our task is properly designed to disentangle these peripheral changes in sensory systems from alterations to the core memory network.

Sustained and transient (de)activations

The activation maps (task > rest) between sustained and transient activities show largely an overlap between activated regions indicating that sustained and transient activations co-occur simultaneously in sensory cortices.[37,38] However, deactivation maps (rest > task) show no overlap (Supplementary Figure A.2). Hence, we did not find any brain region —within or outside of the default network— that simultaneously showed sustained and transient deactivations. The lack of overlap between sustained and transient deactivations is not easily explained by overfitting or competition within the mixed model, as we did find overlap between sustained and transient activities. This is also confirmed by our comparisons of parallel mixed model with the block-only and event-only models (Supplementary Figure A.3). We also found that the majority of brain regions showed transient and not sustained deactivations. We interpret these findings in terms of task-intrusive and spontaneous thoughts.[74–76] Task-induced deactivations are modulated by task demands[77] consistent with transient deactivation in response to the task. This would mean that the higher the task demands, the more deactivation will occur. The relative lack of brain regions that showed sustained deactivation suggests that a stable pattern of reduced activity is very rare, whereas event modeling gives more information about deactivation. This finding is also consistent with other mixed design studies that suggested

mean activity is a relatively poor predictor of task performance[78] because it disregards differences between stimuli and easily overestimates outlier. Finally, it is also possible that task-induced deactivations are modulated by, but not very tightly coupled to, stimulus onset. This would hinder the separation of sustained and transient deactivation. This last explanation is consistent with spontaneous thoughts (e.g., on the task instruction or other distractions coming from the situation in the scanner) that are partially restricted by the cognitive demands but not tightly coupled to stimulus onset.

Reliability

As we did not have data to conduct a test-retest reliability across the complete task, we estimated the task reliability using the second task session as the retest session. ICC analysis for the behavioral overall and sensory-specific data according to Koo and Li[79] showed moderate-to-good reliability. Stimulus-specific ICCs were slightly lower. However, the two task sessions were not identical, as the order of blocks was different and new stimuli of the same conditions were presented. Therefore, we expect to have underestimated the actual ICC values, and we consider the obtained values to be quite plausible. The poor-to-moderate reliability values in the stimulus conditions confirmed our decision to exclude the separate stimulus conditions from ESA analyses. Overall, the moderate-to-good behavioral reliability permits the application of this task as an fMRI paradigm.

Voxel-wise ICC analysis for fMRI data smoothed on group-level shows in general lower ICC values than the behavioral data. These values are similar to those of other studies that reported low or very heterogeneous ICC values for fMRI tasks.[13,80–82] fMRI reliability is influenced by many factors, including scanner noise, physiological noise, cognitive factors and processes, sample size, sample characteristics, and task characteristics.[80,83] We observed in most contrasts high ICC values in regions showing high activity (especially visual and, more specifically, in scene-related areas). This is not completely unexpected, as high fMRI activity can reduce the influence of errors, which results in a decreased within-participant variation, which then leads to an increase in ICC values.[80,81] However, we observed also in some highly activated regions a fairly low ICC (e.g. hippocampus, precuneus, or superior temporal gyrus). This could be a property of the

regions showing a greater variability in the hemodynamic response. In this case, low ICC represents a low congruence between the two halves of activity of the task within participants, although activation was commonly observed over the complete task. Still, we consider our reliability estimates as quite realistic and conclude that even with the parallel presentation, the reliability in comparison with other fMRI studies is not reduced.

LIMITATIONS

Parallel mixed block/event-related fMRI designs measure transient and sustained information within a single, time-efficient paradigm that includes a large number of stimuli. However, this makes mixed designs very complex, and regressors are always affected by both events and blocks. Especially, events are always modeled on a changing baseline. This can make it difficult to interpret the comparison between transient and sustained effects.

Our study was sufficiently powered to calculate sensory-specific and overall ESA. However, we were not able to calculate stimulus-specific ESA within our sample. In the current study, we only included behavioral results from a small sample of older participants. Although their memory performance was slightly worse than that of participants in our younger sample, the hit-rates were close to 50 %. Therefore, we would expect this task to be also applicable in older populations as well.

Finally, we would like to mention that we were able to estimate low-to-moderate reliability using a split-half reliability. These estimates are comparable with those of other fMRI tasks, but reliability should be kept in mind especially when analyzing smaller sample sizes.[83–85]

CONCLUSIONS AND OUTLOOK

The presented parallel mixed design task paradigm enables efficient mapping of a versatile number of contrasts in limited time, making it attractive to acquire task-evoked fMRI in an epidemiological context for large-scale studies. The ability to map sensory activity as well as sensory-specific and unspecific ESA, and the ability to separate sustained and transient activities, can provide new insights into the dynamics of fMRI

across the life span.[38,86] Currently, it remains unclear how multiple modifiable factors like lifestyle, education, or blood pressure, together with non-modifiable factors, like APOE status or gender, determine the brain's functional responses over age and to pathology. Only large-scale population studies that possess sufficient power to dissociate these factors can provide answers to these questions.

Besides these relevant questions, we would like to encourage future studies using the task to evaluate between-scanner reliability and to explore intra-individual variability in more depth. As the task is easily applicable, it does not require special scanning parameters and shows quite strong activation in the relevant brain regions relating to the main functions, we would not assume a huge loss of information if the data were collected with different scanners. As mentioned, the task requires only ten minutes of fMRI scanning time. However, for generating ESA contrasts, an additional 10 to 19 minutes of (post-scan) recognition task needs to be performed. For many large-scale population studies, the limiting factor is the fMRI scanning time. The additional recognition task can be performed during other structural scans, which makes it user-friendly and cost-efficient. Previously, we demonstrated that the scan quality was not affected, or even benefitted, from performing a task inside the MRI.[8] In case no other scans are needed, it is also possible to perform the recognition task outside the scanner, as long as the delay is constant for all participants.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of data protection regulations. On the group-level we uploaded the T-maps of the contrasts described in the article under https://neurovault.org/collections/IABCOPVN/ (persistent identifier: https://identifiers.org/neurovault.collection:4413). All tools used for the post processing are open source and described in detail in section Materials and Methods - Behavioral analysis and Materials and Methods - Functional MRI preprocessing. Task scripts can be assessed under https://www.rheinland-studie.de/data-code/boenniger2020. Requests to access the datasets should be directed to the Rhineland Study's Data Use and Access committee, Prof. Dr. Monique M. B. Breteler, RS-DUAC@dzne.de.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the medical ethics committee of the Medical Faculty of the University of Bonn. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MMB: conceptualization, methodology, software, formal analysis, investigation, writing–original draft, and visualization. KD: formal analysis and writing–review and editing. SH: methodology and writing–review and editing. MS: software and writing–review and editing. TS: software, validation, resources, and writing–review and editing. MMBB: conceptualization, resources, data curation, writing–review and editing, supervision, and funding acquisition. WH: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing–review and editing, visualization, supervision, and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by DZNE institutional funds.

ACKNOWLEDGEMENTS

We thank our study technicians Kerstin Micus, Sarah Jansen, and Jarmila Schmitz for their support in acquiring the data. Further, we thank Fabienne Fox and Santiago Estrada for their feedback on the pre-final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at:

https://www.frontiersin.org/articles/10.3389/fnhum.2020.591721/full#supplementary-material

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

1. Hofman A, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*. 2015;30(8):661-708. doi:10.1007/s10654-015-0082-x

2. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523-1536. doi:10.1038/nn.4393

3. Breteler MMB, Stöcker T, Pracht E, Brenner D, Stirnberg R. MRI in the Rhineland Study: A novel protocol for population neuroimaging. *Alzheimer's Dement*. 2014;10(4):92. doi:10.1016/j.jalz.2014.05.172

4. Damoiseaux JS, Huijbers W. The complementary value of task-evoked and resting-state functional imaging: a commentary on Campbell and Schacter (2016). *Lang Cogn Neurosci*. 2017;32(6):678-680. doi:10.1080/23273798.2016.1248452

5. Campbell KL, Schacter DL. Ageing and the resting state: is cognition obsolete? *Lang Cogn Neurosci*. 2017;32(6):661-668. doi:10.1080/23273798.2016.1227858

6. Davis SW, Stanley ML, Moscovitch M, Cabeza R. Resting-state networks do not determine cognitive function networks: a commentary on Campbell and Schacter (2016). *Lang Cogn Neurosci*. 2017;32(6):669-673. doi:10.1080/23273798.2016.1252847

7. Vanderwal T, Kelly C, Eilbott J, Mayes LC, Castellanos FX. Inscapes: a movie paradigm to improve compliance in functional magnetic resonance imaging. *Neuroimage*. 2015;122:222-232. doi:10.1016/j.neuroimage.2015.07.069

8. Huijbers W, Van Dijk KRA, Boenniger MM, Stirnberg R, Breteler MMB. Less head motion during MRI under task than resting-state conditions. *Neuroimage*. 2017;147:111-120. doi:10.1016/j.neuroimage.2016.12.002

9. Pinel P, Thirion B, Meriaux S, et al. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci*. 2007;8(1):91. doi:10.1186/1471-2202-8-91

10. Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27(4):685-691. doi:10.1002/jmri.21049

11. Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage*. 2013;80:105-124. doi:10.1016/j.neuroimage.2013.04.127

12. Sperling R. Functional MRI studies of associative encoding in normal aging, mild cognitive impairment, and Alzheimer's Disease. *Ann N Y Acad Sci*. 2007;1097(1):146-155. doi:10.1196/annals.1379.009

13. Barch DM, Burgess GC, Harms MP, et al. Function in the Human Connectome: Task-fMRI and individual differences in behavior. *Neuroimage*. 2013;80:169-189. doi:10.1016/j.neuroimage.2013.05.033

14. Nenert R, Allendorfer JB, Szaflarski JP. A model for visual memory encoding. *PLoS One*. 2014;9(10):e107761. doi:10.1371/journal.pone.0107761

15. Sidhu MK, Stretton J, Winston GP, et al. Memory fMRI predicts verbal memory decline after anterior temporal lobe resection. *Neurology*. 2015;84(15):1512-1519. doi:10.1212/WNL.0000000000001461

16. Hayes JM, Tang L, Viviano RP, van Rooden S, Ofen N, Damoiseaux JS. Subjective memory complaints are associated with brain activation supporting successful memory encoding. *Neurobiol Aging*. 2017;60:71-80. doi:10.1016/j.neurobiolaging.2017.08.015

17. Wheeler ME, Petersen SE, Buckner RL. Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci*. 2000;97(20):11125-11129. doi:10.1073/pnas.97.20.11125

18. Daselaar SM, Porat Y, Huijbers W, Pennartz CMA. Modality-specific and modality-independent components of the human imagery system. *Neuroimage*. 2010;52(2):677-685. doi:10.1016/j.neuroimage.2010.04.239

19. Langner R, Kellermann T, Eickhoff SB, et al. Staying responsive to the world: Modality-specific and -nonspecific contributions to speeded auditory, tactile, and visual stimulus detection. *Hum Brain Mapp*. 2012;33(2):398-418. doi:10.1002/hbm.21220

20. Gusnard DA, Raichle ME. Searching for a baseline: Functional imaging and the resting human brain. *Nat Rev Neurosci*. 2001;2(10):685-694. doi:10.1038/35094500

21. Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 1997;17(11):4302-4311. doi:10.1523/JNEUROSCI.17-11-04302.1997

22. Epstein R, Kanwisher N. A cortical representation of the local visual environment. *Nature*. 1998;392(6676):598-601. doi:10.1038/33402

23. Gazzaley A, Cooney JW, Rissman J, D'Esposito M. Top-down suppression deficit underlies working memory impairment in normal aging. *Nat Neurosci*. 2005;8(10):1298-1300. doi:10.1038/nn1543

24. Collins JA, Dickerson BC. Functional connectivity in category-selective brain networks after encoding predicts subsequent memory. *Hippocampus*. 2019;29(5):440-450. doi:10.1002/hipo.23003

25. Srokova S, Hill PF, Koen JD, King DR, Rugg MD. Neural differentiation is moderated by age in scene-selective, but not face-selective, cortical regions. *eNeuro*. 2020;7(3):ENEURO.0142-20.2020. doi:10.1523/ENEURO.0142-20.2020

26. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature*. 2000;403:309-312. doi:10.1038/35002078

27. Belin P, Zatorre RJ, Ahad P. Human temporal-lobe response to vocal sounds. *Cogn Brain Res*. 2002;13(1):17-26. doi:10.1016/S0926-6410(01)00084-2

28. Pernet CR, McAleer P, Latinus M, et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage*. 2015;119:164-174. doi:10.1016/j.neuroimage.2015.06.050

29. Agus TR, Paquette S, Suied C, Pressnitzer D, Belin P. Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Sci Rep*. 2017;7(1):1-7. doi:10.1038/s41598-017-11684-1

30. Zäske R, Awwad Shiekh Hasan B, Belin P. It doesn't matter what you say: FMRI correlates of voice learning and recognition independent of speech content. *Cortex*. 2017;94:100-112. doi:10.1016/j.cortex.2017.06.005

31. Aglieri V, Chaminade T, Takerkart S, Belin P. Functional connectivity within the voice perception network and its behavioural relevance. *Neuroimage*. 2018;183:356-365. doi:10.1016/j.neuroimage.2018.08.011

32. Young AW, Frühholz S, Schweinberger SR. Face and voice perception: understanding commonalities and differences. *Trends Cogn Sci*. 2020;24(5):398-410. doi:10.1016/j.tics.2020.02.001

33. Wagner AD, Schacter DL, Rotte M, et al. Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*. 1998;281(5380):1188-1191. doi:10.1126/science.281.5380.1188

34. Otten LJ, Rugg MD. When more means less: neural activity related to unsuccessful memory encoding. *Curr Biol*. 2001;11(19):1528-1530. doi:10.1016/S0960-9822(01)00454-7

35. Prince SE, Dennis NA, Cabeza R. Encoding and retrieving faces and places: Distinguishing process- and stimulus-specific differences in brain activity. *Neuropsychologia*. 2009;47(11):2282-2289. doi:10.1016/j.neuropsychologia.2009.01.021

36. Velanova K, Jacoby LL, Wheeler ME, McAvoy MP, Petersen SE, Buckner RL. Functional-anatomic correlates of sustained and transient processing components engaged during controlled retrieval. *J Neurosci*. 2003;23(24):8460-8470. doi:10.1523/JNEUROSCI.23-24-08460.2003

37. Visscher KM, Miezin FM, Kelly JE, et al. Mixed blocked/event-related designs separate transient and sustained activity in fMRI. *Neuroimage*. 2003;19(4):1694-1708. doi:10.1016/S1053-8119(03)00178-2

38. Petersen SE, Dubis JW. The mixed block/event-related design. *Neuroimage*. 2012;62(2):1177-1184. doi:10.1016/j.neuroimage.2011.09.084

39. Huijbers W, Pennartz CMA, Rubin DC, Daselaar SM. Imagery and retrieval of auditory and visual information: Neural correlates of successful and unsuccessful performance. *Neuropsychologia*. 2011;49(7):1730-1740. doi:10.1016/j.neuropsychologia.2011.02.051

40. Parsons CE, Young KS, Craske MG, Stein AL, Kringelbach ML. Introducing the Oxford Vocal (OxVoc) Sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Front Psychol*. 2014;5:562. doi:10.3389/fpsyg.2014.00562

41. Sperling RA, Bates JF, Chua EF, et al. fMRI studies of associative encoding in young and elderly controls and mild Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2003;74(1):44-50. doi:10.1136/jnnp.74.1.44

42. Minear M, Park DC. A lifespan database of adult facial stimuli. *Behav Res Methods, Instruments, Comput*. 2004;36(4):630-633. doi:10.3758/BF03206543

43. Huijbers W, Mormino EC, Schultz AP, et al. Amyloid-β deposition in mild cognitive impairment is associated with increased hippocampal activity, atrophy and clinical progression. *Brain*. 2015;138(4):1023-1035. doi:10.1093/brain/awv007

44. Huijbers W, Pennartz CM, Cabeza R, Daselaar SM. When learning and remembering compete: A functional MRI study. *PLoS Biol*. 2009;7(1):e1000011. doi:10.1371/journal.pbio.1000011

45. Peirce JW. PsychoPy-psychophysics software in Python. *J Neurosci Methods*. 2007;162(1-2):8-13. doi:10.1016/j.jneumeth.2006.11.017

46. Yarkoni T, Barch DM, Gray JR, Conturo TE, Braver TS. BOLD correlates of trial-by-trial reaction time variability in gray and white matter: A multi-study fMRI analysis. *PLoS One*. 2009;4(1):e4257. doi:10.1371/journal.pone.0004257

47. Viswanathan S, Abdollahi RO, Wang BA, Grefkes C, Fink GR, Daun S. A response-locking protocol to boost sensitivity for fMRI-based neurochronometry. *Hum Brain Mapp*. 2020;41(12):3420-3438. doi:10.1002/hbm.25026

48. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. doi:10.1037/0033-2909.86.2.420

49. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46. doi:10.1037//1082-989x.1.4.390

50. Andrade A, Paradis A-L, Rouquette S, Poline J-B. Ambiguous results in functional neuroimaging data analysis due to covariate correlation. *Neuroimage*. 1999;10(4):483-486. doi:10.1006/nimg.1999.0479

51. Cohen MA, Horowitz TS, Wolfe JM. Auditory recognition memory is inferior to visual recognition memory. *Proc Natl Acad Sci*. 2009;106(14):6008-6010. doi:10.1073/pnas.0811884106

52. Thelen A, Talsma D, Murray MM. Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*. 2015;138:148-160. doi:10.1016/j.cognition.2015.02.003

53. Gloede ME, Gregg MK. The fidelity of visual and auditory memory. *Psychon Bull Rev*. 2019;26(4):1325-1332. doi:10.3758/s13423-019-01597-7

54. Tomasi D, Caparelli EC, Chang L, Ernst T. fMRI-acoustic noise alters brain activation during working memory tasks. *Neuroimage*. 2005;27(2):377-386. doi:10.1016/j.neuroimage.2005.04.010

55. Peelle JE. Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Front Neurosci*. 2014;8:1-13. doi:10.3389/fnins.2014.00253

56. Baddeley A. Working memory: looking back and looking forward. *Nat Rev Neurosci*. 2003;4:829-839. doi:10.1038/nrn1201

57. Shams L, Seitz AR. Benefits of multisensory learning. *Trends Cogn Sci*. 2008;12(11):411-417. doi:10.1016/j.tics.2008.07.006

58. Quak M, London RE, Talsma D. A multisensory perspective of working memory. *Front Hum Neurosci*. 2015;9(APR):1-11. doi:10.3389/fnhum.2015.00197

59. Saults JS, Cowan N. A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *J Exp Psycholology Gen*. 2007;136(4):663-684. doi:10.1037/0096-3445.136.4.663

60. Collins JA, Olson IR. Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*. 2014;61(1):65-79. doi:10.1016/j.neuropsychologia.2014.06.005

61. Mostafa MM. Brain processing of vocal sounds in advertising: A functional magnetic resonance imaging (fMRI) study. *Expert Syst Appl*. 2012;39(15):12114-12122. doi:10.1016/j.eswa.2012.04.003

62. Johnson JD, Rugg MD. Recollection and the Reinstatement of Encoding-Related Cortical Activity. *Cereb Cortex*. 2007;17(11):2507-2515. doi:10.1093/cercor/bhl156

63. Kim H, Daselaar SM, Cabeza R. Overlapping brain activity between episodic memory encoding and retrieval: Roles of the task-positive and task-negative networks. *Neuroimage*. 2010;49(1):1045-1054. doi:10.1016/j.neuroimage.2009.07.058

64. Gilmore AW, Nelson SM, McDermott KB. A parietal memory network revealed by multiple MRI methods. *Trends Cogn Sci*. 2015;19(9):534-543. doi:10.1016/j.tics.2015.07.004

65. Kim H. Neural activity during working memory encoding, maintenance, and retrieval: A network-based model and meta-analysis. *Hum Brain Mapp*. 2019;40(17):4912-4933. doi:10.1002/hbm.24747

66. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *PNAS*. 2001;98(2):676-682. doi:10.1073/pnas.98.2.676

67. Daselaar SM, Prince SE, Cabeza R. When less means more: deactivations during encoding that predict subsequent memory. *Neuroimage*. 2004;23(3):921-927. doi:10.1016/j.neuroimage.2004.07.031

68. Huijbers W, Schultz AP, Vannini P, et al. The encoding/retrieval flip: Interactions between memory performance and memory stage and relationship to intrinsic cortical networks. *J Cogn Neurosci*. 2013;25(7):1163-1179. doi:10.1162/jocn_a_00366

69. Krieger-Redwood K, Jefferies E, Karapanagiotidis T, et al. Down but not out in posterior cingulate cortex: Deactivation yet functional coupling with prefrontal cortex during demanding semantic cognition. *Neuroimage*. 2016;141:366-377. doi:10.1016/j.neuroimage.2016.07.060

70. Sperling RA, Dickerson BC, Pihlajamaki M, et al. Functional alterations in memory networks in early Alzheimer's Disease. *Neuromolecular Med*. 2010;12(1):27-43. doi:10.1007/s12017-009-8109-7

71. Ewers M, Sperling RA, Klunk WE, Weiner MW, Hampel H. Neuroimaging markers for the prediction and early diagnosis of Alzheimer's disease dementia. *Trends Neurosci*. 2011;34(8):430-442. doi:10.1016/j.tins.2011.05.005

72. Jagust W. Vulnerable neural systems and the borderland of brain aging and neurodegeneration. *Neuron*. 2013;77(2):219-234. doi:10.1016/j.neuron.2013.01.002

73. Fu L, Maes JHR, Kessels RPC, Huijbers W. The neural basis of individual differences in memory performance in young and older adults: Using the encoding/retrieval flip account as framework. *Neurobiol Learn Mem*. 2020;173(107251). doi:10.1016/j.nlm.2020.107251

74. Weissman DH, Roberts KC, Visscher KM, Woldorff MG. The neural bases of momentary lapses in attention. *Nat Neurosci*. 2006;9(7):971-978. doi:10.1038/nn1727

75. Andrews-Hanna JR, Reidler JS, Huang C, Buckner RL. Evidence for the Default Network's Role in Spontaneous Cognition. *J Neurophysiol*. 2010;104(1):322-335. doi:10.1152/jn.00830.2009

76. Christoff K, Irving ZC, Fox KCR, Spreng RN, Andrews-Hanna JR. Mind-wandering as spontaneous thought: a dynamic framework. *Nat Rev Neurosci*. 2016;17(11):718-731. doi:10.1038/nrn.2016.113

77. McKiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR. A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J Cogn Neurosci*. 2003;15(3):394-408. doi:10.1162/089892903321593117

78. Garrett DD, McIntosh AR, Grady CL. Brain signal variability is parametrically modifiable. *Cereb Cortex*. 2014;24(11):2931-2940. doi:10.1093/cercor/bht150

79. Koo TK, Li MY. Cracking the code: Providing insight into the fundamentals of research and evidence-based practice - A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012

80. Bennett CM, Miller MB. How reliable are the results from functional magnetic resonance imaging? *Ann N Y Acad Sci*. 2010;1191(1):133-155. doi:10.1111/j.1749-6632.2010.05446.x

81. Brandt DJ, Sommer J, Krach S, et al. Test-retest reliability of fMRI brain activity during memory encoding. *Front Psychiatry*. 2013;4(163):1-9. doi:10.3389/fpsyt.2013.00163

82. Elliott ML, Knodt AR, Ireland D, et al. What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychol Sci*. 2020;31(7):792-806. doi:10.1177/0956797620916786

83. Noble S, Scheinost D, Constable RT. A guide to the measurement and interpretation of fMRI test-retest reliability. 2020;(October 29). doi:10.31219/osf.io/w3qkf

84. Turner BO, Paul EJ, Miller MB, Barbey AK. Small sample sizes reduce the replicability of task-based fMRI studies. *Commun Biol*. 2018;1(1):62. doi:10.1038/s42003-018-0073-z

85. Bossier H, Roels SP, Seurinck R, et al. The empirical replicability of task-based fMRI as a function of sample size. *Neuroimage*. 2020;212:1-12. doi:10.1016/j.neuroimage.2020.116601

86. Jimura K, Braver TS. Age-related shifts in brain activity dynamics during task switching. *Cereb Cortex*. 2010;20(6):1420-1431. doi:10.1093/cercor/bhp206

3.3 Cognitive examinations in the Rhineland Study

TEST SELECTION, DESCRIPTION, AND IMPLEMENTATION

Cognitive functions are subdivided into a large number of domains.[1] Performance in most of these domains typically decreases with advancing age.[2] In the Rhineland Study, we therefore noted *episodic verbal memory*, *working memory*, *executive functions*, and *processing speed* as the domains in which declines in performance over age were expected. When used in conjunction with age-standardized norms, performance in the domains assist in identifying and marking the progression of neurodegeneration. We also included a measure of *crystallized intelligence*, performance in which is resilient even in the presence of neurodegeneration, to estimate the participants premorbid cognitive ability.[2]

Several factors influenced neuropsychological test selection in the Rhineland Study. We selected tests that are suited for use in prospective cohort studies, given short examination times, rich data yields, and applicability to a wide age range. Moreover, we included tests that relied upon different sensory stimuli and motor responses to maximize the inclusion of participants with impairments. To maintain motivation and reduce possible feelings of failure, we also prioritized tasks that used adaptive testing, whereby participants do not know the maximum possible performance. We preferred tests that are well-studied and with published psychometric validity, so that our measurements would be comparable to other studies and psychometrically sound. Finally, where possible, we measured each domain with multiple tests so that composite domain scores were not affected by specific task-related factors unrelated to that domain.

In the final cognitive test battery, we included eight cognitive tests (Table 1). To assess episodic verbal memory, we used sum of trials 1 to 5 and the delayed recall performance of our version of *Rey's Auditory Verbal Learning Test (AVLT).*[3,4] Working memory was assessed using forward and backward span of a *digit span task* and a *Corsi block-tapping test.*[5] Executive function was assessed using time to complete a *Trail-making Test* (*TMT*) *B*, number of words produced in a *word fluency task* (animals), and error rate in an *eye-tracking antisaccade task.*[6] Processing speed was assessed using time to complete a *TMT A* and mean saccade latency of an *eye-tracking prosaccade task.*[6] Finally, number

of correctly recognized words in the *Mehrfachwahl-Wortschatz Intelligenztest* (*MWT-B; Multiple choice vocabulary test*)[7] was included to assess crystallized intelligence. All tests are commonly used, except for those involving eye-tracking. We included eye-tracking tasks as they require little instructions and are culturally independent (e.g. they do not rely upon language knowledge or culture-specific education).[8] In addition, control of eye movements is impaired in several neurodegenerative diseases.[9]

Table 1. Cognitive test scores included in the composite domain scores.

Cognitive domain	**Test scores**	**Range**
Episodic verbal memory	AVLT – sum trials 1 to 5 (sum of recalls 1 to 5)	0-75 words
	AVLT – time delayed recall (No. of recall 7)	0-15 words
Working memory	Digit span forward (max. span)	3-9 digits
	Digit span backward (max. span)	2-9 digits
	Corsi block-tapping span forward (max. span)	2-9 blocks
	Corsi block-tapping span backward (max. span)	2-9 blocks
Executive function	Word fluency (No. of distinct animals named in one minute)	0-∞ words
	TMT B (time to completion)	0-301 sec
	Error rate of antisaccade task	0-100 %
Processing speed	TMT A (time to completion)	0-301 sec
	Mean saccade latency of prosaccade task	0-∞ ms
Crystallized intelligence	MWT-B (No. of correctly recognized words)	0-37 words

Abbreviations: sec: seconds; AVLT: Rey's Auditory Verbal Learning Test; TMT: Trail-making Test; MWT-B: Mehrfachwahl-Wortschatz Intelligenztest/Multiple choice vocabulary test

All tests were adapted for computer-based testing to reduce manual post-processing and attain data on test performance, duration, and error types. Examinations are administered by trained and certified study technicians using standardized operating procedures. Responses to self-administered tests were recorded using a touchscreen, and supervised

training prior to the main task ensured proper understanding. Interview-based responses were recorded by study assistants and immediately entered into the database.

Participants began the cognitive assessment phase with the AVLT. As it is interview-based (details described in the introduction and Chapter 3.1), study assistants recorded the words participants recalled using a computer with word-completion aids to increase writing speed and correctness. In the period leading up to the delayed recall of the AVLT (20-30 minutes), participants completed non-verbal tasks (eye-tracker, Corsi block-tapping test, TMT).

The protocol for the eye-tracker tasks is described in detail in Coors et al. (2021).[6] Briefly, the prosaccade task is a standard horizontal “step” task in which participants are asked to follow a stimulus as closely as possible with their eyes (no head motion). Each trial involved the stimulus jumping from the center of the monitor to either the right or the left. In the antisaccade task, participants are instructed to look at the stimulus when it is in the central position, but when it jumps to the periphery, they have to look at the opposite position (mirror image).

The Corsi block-tapping test[5] is self-administered on the tablet. Nine blocks are displayed in the tablet and a sequence is given by blocks changing color one at a time. Immediately following the completion of a sequence, participants attempt to repeat the sequence by tapping the blocks in the same order (forward condition) or in the reverse order (backward condition). Should participants perform one of two trials correctly, the length of sequence in the next trial increases. The test ends when an error is made in the sequence of a given length twice.

The TMT (versions A and B) is also self-administered on a touchscreen computer, similar to the version in the PEBL test battery.[10] Participants connect 24 randomly scattered digits (TMT A) or 24 randomly scattered digits and letters (TMT B) as quickly as possible by tapping them in order. Items are connected by lines when a correct digit or letter is tapped, analogous to the original paper-pencil test.[11] If participants do not complete a TMT trial within 301 seconds, the test ends.

Participants also self-administer the MWT-B.[7] To ensure task comprehension, the study assistant observes performance on the first trial and participants perform the rest of the test independently. On each trial, five words (one existing German word and four fictional

words) are presented side by side, and after participants are asked to select the sole real word. When participants select a word or skip the trial, the next row of words appears.

In the animal word fluency task, participants name as many animals as possible within one minute. Answers are audio-recorded for later manual evaluation.

Finally, participants perform a digit span task in which participants are read numerical sequences with an increasing number of digits by study assistants. Participants must then repeat the sequence in the same order (forward task) or in the reverse order (backward task). The task ends if a sequence is not recalled correctly twice.

DATA CLEANING

Using R[12] and RStudio[13], test performance is first visually inspected by the responsible scientist and interpreted with regard to notes on test administration recorded by study assistants (in cases of deviations from standard operating procedures or participant behavior possibly impacting performance). Next, data are checked with respect to Rhineland Study sample performance using an outlier labeling rule (interquartile range multiplier factor of 2.2). Given the age-related nature of cognitive test performance, the outlier labeling rule is applied separately within 10-year age bands. Identified outliers are removed only if further examination of study assistant comments and other cognitive examination results strongly suggest a technical or assessment error.

For the Corsi block-tapping test, performance data were removed if a participant recorded a span less than two, given unlikely score reliability. In these instances, technical issues or problems in the handling of the tablet are assumed. Similarly, data were also removed from the Corsi block-tapping forward test when a participant's backward span was more than four, but their forward span was only two.

For reliability reasons, performance data were removed from eye-tracking prosaccade or antisaccade if a participant recorded fewer than seven valid trials. For the antisaccade task, data were additionally removed if participants had more than four errors and zero corrections, as we assume that the task was not correctly understood.

DATA POST-PROCESSING

R[12] and RStudio[13] were also used for post-processing of the raw test results, but most single test results did not require detailed post-processing. For the digit span and the Corsi block-tapping test, the longest sequence correctly recalled represented each participant's span number and was used as the main outcomes of those tasks. For MWT-B, the sum of all trials correctly answered was the main outcome. For TMT A and B, the time to completion was used as main outcomes.

Manual post processing was necessary in the word fluency test and the AVLT. For the word fluency test, we first created two transcriptions of each audio-file. Manual review was undertaken in cases where the two transcriptions did not match exactly. Typing errors were corrected using a table that matched each error to the correct animal/word. Using another table, the responses were classified as animals or errors. These two tables are updated with each word that is not listed in either the correction table or the classification table. For its main outcome, the total number of correct animals was used.

The AVLT used a similar system like that implemented for the word fluency test. There was a list of the learned words and a table in which all remaining words that participants named were listed. Each new word was checked manually, typing errors were corrected, and the words were added to the table. Then, a number of variables were created which are described in detail in Chapter 3.1.

The analysis of eye-tracking data required significant post-processing, which is described in detail elsewhere (Coors et al., 2021).[6]

To summarize the single cognitive test scores as domain scores (Table 1, Figure 1), we first adjusted cognitive tests with skewed distributions (TMT A, TMT B, and mean latency of the prosaccade task) using $\log_{10}$ transformations. Second, we reversed TMT A and B time, mean latency of the prosaccade, and the error rate of the antisaccade task, so that higher values represented better performance for all cognitive tests. Third, we z-standardized each cognitive test variable against the sample mean and standard deviation. Participants who did not report German as their first language or who were severely cognitively impaired due to traumatic brain injury or other non-age-related disease did not contribute to the mean and standard deviation statistics. However, z-

scores were still calculated for these participants based on the mean and standard deviation of the contributing sample.

For the fluid cognitive domains (executive function, processing speed, working memory and episodic verbal memory), individual contributing standardized scores were averaged to obtain composite domain z-scores. Crystallized intelligence was represented only by the MWT-B z-score (Table 1, Figure 1). Higher-order composite scores for total memory and global cognitive functions were further averaged from the contributing domain scores. Participants had to have valid scores on at least 50 % of the individual tests that contributed to a domain for a composite domain score to be calculated. For the higher order composite scores (total memory, global cognitive function), participants were required to have composite domain scores for each included fluid domain.

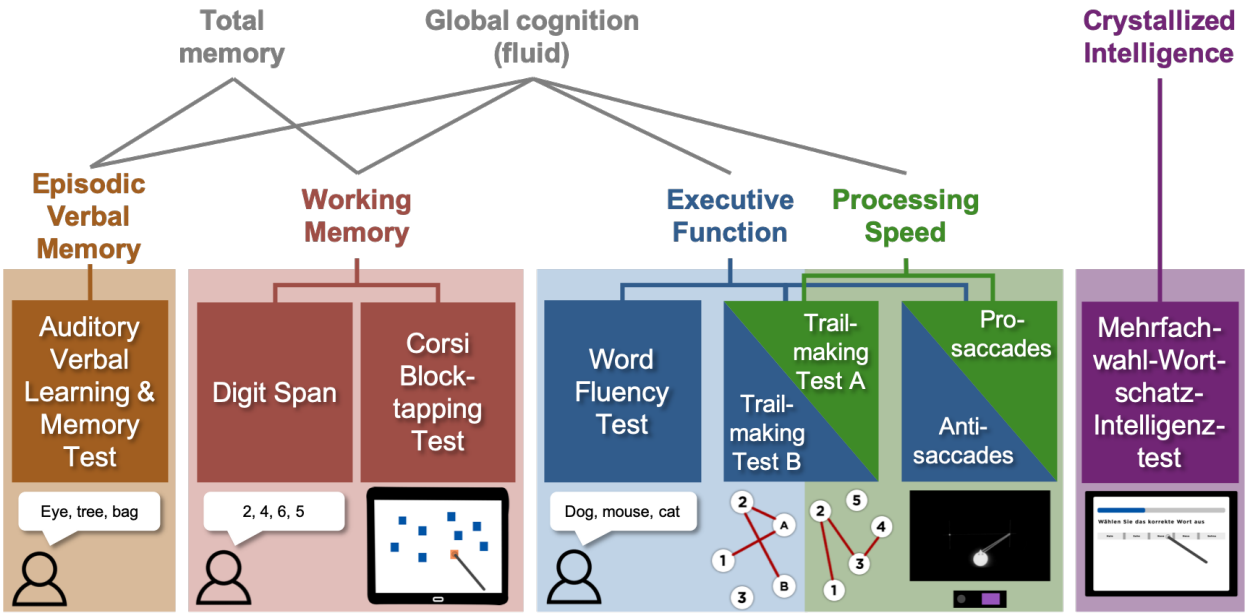

Figure 1. Cognitive test battery and cognitive domain scores of the Rhineland Study. Mehrfachwahl-Wortschatz Intelligenztest: Multiple choice vocabulary test.[a]

[a] Figure originally created by Natascha Merten and Meta Miriam Bönniger, first published in Merten, N. The Auditory System and Its Relation to Cognitive Function in the Process of Aging [PhD thesis]. Bonn, Germany: Faculty of Medicine University of Bonn; 2019. Further modified by Meta Miriam Bönniger.

REFERENCES

1. Harvey PD. Domains of cognition and their assessment. *Dialogues Clin Neurosci*. 2019;21(3):227-237. doi:10.31887/DCNS.2019.21.3/pharvey
2. Deary IJ, Corley J, Gow AJ, et al. Age-associated cognitive decline. *Br Med Bull*. 2009;92(1):135-152. doi:10.1093/bmb/ldp033
3. Rey A. Mémorisation d'une série de 15 mots en 5 répetitions. In: Rey A, ed. *L'examen Clinique En Psychologie*. 3rd ed. Presses Universitaries de France; 1970:141-193.
4. Taylor EM. 8. Learning and Memory. In: Taylor EM, ed. *Psychological Appraisal of Children with Cerebral Defects*. 3rd ed. Harvard University Press; 1968:417-437.
5. Corsi PM. Human memory and the medial temporal region of the brain. Published online 1972.
6. Coors A, Merten N, Ward DD, Schmid M, Breteler MMB, Ettinger U. Strong age but weak sex effects in eye movement performance in the general adult population: Evidence from the Rhineland Study. *Vision Res*. 2021;178:124-133. doi:10.1016/j.visres.2020.10.004
7. Lehrl S. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. Spitta Verlag GmbH & Co. KG; 2005.
8. Ettinger, U., Klein C. Eye Movements. In: Reuter M, Montag C, eds. *Neuroeconomics*. Springer-Verlag; 2016:481-502.
9. Anderson TJ, MacAskill MR. Eye movements in patients with neurodegenerative disorders. *Nat Rev Neurol*. 2013;9(2):74-85. doi:10.1038/nrneurol.2012.273
10. Mueller ST, Piper BJ. The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *J Neurosci Methods*. 2014;222:250-259. doi:10.1016/j.jneumeth.2013.10.024
11. Reitan RM. *Trail Making Test Manual for Administration and Scoring*. Reitan Neuropsychological Laboratory; 1992.
12. R Core Team. R: A language and environment for statistical computing. Published online 2019. https://www.r-project.org/

13. RStudio Team. RStudio: Integrated development for R. Published online 2019. http://www.rstudio.com/

4 Potential biomarkers and risk factors of neurodegeneration and cognitive functions

4.1 Association of retinal layer measurements and adult cognitive function - A population-based study

David D. Ward[1], Matthias M. Mauschitz[1,2], Meta M. Bönniger[1], Natascha Merten[1,4], Robert P. Finger[2], and Monique M.B. Breteler[1,3]

[1]Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

[2]Department of Ophthalmology, Faculty of Medicine, University of Bonn, Germany

[3]Institute for Medical Biometry, Informatics and Epidemiology, Faculty of Medicine, University of Bonn, Germany

[4]Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin **-** Madison, USA

Journal: Neurology 2020; 95:31144-e1152,
doi:10.1212/WNL.0000000000010146

ABSTRACT

Objective: To quantify the associations of peripapillary retinal nerve fiber layer (pRNFL) thickness and macular ganglion cell layer (mGCL) volume with cognitive functioning and to investigate how demographic and vascular health factors affect these associations in a population-based sample of adults.

Methods: The sample included the first 3,000 participants (age range 30-95 years) of the Rhineland Study (recruited from March 2016 to December 2018) who underwent spectral-domain optical coherence tomography and cognitive assessment at one of two identical study centers in Bonn, Germany. We used multiple linear regression models to examine the relationships between retinal layer measurements and cognitive functioning after adjustment for confounders, and we examined the moderating effects of demographic and vascular health factors.

Results: The analytical sample included 2,483 participants who were 54.3 years old (SD 13.8 years) on average. After full adjustment, each 1-SD decrease in mGCL volume was associated with a greater decrease in global function than that of pRNFL thickness (β = −0.048, 95 % confidence interval (CI) [-0.077, -0.018] vs β = -0.021, 95 % CI [-0.049, 0.007]). These relationships increased in strength with advancing age, were stronger in participants with hypertension, and were reversed in current smokers relative to nonsmokers.

Conclusions: mGCL volume is more strongly related to adult cognitive functioning than pRNFL thickness, making it a better potential biomarker of neurodegeneration. Age and vascular health factors play important roles in determining the strength and direction of this association.

INTRODUCTION

Thinning of the retinal nerve fiber layer (RNFL) and the ganglion cell layer (GCL) of the retina, assessed by optical coherence tomography (OCT), is a purported biomarker of neurodegeneration.[1] Thinner RNFL and GCL have been associated with smaller brain volumes[2] and an increased risk of cognitive decline and dementia.[3,4] In cross-sectional analyses of healthy individuals, thinner RNFL has been associated with worse cognitive functioning in multiple cognitive domains.[5,6]

It is likely that a multitude of factors affect how closely the condition of the retina reflects level of brain function. Due to retinal changes occurring in accordance with the development of brain neurodegenerative processes,[1] retinal measurements may be better indicators of cognitive functioning in later than in earlier adulthood. In addition, vascular health factors such as hypertension, smoking, diabetes mellitus, and stroke may also affect the retina-brain relationship because they are important determinants of both retinal thickness[7] and later-life dementia risk.[8] Despite this, these probable moderating effects, which may affect the validity of using OCT as a biomarker of brain status, remain largely untested.

In this study, we sought to investigate how closely measurements of retinal structures by spectral-domain OCT (SD-OCT) reflect level of cognitive functioning and to identify factors that affect these relationships. Specifically, our first aim was to determine the degree to which thinner peripapillary RNFL (pRNFL) thickness and less macular GCL (mGCL) volume are associated with worse cognitive functioning. Our second aim was to determine how age, sex, hypertension, smoking, diabetes mellitus, and stroke influence these relationships.

METHOD

Participants

The data used for this investigation were obtained from the first 3,000 participants of the Rhineland Study (age range 30-95 years). The Rhineland Study is a population-based prospective cohort study of community-dwelling individuals and is focused predominantly on generating new knowledge of causes, biomarkers, and pathophysiology of age-related

neurodegenerative conditions. Participants in the Rhineland Study are selectively recruited from two municipal districts in Bonn, Germany, and are primarily white of European descent. Those in the current study were recruited progressively from March 2016 to December 2018. At baseline and at each planned follow-up visit, participants complete an 8-hour in-depth multidomain phenotypic assessment of anthropometry, physical activity and fitness, cardiovascular health, brain imaging, cognitive testing, neurologic functioning, ophthalmologic health and functioning, and other sensory systems. No financial incentives were offered for study participation.

Of the initial 3,000 participants, a total of 410 participants were removed from the analyzed sample due to predetermined exclusion criteria: 207 individuals did not speak German as a first language, which reduced the validity of their cognitive test data; 94 individuals had a diagnosis of macular degeneration; 91 individuals had a diagnosis of glaucoma; 13 individuals had a diagnosis of retinopathy; 4 individuals had a diagnosis of dementia; and 1 individual had incurred a severe traumatic brain injury. Participants with missing data from the OCT (n = 62) or the cognitive testing (n = 37) were also excluded. An additional 8 individuals were removed due to age-specific outlying OCT or cognitive testing data, which were identified with the application of an outlier-labeling rule[9] with an interquartile range multiplier of 3.0. The final analyzed sample comprised 2,483 participants.

Standard protocol approvals, registrations, and patient consents

Approval to undertake the study using humans was obtained from the ethics committee of the University of Bonn, Medical Faculty. The study was carried out in accordance with the recommendations of the International Council for Harmonization Good Clinical Practice standards. We obtained written informed consent from all participants in accordance with the Declaration of Helsinki.

Assessment of cognitive functioning

Participants completed a 50-minute cognitive test battery that assessed 5 core cognitive domains across multiple modalities. Executive function was assessed with a verbal fluency test (total number of animals named in 60 seconds), Trail-Making Test Part B (time

to completion), and an eye-tracker antisaccade task (percentage of erroneous prosaccades). Processing speed was assessed with Trail-Making Test Part A (time to completion) and an eye-tracker prosaccade task (mean latency). Working memory was assessed with a Corsi block tapping test (sum of forward and backward span) and a digit span test (sum of forward and backward span). Verbal episodic memory was assessed with two components of a verbal learning and memory test (immediate recall trials 1–5 total, delayed recall). Crystallized intelligence was assessed with a multiple-choice vocabulary intelligence test (total real words correctly identified). All tests were administered in German by trained study technicians according to standardized protocols. The scores of cognitive tests with skewed distributions were adjusted with log transformations, and some cognitive test scores were inverted so that higher scores represented better performance.

Ophthalmologic assessment

We measured retinal layers using a Spectralis SD-OCT (Heidelberg Engineering, Heidelberg, Germany) and previously described automated segmentation algorithms.[10] In short, assessment of them GCL volume was based on a 20° x 20° volume scan (97 horizontal B scans with 20 automatic real time frames), and data were extracted with the Heidelberg Eye Explorer software. mGCL volumes were adjusted for the number of scans before use in statistical analyses. pRNFL thickness was measured with a 3.5 mm circular scan around the optic nerve head, the positioning of which was based on each individual's anatomy using the Anatomical Positioning System.[11] This measure of global pRNFL was created from averaging up to 100 B scans and is representative of the six sectors surrounding the optic nerve head.[11] We measured pRNFL in units of thickness (micrometers) with a circular scan and mGCL in units of volume (cubic millimeters) with a volume scan because these are standard methods in both routine clinical and research settings.[12,13] In the absence of contraindications, participants were dilated for imaging with the standard mydriatic agents tropicamide and phenylephrine. Refraction and best-corrected visual acuity were measured with an automated refractometer (Ark-1s, NIDEK Co, Tokyo, Japan). Spherical equivalent was calculated as the spherical value and half of the cylindrical value. Diagnoses of glaucoma, macular degeneration, and retinopathy were

collected by self-report during a medical interview. If available, the ophthalmologic data for the right eye were used for each individual. In the absence of valid data for the right eye, data from the participant's left eye were used.

Assessment of covariates

Participants completed a range of questionnaires that took on both structured interview and self-administered forms. Highest education level was determined with the International Standard Classification of Education 2011 and was coded as low (lower secondary education or below), middle (upper secondary education to undergraduate university level), and high (postgraduate university study). Whether participants had received a diagnosis of dementia or had a history of stroke was determined during a medical interview, and smoking status (current smoker, nonsmoker), age, and sex were determined via self-report. Prevalent hypertension was defined as current antihypertension medication use or by mean systolic blood pressure ≥140 mmHg or mean diastolic blood pressure ≥90 mmHg. Prevalent diabetes mellitus was determined by current antidiabetic medication use, hemoglobin A_{1c} ≥6.5 %, or fasting glucose ≥126 mg/dL (7.0 mmol/L).

Statistical analysis

Within the analytical sample, we first calculated standardized scores for each individual cognitive test and each retinal layer measurement. All other continuous variables were mean centered. Cognitive test z-scores were then averaged to produce domain scores of fluid cognitive abilities (executive function, processing speed, working memory, verbal episodic memory); the single multiple-choice vocabulary intelligence test z-score was used to represent crystallized intelligence. The fluid cognitive domain scores were then averaged to produce a measure of global function. Global function was our main outcome variable; individual cognitive subdomains were examined to understand the basis of any global function association. We used separate multiple linear regression models to quantify the change in cognitive functioning z-score per 1-SD decrease in either pRNFL thickness or mGCL volume, with the uncertainty of the estimate represented through 95 %

confidence intervals (CIs). Age, sex, education, estimated refractive error (spherical equivalent), best-corrected visual acuity, smoking status, hypertension, diabetes mellitus, and history of stroke were included as covariates in all statistical models on the basis of possibly confounding the relationship between retinal layer measurement and cognitive functioning.

We first determined the age-adjusted and full-adjusted associations of each retinal layer measurement with cognitive functioning across the entire analytical sample. Next, we examined whether the association of retinal layer measurement and cognitive functioning differed between prespecified subgroups by including interaction terms in separate full-adjusted models. On the basis of a recent meta-analysis of the systematic determinants of pRNFL thickness[7] (mGCL volume was not assessed), subgroups of interest included sex (men, women), history of hypertension (yes, no), smoking status (current smoker, nonsmoker), and history of stroke (yes, no). Diabetes mellitus (yes, no), which is another major vascular risk factor, was also evaluated. In addition, due to the importance of educational level with regard to risk of cognitive impairment and dementia,[8] we coded a bivariate education-level variable (low: lower secondary education or below and upper secondary education to undergraduate university level; high: postgraduate university study) for use in interaction models. Associations of retinal layer measurements and cognitive functioning for each subgroup were determined using the intercept and interaction term coefficient from within each model, and no stratified models were computed. Finally, to quantify the degree to which the associations of retinal layer measurements and cognitive functioning changed across the adult lifespan, we added interaction terms of age and retinal layer to full-adjusted models. To clarify the basis of any retinal layer by age interaction, the full-adjusted associations of retinal layer measurements and cognitive functioning were then calculated within each age group (30-39, 40-49, 50-59, 60-69, 70-79, 80-94 years) separately.

We used the Hmisc R package version 4.1-1[14] to impute missing covariate data (predictive mean matching, 15 imputations). Covariate data were missing with respect to estimated refractive error (n = 8, 0.3 %), education (n = 15, 0.6 %), best-corrected visual acuity (n = 7, 0.3 %), hypertension (n = 44, 1.8 %), smoking status (n = 15, 0.6 %), and diabetes mellitus (n = 63, 2.5 %). Both adding an age^2 covariate in each linear model and fitting

generalized additive models with smoothed terms were investigated but discarded on the basis of parsimony and not affecting the interpretation of findings. All statistical analyses were undertaken with an α level of 0.05 using R version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria).[15]

Data availability

The datasets for this manuscript are not publicly available because of data protection regulations. Access to data can be provided to scientists in accordance with the Rhineland Study's Data Use and Access Policy. Requests for information or to access the datasets should be directed to RS-DUAC@dzne.de.

RESULTS

Demographics

The characteristics of the final analyzed sample of 2,483 individuals are presented in Table 1. Age ranged from 30 to 94 years, and the sample was highly educated. Included participants were younger ($p < .001$) and more educated ($p < .001$) than those who were excluded. No differences between the included and excluded participants existed after adjustment for age in the proportion of men/women, smoking status, hypertension, diabetes mellitus, or history of stroke. On average, the analyzed sample had a pRNFL thickness of 100.0 µm (SD = 10.8 µm) and an mGCL volume of 1.1 mm^3 (SD = 0.1 mm^3). Participants tended to be more myopic than hyperopic (spherical equivalent, M = -0.6, SD = 2.5), and vision was normal (best-corrected visual acuity, M = 1.1, SD = 0.2). The thickness of the pRNFL was positively correlated with the volume of the mGCL (r = 0.68).

Table 1. Sample characteristics

Characteristic	Excluded participants (n = 517)	Analyzed sample (n = 2,483)
Age, years		
Mean (SD)	60.0 (16.1)	54.3 (13.8)
Range	30–95	30–94
Sex, n (%)		
Women	294 (57)	1,400 (56)
Men	223 (43)	1,083 (44)
Education level, n (%)		
High (ISCED-11: 7-8)	239 (48)	1,342 (54)
Middle (ISCED-11: 3-6)	233 (47)	1,089 (44)
Low (ISCED-11: 0-2)	29 (6)	37 (1)
Smoking status, n (%)		
Non-smoker	437 (86)	2,141 (87)
Current smoker	74 (14)	327 (13)
Prevalent hypertension, n (%)		
No	257 (51)	1,516 (62)
Yes	249 (49)	923 (38)
Prevalent diabetes, n (%)		
No	462 (92)	2,307 (95)
Yes	42 (8)	113 (5)
History of stroke, n (%)		
No	498 (96)	2,410 (97)
Yes	19 (4)	73 (3)

Abbreviation: ISCED = International Standard Classification of Education. Prevalent hypertension was defined by either current antihypertension medication use or by mean systolic blood pressure ≥140 mmHg and mean diastolic blood pressure ≥90 mmHg. Prevalent diabetes was determined by current antidiabetic medication use, hemoglobin A_{1C} ≥6.5 or fasting glucose ≥126 mg/dL (7.0 mmol/L).

Associations of retinal layer measurements and cognitive functioning

We observed multiple relationships between retinal layer measurements and performance in different cognitive domains across the whole sample, with thinner retinal layers most commonly associated with worse cognitive functioning. The standardized associations of retinal layer measurements and cognitive functioning, adjusted for age and for age, sex, education, estimated refractive error (spherical equivalent), best corrected visual acuity, smoking status, hypertension, diabetes mellitus, and history of stroke, are presented in Figure 1. Each 1-SD decrease in mGCL volume was associated with significantly worse global function after adjustment for age and full adjustment. For the cognitive subdomains, significant full-adjusted associations were present between less mGCL volume and worse processing speed, as well as worse verbal episodic memory. Associations of pRNFL thickness with cognitive functioning were in the same direction as those of mGCL volume but were weaker and not statistically significant when fully adjusted for potential confounders.

Interactions among retinal layer measurements, demographic and vascular health factors, and cognitive functioning

We next examined whether the associations of pRNFL thickness and mGCL volume with cognitive functioning differed by key demographic and vascular health factors. The standardized associations of retinal layer measurements and global function, stratified by subgroup and presented alongside the significance test for the associated retinal layer x subgroup interaction term, are given in Table 2. We observed significant moderating effects of smoking status and hypertension on the relationships between retinal layer measurements and global function. For smoking status, the expected relationship of less mGCL volume and thinner pRNFL thickness with worse global function was observed in nonsmokers, but the opposite relationship was observed in smokers. In smokers, thinner retinal layers were associated with better global function. For hypertension, less mGCL volume was more strongly associated with worse global function in those with hypertension than in those without hypertension. A similar but weaker and not statistically significant interaction was observed for pRNFL thickness. For the cognitive subdomains and in the same direction as global function, smoking status significantly

moderated the relationships of both pRNFL thickness and mGCL volume with executive function and verbal episodic memory. The presence of hypertension significantly strengthened the relationship between mGCL volume and executive function, and the presence of diabetes mellitus significantly strengthened the relationship between pRNFL thickness and working memory. The associations of retinal layer measurements and cognitive functioning were not significantly different between subgroups for any other cognitive subdomain, and sex, education level, and stroke did not significantly affect any relationship.

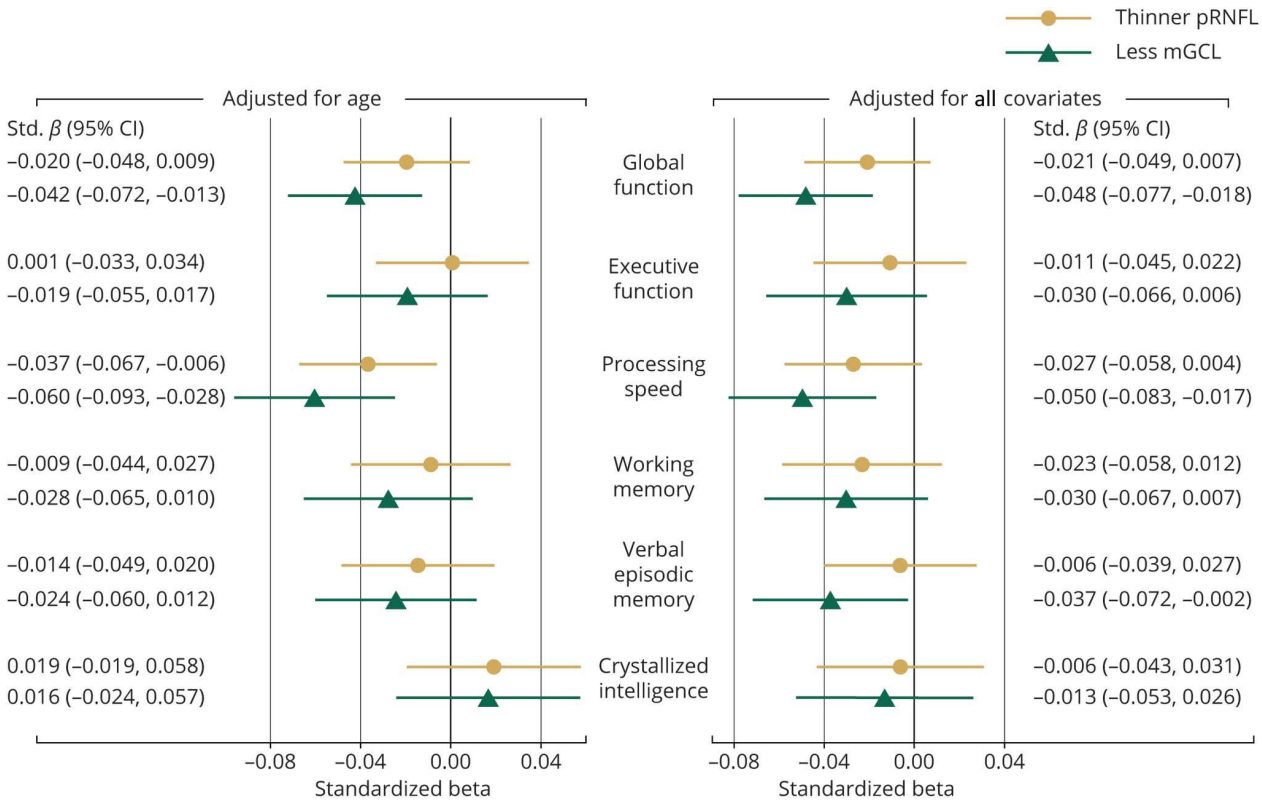

Figure 1. Standardized associations of pRNFL thickness (yellow) and mGCL volume (green) with cognitive functioning. Values represent the mean change in cognitive domain z-score per 1-SD decrease in retinal layer measurement (n = 2,483). Covariates included age, sex, education, estimated refractive error (spherical equivalent), best-corrected visual acuity, smoking status, hypertension, diabetes mellitus, and history of stroke. CI = confidence interval; mGCL = macular ganglion cell layer; pRNFL = peripapillary retinal nerve fiber layer.

Table 2. Standardized associations of retinal layer measurements and global function by subgroup

Subgroup	n	pRNFL thickness		mGCL volume	
		β value (95 % CI)	Interaction *p* value	*β* value (95 % CI)	Interaction *p* value
Sex					
Women	1400	-0.035 (-0.072, 0.002)	.245	-0.056 (-0.094, -0.017)	.489
Men	1083	-0.003 (-0.057, 0.051)		-0.037 (-0.090, 0.017)	
Education level					
High	1357	-0.015 (-0.052, 0.021)	.669	-0.042 (-0.082, -0.003)	.746
Low	1126	-0.027 (-0.081, 0.027)		-0.051 (-0.105, 0.002)	
Smoking status					
Non-smoker	2141	-0.038 (-0.067, -0.008)	.002	-0.066 (-0.097, -0.035)	< .001
Current smoker	327	0.087 (0.009, 0.165)		0.080 (0.000, 0.160)	
Hypertension					
No	1516	-0.009 (-0.045, 0.026)	.329	-0.018 (-0.056, 0.020)	.019
Yes	923	-0.037 (-0.091, 0.018)		-0.084 (-0.140, -0.029)	

Continuation Table 2

Subgroup	n	pRNFL thickness		mGCL volume	
		β value (95 % CI)	Interaction *p* value	*β* value (95 % CI)	Interaction *p* value
Diabetes mellitus					
No	2307	-0.015 (-0.044, 0.013)	.108	-0.042 (-0.072, -0.011)	.209
Yes	113	-0.112 (-0.230, 0.006)		-0.112 (-0.220, -0.003)	
Stroke					
No	2410	-0.015 (-0.044, 0.013)	.065	-0.043 (-0.073, -0.013)	.126
Yes	73	-0.153 (-0.299, -0.007)		-0.158 (-0.305, -0.011)	

Abbreviations: CI = confidence interval; mGCL = macular ganglion cell layer; pRNFL = peripapillary retinal nerve fiber layer. Values represent the change in global function z-score (95 % CI) per 1-SD decrease in pRNFL thickness or mGCL volume. Subgroup associations were calculated from models that included interaction terms of factor x retinal layer measurement, and were adjusted for age, sex, education, estimated refractive error (spherical equivalent), best-corrected visual acuity, smoking status, hypertension, diabetes mellitus and history of stroke.

Interactions among age, retinal layer measurements, and cognitive functioning

When age by retinal layer interaction terms were added to the full-adjusted models, we found that each additional year of age was associated with an increase in the strength of association between less mGCL volume and thinner pRNFL thickness and worse global function (β = -0.003, 95 % CI [-0.005, -0.001], p = .002 and β = -0.002, 95 % CI [-0.004, 0.000], p = .014, respectively). Among the cognitive subdomains, each year of age was associated with a significantly stronger relationship between less mGCL volume and worse executive function (β = -0.003, 95 % CI [-0.006, -0.001], p = .004), as well as worse verbal episodic memory (β = -0.002, 95 % CI [-0.005, 0.000], p = .026). For global function, the full-adjusted standardized associations of retinal layer measurements and cognitive performance stratified by age group are presented in Figure 2. Age stratified analysis showed that, in general, the associations of retinal layer measurements and cognitive functioning emerged after the age of 50 years, weakened between 70 and 79 years of age, and became strongest after the age of 80 years.

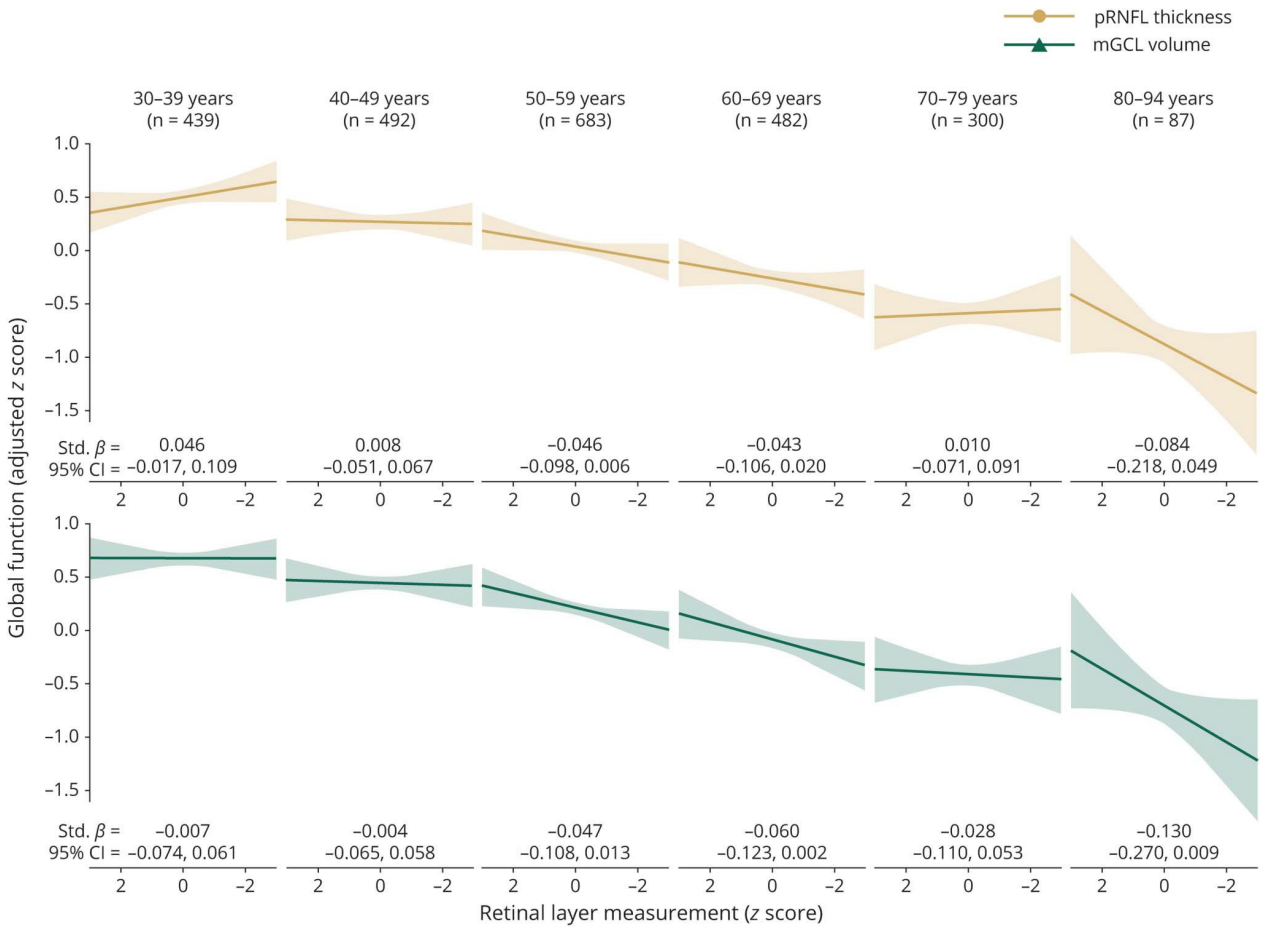

Figure 2. Associations of retinal layer measurements and global function differ by age. Data presented here represent the age-specific change in global function (z-score) as peripapillary retinal nerve fiber layer (pRNFL) thickness (yellow) and macular ganglion cell layer (mGCL) volume (green) z-scores decrease. Plotted slopes were derived from models that were adjusted for sex, education, estimated refractive error (spherical equivalent), best-corrected visual acuity, smoking status, hypertension, diabetes mellitus, and history of stroke. The presented values represent standardized β coefficients with 95 % confidence intervals (CIs) and were calculated from models that were also adjusted for age within each age group.

DISCUSSION

This study was designed to determine the degree to which thinner pRNFL and less mGCL volume are associated with worse cognitive functioning and whether demographic or vascular health factors affect these relationships. To evaluate this, we quantified the associations of retinal layer measurements and multidomain cognitive functioning and examined the interactive effects of age, sex, education level, hypertension, smoking status, diabetes mellitus, and history of stroke. From our sample of 2,483 community-

dwelling individuals 30 to 94 years of age, we report three main findings. First, mGCL volume is more closely associated with cognitive functioning than pRNFL thickness. Second, associations of retinal layer measurements and cognitive functioning are dependent on age, emerging in midlife and becoming strongest after the age of 80 years. Third, hypertension and current smoking affect the relationship between retinal layer measurements and cognitive functioning. Understanding which retinal structures are most closely related to cognitive functioning and in whom these associations are strongest provides important information about the potential utility and suitable application of OCT-derived retinal measurements as biomarkers of brain status.

Previous population-based research using SD-OCT has reported worse cognitive functioning in those with thinner RNFL[3,4,6] and less GCL-inner plexiform layer[3] in cross-sectional analyses. In our total sample (mean age 54.3 years) and after adjustment for age, sex, education level, spherical equivalent, best-corrected visual acuity, hypertension, smoking status, diabetes mellitus, and history of stroke, less mGCL volume was linearly associated with worse global function, processing speed, and verbal episodic memory. In contrast, pRNFL thickness was not significantly associated with performance in any cognitive domain. Overall, we observed a trend of thinner pRNFL thickness and less mGCL volume being associated with worse cognitive functioning in most cognitive domains, with the size of the effects tending to be stronger for the mGCL than for the pRNFL. Nonetheless, it must be noted that the size of the linear relationships was small, with standardized β values after full adjustment ranging from -0.006 for the association of thinner pRNFL thickness with worse verbal episodic memory and worse crystallized intelligence to -0.050 for the association of less mGCL volume with worse processing speed. The weak associations of retinal layer measurements and cognitive functioning in our overall sample suggest that mGCL volume and pRNFL thickness likely possess limited utility as biomarkers of brain function in relatively healthy adults before older age.

Age-dependent associations of retinal layer measurements and cognitive functioning were evident within our data. Only one previous population-based study has explored whether an age dependency in the effect of retinal layer measurement on cognitive functioning exists, albeit quantifying only the pRNFL and using a previous-generation retinal imaging technique.[16] In that study, in a sample of 1,485 individuals between 18 and 85 years of

age, the authors reported that the associations of pRNFL thickness and cognitive functioning weakened with age. In contrast, we found the opposite: the association of thinner retinal layers and worse cognitive functioning emerged after the age of 50 years, weakened between the ages of 70 and 79 years, and was strongest in those >80 years of age. Observing the strongest relationships between retinal layer measurements and cognitive functioning in our oldest participants is relevant in the context of neurodegeneration; two longitudinal studies recently demonstrated that thinner RNFL is associated with both an increased risk of cognitive decline[4] and incident dementia.[3] The few participants in our study with a diagnosis of dementia were excluded from the analytical sample (n = 4), indicating that the associations of retinal layer measurements and cognitive functioning are not necessarily dependent on the inclusion of individuals with clinical cognitive dysfunction. The reasons we observed the strongest retina-cognition relationships in our oldest participants are not clear, and some speculation is required. For instance, the greatest variance in the presence and severity of brain and retinal pathologies is expected in older individuals. Many of these detrimental physiologic changes in the brain are not destined to cause dementia, but they could cause nonclinical declines in cognitive functioning, leading to stronger retina-cognition associations. An alternative reason is that prodromal neurodegenerative diseases are most likely to be present in older age, which could be partially driving the strength of the associations in the oldest group.

We observed that vascular health factors, specifically smoking status, hypertension, diabetes mellitus, and history of stroke, tended to affect the degree to which retinal layer measurements were associated with cognitive functioning. To the best of our knowledge, this is the first time that such moderating effects have been reported. Both hypertension and stroke have been associated with a thinner pRNFL,[7] and these factors, along with diabetes mellitus, were associated with a strengthening of the relationship between retinal layer measurements and cognitive functioning in our study. This indicates that the expression of a relationship between retinal layer measurements and cognitive functioning may partially depend on a relatively higher rate of accumulation of vascular pathology; in those individuals without significant declines in their vascular health, retinal layer measurements may provide limited information on brain status. Indeed, the integrity of the retina is intimately linked to the functionality of its vasculature, with individual risk for ocular

diseases, particularly glaucoma,[17] heightened in those with greater vascular risk. We unexpectedly observed thinner pRNFL thickness and less mGCL volume to be associated with better cognitive functioning in smokers only. In contrast to hypertension and stroke, current smoking has been associated with a thicker pRNFL, possibly because smoking-related axonal degeneration may thicken retinal layers through reduced axonal flow or axonal swelling in the retina.[7] Smoking also has antagonistic effects on cognitive functioning, having been associated with both a faster global cognitive decline[18] and a cognitive advantage due to exposure to nicotine.[19] Therefore, an explanation for our findings is that a thicker retinal layer in smokers is a marker of having incurred more smoking-related vascular pathology, leading to thinner layers being associated with better cognitive functioning in smokers only. An alternative explanation is that current smoking status is a proxy for heightened nicotine exposure, leading to thinner retinal layers but better cognitive performance in some individuals. Of course, another explanation is that this interaction represents a chance finding, and replication in other cohorts is required.

Our results should be interpreted with respect to a number of limitations. First, this study examined cross-sectional relationships exclusively. Therefore, we could not explore whether thinner retinal layers are associated with a greater risk of incident cognitive decline or dementia. Second, due to the small number of participants in our sample with a diagnosis of dementia, we excluded these individuals before the final analysis. Therefore, we cannot report the strength of relationships between pRNFL thickness and mGCL volume across the continuum of cognitive functioning from healthy to clinically impaired. Third, we cannot discount the possibility that a survivor bias led to an underestimation of the associations of retinal layer measurements and cognitive functioning in those 70 to 79 years of age, but this explanation is discrepant with the observation that the strongest effects were found in those >80 years of age. Fourth, differences in the strengths of observed retinal measurement-cognition relationships could, in part, reflect a differential tendency of thickness and volume measures to decline in response to smaller pathologic changes, regardless of retinal location. This does not, however, affect our conclusions with regard to which retinal measurements are most viable as biomarkers of brain function. Fifth, our participants were predominantly white of European descent, and confirmation of our findings in other ethnic groups is required. Finally, we did not adjust α levels for multiple comparisons when denoting statistically

significant effects, and some level of chance findings, particularly with regard to cognitive subdomains and subgroup differences, is a possibility. Still, our study possessed a number of strengths. We were able to analyze a large community-dwelling sample from the Rhineland Study, which included a very wide adult age range (30-94 years). This allowed us to investigate age-related interactions between retinal layer measurements and cognitive functioning across the adult lifespan, which has received minimal prior research attention. Due to the availability of data on a wide range of cognitive tests, we also were able to use composite domain summary scores as our outcome measures, reducing the risk of spurious findings within single cognitive tests.

Our findings serve to provide greater clarity with regard to which retinal structures are most closely related to cognitive functioning and in whom these relationships are strongest. Such knowledge is important in furthering the development and suitable application of OCT-derived retinal measurements as biomarkers of brain status. We found evidence that retinal layer measurements are most indicative of brain function in later life and that vascular health factors play important roles in determining how the associations of retinal layer measurements and cognitive functioning manifest.

ACKNOWLEDGEMENT

The authors thank Andre Medek, Benjamin Meier, Thomas Schmidt, and Christof Meigen for their contributions in the process of data handling and management.

STUDY FUNDING

The Rhineland Study at the German Center for Neurodegenerative Diseases is funded predominantly by the Federal Ministry of Education and Research and the Ministry of Culture and Science of the German State of North Rhine-Westphalia.

DISCLOSURE

D. Ward, M. Mauschitz, M. Bönniger, and N. Merten report no disclosures relevant to the manuscript. R. Finger reports personal fees from Novartis, Bayer, Roche/Genentech,

Santhera Pharmaceuticals, Opthea, Inositec, Ellex, and Alimera Sciences and research grants from Novartis, CenterVue, and Zeiss. M. Breteler reports no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

REFERENCES

1. den Haan J, Verbraak FD, Visser PJ, Bouwman FH. Retinal thickness in Alzheimer's disease: A systematic review and meta-analysis. *Alzheimer's Dement Diagnosis, Assess Dis Monit*. 2017;6:162-170. doi:10.1016/j.dadm.2016.12.014

2. Mutlu U, Bonnemaijer PWM, Ikram MA, et al. Retinal neurodegeneration and brain MRI markers: The Rotterdam Study. *Neurobiol Aging*. 2017;60:183-191. doi:10.1016/j.neurobiolaging.2017.09.003

3. Mutlu U, Colijn JM, Ikram MA, et al. Association of retinal neurodegeneration on optical coherence tomography with dementia: A population-based study. *JAMA Neurol*. 2018;75(10):1256-1263. doi:10.1001/jamaneurol.2018.1563

4. Ko F, Muthy ZA, Gallacher J, et al. Association of retinal nerve fiber layer thinning with current and future cognitive decline: A study using optical coherence tomography. *JAMA Neurol*. 2018;75(10):1198-1205. doi:10.1001/jamaneurol.2018.1578

5. Khawaja AP, Chan MPY, Yip JLY, et al. Retinal nerve fiber layer measures and cognitive function in the EPIC-Norfolk cohort study. *Investig Ophthalmol Vis Sci*. 2016;57(4):1921-1926. doi:10.1167/iovs.16-19067

6. Jones-Odeh E, Yonova-Doing E, Bloch E, Williams KM, Steves CJ, Hammond CJ. The correlation between cognitive performance and retinal nerve fibre layer thickness is largely explained by genetic factors. *Sci Rep*. 2016;6:34116. doi:10.1038/srep34116

7. Mauschitz MM, Bonnemaijer PWM, Diers K, et al. Systemic and ocular determinants of peripapillary retinal nerve fiber layer thickness measurements in the European Eye Epidemiology (E3) population. *Ophthalmology*. 2018;125(10):1526-1536. doi:10.1016/j.ophtha.2018.03.026

8. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet*. 2017;390(10113):2673-2734. doi:10.1016/S0140-6736(17)31363-6

9. Hoaglin DC, Iglewicz B, Tukey JW. Performance of some resistant rules for outlier labeling. *J Am Stat Assoc*. 1986;81(396):991-999. doi:10.1080/01621459.1986.10478363

10. Mauschitz MM, Holz FG, Finger RP, Breteler MMB. Determinants of macular layers and optic disc characteristics on SD-OCT: The Rhineland Study. *Transl Vis Sci Technol*. 2019;8(3). doi:10.1167/tvst.8.3.34

11. Mauschitz MM, Roth F, Holz FG, Breteler MMB, Finger RP. The impact of lens opacity on SD-OCT retinal nerve fiber layer and Bruch's membrane opening measurements using the Anatomical Positioning System (APS). *Investig Ophthalmol Vis Sci*. 2017;58(5):2804-2809. doi:10.1167/iovs.17-21675

12. Gardiner SK, Fortune B, Demirel S. Localized changes in retinal nerve fiber layer thickness as a predictor of localized functional change in glaucoma. *Am J Ophthalmol*. 2016;170:75-82. doi:10.1016/j.ajo.2016.07.020

13. Sadda SR, Guymer R, Holz FG, et al. Consensus definition for atrophy associated with age-related macular degeneration on OCT: Classification of atrophy report 3. *Ophthalmology*. 2018;125(4):537-548. doi:10.1016/j.ophtha.2017.09.028

14. Harrel F. Hmisc: Harrell Miscellaneous. R Package Version 41-1. Published 2018. Accessed March 15, 2019. CRAN.R-project.org/package=Hmisc

15. R Core Team. A language and environment for statistical computing. Vienna: R foundation for statistical computing.

16. van Koolwijk LME, Despriet DDG, van Duijn CM, et al. Association of cognitive functioning with retinal nerve fiber layer thickness. *Investig Ophthalmol Vis Sci*. 2009;50:4576-4580. doi:10.1167/iovs.08-3181

17. Choi J, Kook MS. Systemic and ocular hemodynamic risk factors in glaucoma. *Biomed Res Int*. 2015;2015:141905. doi:10.1155/2015/141905

18. Sabia S, Elbaz A, Dugravot A, et al. Impact of smoking on cognitive decline in early old age: The Whitehall II cohort study. *Arch Gen Psychiatry*. 2012;69(6):627-635. doi:10.1001/archgenpsychiatry.2011.2016

19. Levin ED, McClernon FJ, Rezvani AH. Nicotinic effects on cognitive function: Behavioral characterization, pharmacological specification, and anatomic localization. *Psychopharmacology (Berl)*. 2006;184(3-4):523-539. doi:10.1007/s00213-005-0164-7

4.2 Perceived stress but not hair cortisol concentration is related to adult cognitive performance

Sadia E. Oumohand,[1] David D. Ward,[1] Meta M. Boenniger,[1] Natascha Merten,[1] Clemens Kirschbaum,[2] and Monique M.B. Breteler[1,3]

[1]Population Health Sciences, German Center for Neurodegenerative Diseases (DZNE), Venusberg-Campus 1, Building 99, 53127 Bonn, Germany

[2]Faculty of Psychology, Technische Universität Dresden, 01062 Dresden, Germany

[3]Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), Faculty of Medicine, University of Bonn, Venusberg-Campus 1, Building 11, 53127 Bonn, Germany

Journal: Psychoneuroendocrinology 2020; 121:104810,
doi: 10.1016/j.psyneuen.2020.104810

ABSTRACT

Chronic stress detrimentally affects cognition but evidence from population-based studies is scarce and largely based on one-dimensional stress assessments. In this study, we aimed to investigate associations of subjective and psychological chronic stress measures with cognition in a population-based sample of adults aged 30-95 years from the Rhineland Study. Participants completed the Perceived Stress Scale (subjective measure) and a cognitive test battery (n = 1,766). Hair cortisol concentration (physiological measure) was assessed by liquid chromatography tandem mass spectrometry in 1,098 participants. Cross-sectional associations between the two measures of chronic stress and cognition were investigated using multivariable linear regression models. Subjective and physiological measures of chronic stress were not associated with each other (B = 0.005, 95 % CI [-0.005, 0.015]). Participants with higher perceived stress and specifically lower perceived self-efficacy performed worse in all cognitive domains (effect sizes ranged from β = -0.129, 95 % CI [-0.177, -0.080] to -0.054, 95 % CI [-0.099, -0.009]; and from β = 0.052, 95 % CI [0.005, 0.098] to 0.120, 95 % CI [0.072, 0.167], respectively). Relationships between subjective chronic stress measures and executive functioning were stronger in men compared to women (interaction β = -0.144, 95 % CI [-0.221, -0.067]). Relationships between perceived stress and working memory, and between perceived self-efficacy and executive functioning, processing speed, verbal episodic and working memory, increased with older age. Hair cortisol concentration was not associated with performance in any cognitive domain. Our results suggest that subjective and physiological measures capture different aspects of chronic stress in the general population.

INTRODUCTION

Chronic stress is commonly known for its detrimental effects on cognitive functioning, and it may facilitate the onset and progression of cognitive decline and dementia later in life.[1–4] However, empirical evidence from population-based studies is scarce and difficult to aggregate. Reasons for this include the wide range of one-dimensional stress assessment tools, i.e. those tools that measure only psychological or physiological stress exposure or response despite stress representing a multidimensional concept,[5,6] as well as the unclear associations among individual stress measures,[5] and varying associations of stress measures with cognitive outcomes.[e.g.7–9]

Chronic stress depends strongly on an individual's appraisal or perception of a stressor as such[10] and thus is often assessed with subjective measures, such as the Perceived Stress Scale.[11] Elevated perceived stress has been repeatedly associated with worse cognitive functioning, and may be a risk factor for cognitive decline.[8,12–15] The relationship between cognition and subdimensions of perceived stress (i.e. perceived helplessness and perceived self-efficacy) has received less attention. Yet, existing evidence suggests that they may be differentially related to cognition in older adults. Higher perceived self-efficacy (the perception of one's ability to cope with stressors) has been linked to better language, memory, processing speed and executive functioning abilities, whereas higher perceived helplessness (the negative affect related to stress experiences), appears unrelated to cognition.[8,16,17]

Besides subjective measures, quantifiable physiological markers of stress, such as the glucocorticoid hormone cortisol, might allow for a more objective assessment of chronic stress. Cortisol is secreted by the hypothalamic-pituitary-adrenal (HPA) axis as part of the physiological stress response. It can pass the blood-brain barrier and drive plasticity changes in several brain structures involved in cognitive functioning, including the hippocampus, amygdala and prefrontal cortex.[18,19] Elevated cortisol concentrations have been associated with worse cognitive functioning.[20–25] However, the majority of previous studies have relied on cortisol measurements from blood, saliva or urine, which are sensitive to short-term changes in cortisol concentration. Reliable long-term estimates in blood, saliva, or urine can only be attained through repeated measurements, but these often result in laborious sampling protocols impractical in the context of population-based

studies. A newer, more resource-efficient method is the measurement of cortisol deposits in scalp hair (hair cortisol concentration, HCC), which reflects cumulative cortisol exposure up to several months prior to sampling.[26] To date, investigations on the relation between HCC and cognition in healthy adults are scarce and findings remain inconclusive.[9,27–29] Some have reported no association between HCC and cognitive performance,[e.g.9,28] while others have reported that higher HCC are associated with worse cognitive performance,[e.g.27] as well as with better learning and memory.[29] Such work highlights the need for further investigation.

Although both subjective and physiological measures have been used to assess chronic stress, their relation with each other it is still under debate.[6] The latest meta-analysis in 10,289 people[30] could not confirm an association between perceived stress and HCC. However, most of the included individual samples were relatively small (n < 50). Of the few existing larger and population- or community-based studies in adults, only one reported an absence of association[31] while the majority, including two newer studies not included in the meta-analysis,[32,33] have reported either weak positive linear[32–34] or curvilinear relationships.[35] In addition, the relevance of HCC as a biomarker of chronic stress with regard to cognitive health in the general population remains to be evaluated. In this study, we aimed to investigate if subjective and physiological measures capture connected aspects of chronic stress and how they relate to performance in a range of cognitive domains across the adult life span in a large population-based sample. We hypothesized that higher perceived stress is associated with higher HCC and that both higher perceived stress and higher HCC are associated with worse cognitive performance.

METHODS

Study population

We used cross-sectional data from the Rhineland Study, an on-going population-based prospective cohort study in Bonn, Germany.[36] Participants are exclusively recruited by invitation from two geographically defined residence areas around two designated study centers. Participants were required to be 30 years or older and have sufficient knowledge of the German language to consent to study participation. As part of their study visit, participants completed self-administered questionnaires, underwent 50 minutes of

cognitive testing, and provided hair samples. Approval to undertake the study was received from the ethics committee of the University of Bonn, Medical Faculty. The study was carried out in accordance with the recommendations of the International Council for Harmonization (ICH) Good Clinical Practice (GCP) standards (ICH-GCP). We obtained written informed consent from all participants in accordance with the Declaration of Helsinki. No financial incentives were offered for study participation.

Our study sample was based on the first 2,000 participants of the Rhineland Study, who completed their study visits between March 7, 2016 and June 8, 2018. We excluded participants who did not speak German as a first language (n = 141), because most of our cognitive tests were validated for German-speaking participants only. We further excluded participants for whom educational status was unknown (n = 21). Of the remaining participants, 1,766 had complete cognitive and subjective stress data. Of those, hair samples were obtained in 1,098 participants (62.2 %). Reasons for exclusion from hair sample collection were insufficient hair length (< 3 cm or bald; n = 588, 88.0 %), refusal to sampling (n = 48, 7.2 %), or hairstyle (dreadlocks or extensions, n = 3, 0.4 %). Further, in the first months of the study (March - June 2016), hair sampling was not yet part of a participants' study visit (n = 29, 4.3 %).

Assessment of chronic stress

Subjective measure

We used the 10-item version of Cohen's Perceived Stress Scale (PSS) as a subjective measure of chronic stress. Participants indicated the degree to which they perceived situations in their life within the last month as unpredictable, uncontrollable, and overloaded.[11,37] Participants rated each item on a 5-point Likert-type scale from "never" to "very often" on a computer tablet. After reversing positively scored items, an overall perceived stress score was calculated with higher values reflecting higher levels of stress (range 0–40). Additionally, two subscores, perceived helplessness and perceived self-efficacy, were calculated.[8,16] The perceived helplessness score comprised the sum of the six negatively worded items. Higher values reflect higher perceived helplessness (range 0–24). The perceived self-efficacy score was derived from the sum of the four positively worded items. Higher values reflect higher perceived self-efficacy (range 0–16). The

perceived stress score and the subscores showed good reliability in our sample (Cronbach's α = 0.86 for perceived stress and perceived helplessness, Cronbach's α = 0.71 for perceived self-efficacy).

Physiological measure

We used HCC as a physiological measure of chronic stress. We cut hair strands of approximately 3 mm diameter from participants' scalps at the posterior vertex. Hair was stored in aluminum foil in a dry and dark place until analysis. Under the assumption of a hair growth rate of approximately 1 cm per month,[38] the scalp-nearest 3 cm of hair, reflective of a 3-month-period before sampling, were analyzed by liquid chromatography tandem mass spectrometry (LC–MS/MS) in two batches at TU Dresden, as previously described by Gao and colleagues.[39] Hair strands were shortened to 3 cm and washed twice for three minutes in 2.5 mL isopropanol. After drying for at least 12 h under a fume hood, 7.5 mg hair from each sample was cut to smaller pieces. For cortisol extraction, they were incubated at room temperature for 18 h in 1.8 mL LC–MS-grade methanol (Carl Roth, Karlsruhe, Germany). Then, 1.6 mL of the supernatant was transferred into a new tube. All liquid components were evaporated for approximately 20 minutes at 50 °C under a constant stream of nitrogen. Finally, 225 µL of a mixture of methanol and water (50:50, v/v) was added to the dried sample. Of the resulting solution, 100 µL were used for LC–MS/MS analysis.

Cognitive assessment

Cognitive tests were administered in German by certified study technicians according to standardized protocols. Crystallized intelligence was measured with the 37-item multiple choice vocabulary test (Mehrfachwahl-Wortschatz Intelligenztest, MWT-B) in which participants were instructed to identify a correct German word among four imaginary words.[40] Processing speed was assessed using a numbers-only trail-making test (TMT A, time to completion in seconds) and an eye-tracker prosaccade task (mean latency to look towards a cue in milliseconds). Executive functioning was assessed using a numbers-and-letters trail-making test (TMT B, time to completion in seconds), an eye-tracker

antisaccade task (percentage of trials in which participants erroneously looked towards a cue rather than away from it), and a categorical verbal fluency task (total number of animals named in 60 s). Verbal episodic memory was assessed using a verbal learning and memory test adapted from Rey[41] (number of correctly remembered words in immediate recall (recalls 1–5) and delayed recall). Working memory was assessed using a Corsi block tapping test adapted from the PEBL battery[42] (sum of forward and backward span) and a digit span test (sum of forward and backward span). We used adapted versions of these reliable and valid cognitive tests, including only those in our testing protocol with well-established psychometric properties. As such, we did not reassess the reliability and validity specific to our dataset. We calculated composite cognitive domain scores from cognitive tests that assessed theoretically similar cognitive processes and ensured that intra-domain tests were sufficiently correlated.

Other variables

We a priori identified age, sex, and education as possible confounders, because previous literature has reported systematic differences in both cognition and stress as a result of age, sex and education.[e.g.27,30] In sensitivity analyses, we additionally accounted for hair-related factors and corticosteroid medication use. Education was defined as the highest educational attainment participants indicated in a self-administered questionnaire. Based on the International Standard Classification of Education (ISCED) 2011, educational attainment was classified as low (completed lower secondary education or below), middle (completed upper secondary education up to completed Bachelor's degree or equivalent), or high (completed Master's degree or equivalent up to completed doctoral degree or equivalent). Hair-related factors were obtained in a short interview along with the hair sample and included current natural hair color (light = blond and red, dark = brown and black, white/grey), average hair washes per week (rarely = three times or less, often = more than three times), hair treatment (permanent waves, dyed, tinted or bleached hair within 3 months prior to sampling), and season. The season in which a hair sample was grown was derived from the preceding three-month period including the month of acquisition. A hair sample was categorized to one of four seasons if at least two of the months fell into the predefined range of winter (December - February), spring (March -

May), summer (June - August), or fall (September - November). Corticosteroid medication use was defined as any use, regular or as needed, within the previous 12 months of drugs from the Anatomical Therapeutic Chemical (ATC) Classification System classes A01AC, A07EA, C05AA, D01AC20, D07, D10AA, G01B, H02, M01BA, N02CB, R01AD, R03AK, R03AL, S01BA, S01BB, S01CA, S01CB, S02B, S02C, S03B, or S03C.

Statistical analysis

Hair samples in which cortisol concentration was below detection limit (n = 7) were assigned the lowest available cortisol value from the dataset. Skewed variables were transformed (log normal for HCC, $\log_{10}$ for TMT A, TMT B and prosaccade mean latency) in order to approximate normal distributions, and age was mean centered. Values for TMT and eye-tracking tests were reversed so that higher values represent better performance across all cognitive tests. To compute cognitive domain scores, individual cognitive tests were z-standardized and then all within-domain cognitive test z-scores were averaged. For a participant to receive a cognitive domain score, valid results on at least 50 % of individual component tests were required. We tested for possible group differences in characteristics between participants who provided a hair sample and those who did not using logistic linear regression.

We investigated the association between subjective and physiological measures of chronic stress and between chronic stress and cognitive performance using multiple linear regression models. In our sample, the two PSS subscores, perceived helplessness and perceived self-efficacy, were moderately correlated with each other (r = -0.6). To avoid collinearity effects, the associations of perceived stress, perceived helplessness and perceived self-efficacy with cognition were assessed in separate models. To quantify the relationship between subjective and physiological measures of chronic stress, we used models with HCC as the dependent and each PSS score as the independent variable, while adjusting for age and sex. We further tested if any of the relationships differed between men and women or across age by including interaction terms. To assess the relationship between chronic stress and cognitive performance, we used separate models for each cognitive domain, with cognitive performance as the dependent and each measure of stress (HCC, perceived stress, perceived helplessness, or perceived self-

efficacy) as the independent variable, while adjusting for age, sex and education. We further tested if any of the relationships differed between men and women or across age by including interaction terms. Because standardized effects per year of age were small, the results for the interactions were presented as standardized effects per decade increase in age. To be able to compare effects within the same subsample, both stress measures and cognitive domain scores were z-standardized using mean and standard deviation of the complete sample before they were entered into the models.

To test the robustness of associations, we tested if relationships between PSS scores and cognition differed in the smaller subsample of participants for whom hair was available. Further, we separately adjusted all models including HCC additionally for hair color, washing frequency, treatment, season, analysis batch, and corticosteroid medication use, because they had been most consistently identified to affect the HCC levels.[e.g.27,30] All analyses were performed in R version 3.4.3[43] and RStudio version 1.1.383[44] using the packages psych[45] and stats.

RESULTS

Characteristics of the sample

Sample characteristics are shown in Table 1. In comparison to participants who did not provide a hair sample, participants from whom a hair sample was taken were more likely to be women, were less likely to have white or grey hair, washed their hair less often, and were more likely to report the use of hair treatments.

Association between subjective and physiological measures of chronic stress

After adjusting for the effects of age and sex, perceived stress was not associated with HCC (B = 0.005, 95 % CI [-0.005, 0.015], p = .339). The same held true for perceived helplessness (B = 0.004, 95 % CI [-0.010, 0.018], p = .606) and perceived self-efficacy (B = -0.017, 95 % CI [-0.041, 0.006], p = .154). We found no interaction of any of these relationships with sex or age (p > .100, not shown). All results were virtually unchanged when models were further adjusted for hair color, washing frequency, hair treatment use, season, analysis batch and corticosteroid medication use.

Table 1. Participant characteristics

	All participants n = 1,766	Hair available n = 1,098	Hair unavailable n = 668	Group difference[a]
Age, mean years (SD)	54.5 (13.7)	54.8 (13.5)	54.0 (14.1)	.257
Women, n (%)	989 (56.0)	853 (77.7)	136 (20.4)	<.001
Stress assessment				
HCC, median pg/mg (IQR)	-	5.55 (8.42)	-	
Perceived stress, mean (SD)	13.9 (6.1)	14.5 (6.2)	13.0 (5.8)	.183
Perceived helplessness, mean (SD)	8.7 (4.3)	9.1 (4.3)	8.0 (4.1)	.249
Perceived self-efficacy, mean (SD)	10.8 (2.5)	10.6 (2.6)	11.0 (2.5)	.213
Cognitive assessment (z-score)				
Crystallized intelligence, mean (SD)	0 (1.0)	0 (0.9)	0 (1.0)	.588
Executive functioning, mean (SD)	0 (0.7)	0 (0.7)	0.1 (0.7)	.181
Processing speed, mean (SD)	0 (0.8)	0 (0.8)	0 (0.8)	.741
Verbal episodic memory, mean (SD)	0 (0.9)	0.1 (0.9)	-0.1 (0.9)	.309
Working memory, mean (SD)	0 (0.7)	0 (0.7)	0.1 (0.7)	.575
Education				
Lower secondary or below, n (%)	31 (1.8)	26 (2.4)	5 (0.7)	ref
Upper secondary to bachelor's degree, n (%)	783 (44.3)	506 (46.1)	277 (41.5)	.672
Master's degree or above, n (%)	952 (53.9)	566 (51.5)	386 (57.8)	.840
Season				
Winter, n (%)	544 (31.3)	340 (31.0)	204 (31.9)	ref
Spring, n (%)	407 (23.4)	252 (23.0)	155 (24.2)	.234
Summer, n (%)	325 (18.7)	213 (19.4)	112 (17.5)	.327
Fall, n (%)	462 (26.6)	293 (26.7)	169 (26.4)	.752

Continuation Table 1.

	All participants n = 1,766	Hair available n = 1,098	Hair unavailable n = 668	Group difference[a]
Hair color				
Light, n (%)	491 (28.3)	329 (30.0)	162 (25.5)	ref
Dark, n (%)	552 (31.9)	364 (33.2)	188 (29.6)	.322
White or grey, n (%)	690 (39.8)	404 (36.8)	286 (45.0)	.021
Hair washed >3 times/week, n (%)	963 (55.5)	492 (44.8)	471 (73.9)	<.001
Hair treated, n (%)	553 (32.0)	490 (44.7)	63 (9.9)	.005
Corticosteroid users, n (%)	214 (12.7)	143 (13.6)	71 (11.2)	.468

Abbreviations: SD: standard deviation, HCC: hair cortisol concentration, IQR: interquartile range, ref: reference.
[a]p-values derived from age- and sex-adjusted logistic regressions (group differences in age and sex were adjusted only for the respective other).

Associations between chronic stress and cognitive performance

After adjusting for age, sex and education, the two measures of chronic stress were differentially associated with cognitive performance in the five measured domains (Figure 1). Participants with higher perceived stress performed significantly worse across all cognitive domains. Regarding the subscores of the PSS, participants with higher perceived self-efficacy performed significantly better in all cognitive domains. For participants with higher perceived helplessness, associations went into the same direction as for overall perceived stress, but associations were less strong and reached statistical significance for executive functioning and working memory only. We observed no marked differences in these associations in the subset of participants for whom hair was available (Supplementary Table A.1). HCC was not related to performance in any cognitive domain either with or without further adjustment for hair color, washing frequency, hair treatment use, season, analysis batch, and corticosteroid medication use (not shown).

The inclusion of interaction terms of stress measures with sex or age did not change the direction of the presented associations (Table 2). Associations between perceived stress, perceived helplessness and perceived self-efficacy and executive functioning were stronger in men compared to women (difference in β ranged from -0.144 for negative to 0.120 for positive associations). With each decade of age, associations between perceived stress and working memory and between perceived self-efficacy and executive functioning, processing speed, verbal episodic memory and working memory became stronger (change in β ranged from -0.032 for negative to 0.044 for positive associations). With higher age, the association between perceived helplessness and executive functioning weakened slightly. The results were only marginally different in the subset of participants who provided a hair sample (Supplementary Table A.2). Associations between HCC and cognitive performance did not differ between sexes and did not change with age either with or without further adjustment for hair color, washing frequency, hair treatment use, season, analysis batch, and corticosteroid medication use (not shown).

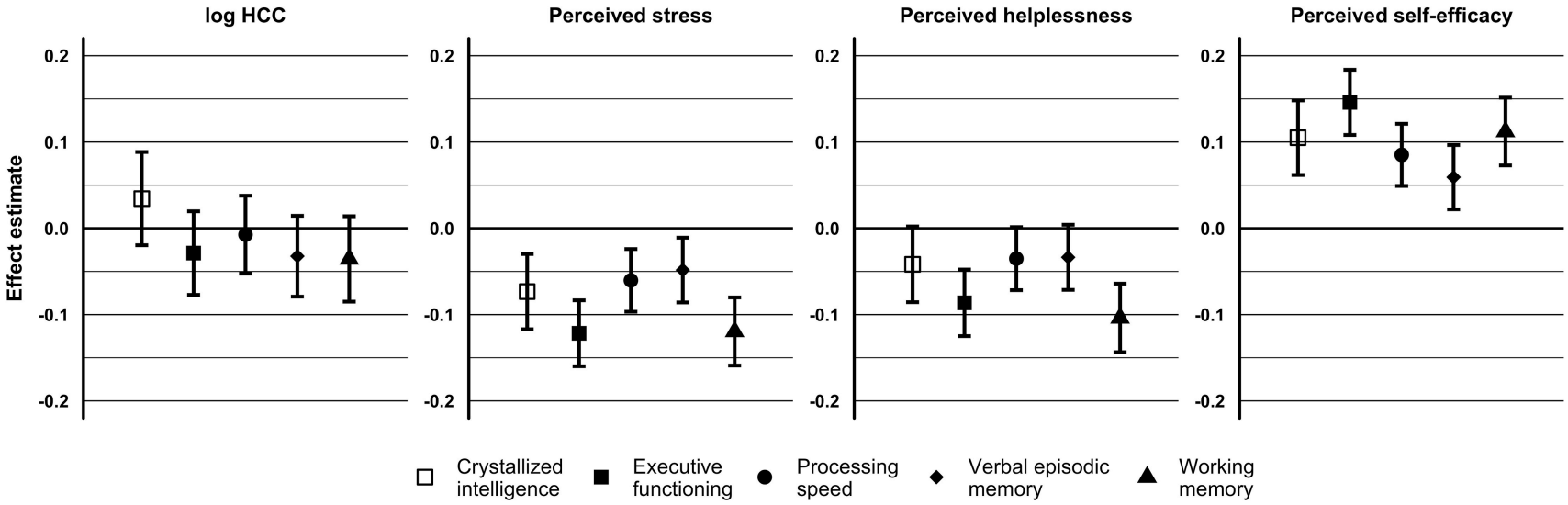

Figure 1. Associations between subjective and physiological measures of chronic stress on cognitive performance. Standardized associations and 95 % confidence intervals of chronic stress measures and cognitive domain performance as determined by multiple linear regressions models. Values represent change in cognitive domain Z-score per one standard deviation increase in stress measure after adjusting for age, sex and education. Models including HCC were based on n = 1,098 participants, models including PSS scores were based on n = 1,766 participants. Abbreviations: HCC: hair cortisol concentration.

Table 2. Interaction effects of subjective and physiological chronic stress measures with age and sex on cognitive performance.

	Crystallized intelligence		Executive functioning		Processing speed		Verbal episodic memory		Working memory	
Interaction	β (95% CI)	*p*	β (95% CI)	*p*	β (95% CI)	*p*	β (95% CI)	*p*	β (95% CI)	*p*
HCC*sex	-0.033 (-0.172;0.107)	.645	0.052 (-0.073;0.177)	.418	0.056 (-0.060;0.173)	.344	0.038 (-0.083;0.159)	.535	0.059 (-0.069;0.187)	.364
PSS*sex	-0.045 (-0.133;0.044)	.322	-0.144 (-0.221;-0.067)	<.001	-0.064 (-0.137;0.010)	.088	-0.046 (-0.122;0.030)	.237	-0.017 (-0.097;0.063)	.674
PSS-H*sex	-0.039 (-0.127;0.050)	.394	-0.123 (-0.201;-0.045)	.002	-0.044 (-0.118;0.030)	.240	-0.027 (-0.103;0.049)	.491	-0.002 (-0.082;0.078)	.962
PSS-SE*sex	0.031 (-0.056;0.117)	.487	0.120 (0.044;0.195)	.002	0.068 (-0.004;0.140)	.063	0.057 (-0.018;0.131)	.135	0.025 (-0.054;0.103)	.538
HCC*age	-0.018 (-0.058;0.022)	.378	0.016 (-0.020;0.052)	.388	-0.011 (-0.044;0.023)	.530	0.001 (-0.034;0.035)	.964	0.010 (-0.026;0.047)	.577
PSS*age	0.007 (-0.025;0.038)	.683	0.003 (-0.025;0.031)	.816	-0.012 (-0.038;0.015)	.392	-0.023 (-0.050;0.005)	.107	-0.032 (-0.061; -0.003)	.029
PSS-H*age	0.011 (-0.020;0.043)	.487	0.030 (0.002;0.057)	.033	0.008 (-0.018;0.035)	.527	-0.008 (-0.035;0.019)	.545	-0.010 (-0.038;0.019)	.495
PSS-SE*age	-0.006 (-0.037;0.024)	.686	0.027 (0.000;0.054)	.048	0.032 (0.007;0.058)	.014	0.032 (0.005;0.058)	.020	0.044 (0.016;0.072)	.002

Continuation Table 2.

Standardized associations and 95 % confidence intervals of subjective chronic stress measures and cognitive domain performance as determined by multiple linear regression models. For interactions with sex, values represent the difference in the association of stress measure and cognitive domain *z*-score between women and men (reference = women). For interactions with age, values represent the change in the association of chronic stress measure and cognitive domain z-score per 10 years of age. All associations were adjusted for age, sex, and education. All models were based on n = 1,098 participants who provided a hair sample. *Abbreviations*: CI: confidence interval, PSS: Perceived stress, PSS-H: perceived helplessness, PSS-SE: perceived self-efficacy.

DISCUSSION

We aimed to investigate if subjective and physiological measures capture connected aspects of chronic stress and how they relate to cognitive performance in the Rhineland Study. We found that, in our large population-based sample, subjective and physiological measures of chronic stress were not related to each other across the adult lifespan. Higher perceived stress was associated with worse performance in multiple cognitive domains. Of the subdimensions of perceived stress, perceived self-efficacy showed stronger relationships with cognitive outcomes than perceived helplessness. HCC was not related to performance in any cognitive domain. Our results suggest that subjective and physiological measures capture separate aspects of chronic stress, which differentially relate to cognitive functioning.

The absence of a relationship between subjective and physiological measures of chronic stress in our population may be explained by the stress-intensity hypothesis.[30] According to this hypothesis, a coupling of perceived stress and cortisol levels occurs only above a certain threshold of stress intensity. As HCC is a long-term cortisol measure, circulating cortisol needs to be elevated for several weeks or months to manifest in elevated hair levels, requiring a person to encounter intense stressors constantly or very often. Such a high stressor burden is not evident in our population-based sample. Inconsistencies between our results and findings from other larger and community- and population-based studies could stem from differences between study samples. Previous studies included larger proportions of people with a lower average educational level[33] or with higher perceived stress,[32,34,35] indicating a higher stressor burden. While the results from the current study suggest that HCC is not associated with chronic stress measures in groups reporting low levels of stress, HCC may be a useful tool to assess chronic stress and its effects on the body in groups with a relatively high stress burden.

Despite a moderate to low perceived stress intensity in our sample, we observed associations of between higher perceived stress –and specifically lower perceived self-efficacy– with worse performance across all examined cognitive domains. The relationships between subjective chronic stress measures and cognitive performance were consistently stronger among the executive function and working memory domains. The latter is in line with the existing literature on the effects of chronic stress on memory.[19]

Our findings further find support for the existence of a similarly strong effect on executive functioning matching the observation of worse executive functioning in stress-related disorders, such as depression.[46] Different relationships of the two subscores of the PSS, perceived helplessness and perceived self-efficacy, with cognition underline the presumed two-factor structure of the PSS across the adult age range.[8,16] Although associations between the two subscores and cognition showed a similar pattern, only perceived self-efficacy was significantly associated with all cognitive domains measured. Therefore, perceived self-efficacy seemingly plays a more important role than perceived helplessness in the association between perceived stress and cognition. However, it remains to be clarified whether lower perceived self-efficacy directly relates to worse cognitive functioning or whether it moderates performance in cognitive testing situations. Lastly, given our cross-sectional design, the associations between perceived stress measures and cognition, especially crystallized intelligence, could have also arisen from cognitive ability shaping stress perception. The latter would be in line with previous findings about crystallized intelligence being a relatively stable construct.[40,47] Longitudinal studies will allow further clarification of the causality underlying the associations between perceived stress and cognitive performance.

We observed both sex and age to modify the relationship between subjective chronic stress measures and executive functioning. Specifically, the associations between subjective measures and executive functioning were significantly stronger in men compared to women. Men generally report lower levels of stress in the PSS than women,[16,37] but definite explanations for this observation are lacking. In the context of gender norms, men might, despite feeling similarly stressed, report that they perceive less stress than women. If that were the case, we would expect a stronger relationship between PSS scores and all cognitive domains in men. However, only the associations between PSS measures and executive functioning were more pronounced in men and further investigation into why the sex effect was limited to this one domain is warranted. With regard to age, the associations between perceived self-efficacy and cognitive performance were stronger in older-aged participants. Although the reported interaction effects were relatively small, perceived self-efficacy seems to play a role in cognitive functioning among older individuals and interventions to strengthen self-efficacy could potentially ameliorate the negative effects of stress on cognition, especially in the elderly.

Although not observed for perceived self-efficacy, we found the association between perceived helplessness and executive functioning to be less pronounced in older participants. It remains to be investigated whether older adults either experience less stress-related negative emotions that would impact cognition or whether their cognitive performance is less affected by these emotions, for example as a consequence of different coping strategies or higher resilience compared to younger adults.

HCC was not associated with performance in any cognitive domain. Our findings are in line with the majority of studies in non-clinical adult samples.[9,28] Two contrasting studies showed positive[27] and negative relationships between HCC and cognition.[29] While a positive association is uncommon, the negative association seems more likely, given the majority of previous literature on cortisol measures and cognitive functioning.[3] Differences in the study samples could explain the variation in results. Compared to the two studies that found significant relationships, our sample was relatively healthy and highly educated. It is possible that negative effects of elevated cortisol on cognition only become evident in individuals with other concurrent vulnerabilities, such as persistent stress-related psychiatric illnesses (e.g. depression) or a lower educational level.

Limitations of our study include the cross-sectional nature of the data, which did not allow testing for temporality in the observed relationships. Although longitudinal investigations point towards an effect of perceived stress on cognition,[7,12] we cannot rule out that cognitive ability in turn shaped stress perception in our sample. Furthermore, it is possible that perceived stress affected the test performance itself through an influence on mood and motivation. This influence might have inflated the strength of associations between subjective chronic stress measures and cognitive functions. A further limitation of the present study was that only 3 cm hair segments were available for analysis. This segment length was chosen because of the trade-off between temporal resolution and analysis costs. While HCC are highly intercorrelated between the three 1 cm proximal hair segments, analytical costs triple compared to a single 3 cm segment analysis. Such costs would have been prohibitively high for the current research project. Due to the 3 cm hair length requirement in our population, hair samples were available from a larger proportion of women than men. Similarly, unbalanced ratios have been observed in previous community- and population-based samples.[27,32–35] More men than women did not meet

the minimum hair length criterion because of short hairstyles or baldness. Although hair samples of only 1 cm would have matched the PSS measurement period better and could have resulted in a more balanced sample, the collection of sufficient hair material (5–7.5 mg minimum) would have remained difficult. Men with very short or near-to-baldness hair and also some women might not accept the amount of hair that would have needed to be cut. Due to the unavailability of hair samples in part of our population, our analysis samples for HCC and PSS differed in size and effect estimates for the reported associations with cognitive performance may not be directly comparable. However, when we restricted our statistical analyses to the subset of participants who provided a hair sample, associations between subjective chronic stress measures and cognitive outcomes were virtually unchanged. We found no association between perceived stress measures and HCC, yet we may have underestimated the association between them due to the moderate to low stress level observed in our study population. Finally, we only included age, sex, and education, as covariates in our models, and residual confounding may have occurred. However, many further possible confounders, such as BMI, cardiovascular health or depressive symptoms, may also be in the causal pathway and be mediators of stress effects on cognition. They were reported to worsen after stressful experiences and elevated cortisol.[4,48,49] Therefore, we did not include those possible mediators in our analyses.

To the best of our knowledge, this is the first population-based study that has investigated the relationship between perceived stress, HCC and a wide range of cognitive functions. Previous studies have used either one-dimensional approaches to study stress, applied more momentary cortisol assessments from blood or saliva as a proxy for longer-term cortisol concentrations, or investigated only a limited range of cognitive outcome measures. Further, our study underlines the importance of examining the two-factor structure of the PSS in relation to health outcomes across the adult age range.

CONCLUSION

Our results add to the body of evidence suggesting that subjective and physiological measures might capture separate aspects of chronic stress, which differentially relate to cognitive functioning in the general population. Assessing stress with multimodal methods

in the future will further deepen our understanding of the impact that chronic stress may have on human health and wellbeing.

AUTHOR CONTRIBUTIONS

MMBB and SEO: Designed the study.

SEO, MMB, and NM: Trained study personnel and oversaw and took part in data acquisition.

DDW, MMB, and NM: Conceptualized the cognitive domain scores.

CK: Provided the analysis of hair samples.

SEO: Performed data analyses and drafted the manuscript.

All authors read and improved subsequent versions of the manuscript.

FUNDING

This work was supported by the Federal Ministry of Education and Research, Germany, (grant number 01EA1601) in the framework of the Joint Programming Initiative "A Healthy Diet for A Healthy Life". The funding source was not involved in the study design, the collection, analysis and interpretation of data, the writing of the manuscript and the decision to submit the article for publication.

DECLARATION OF COMPETING INTEREST

None.

ACKNOWLEDGEMENTS

The authors want to especially thank all participants of the Rhineland Study and the study personnel involved in the extensive data collection for this study. Furthermore, the authors would like to thank Dr. Ahmad Aziz for his advice on statistical matters and Dr. Wei Gao

for the personal communication about methodological details regarding the mass spectrometry analysis.

REFERENCES

1. Dong H, Csernansky JG. Effects of stress and stress hormones on amyloid-β protein and plaque deposition. *J Alzheimer's Dis*. 2009;18(2):459-469. doi:10.3233/JAD-2009-1152

2. Justice NJ. The relationship between stress and Alzheimer's disease. *Neurobiol Stress*. 2018;8:127-133. doi:10.1016/j.ynstr.2018.04.002

3. Lupien SJ, McEwen BS, Gunnar MR, Heim C. Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nat Rev Neurosci*. 2009;10(6):434-445. doi:10.1038/nrn2639

4. Marin MF, Lord C, Andrews J, et al. Chronic stress, cognitive functioning and mental health. *Neurobiol Learn Mem*. 2011;96(4):583-595. doi:10.1016/j.nlm.2011.02.016

5. Epel ES, Crosswell AD, Mayer SE, et al. More than a feeling: A unified view of stress measurement for population science. *Front Neuroendocrinol*. 2018;49:146-169. doi:10.1016/j.yfrne.2018.03.001

6. Weckesser LJ, Dietz F, Schmidt K, Grass J, Kirschbaum C, Miller R. The psychometric properties and temporal dynamics of subjective stress, retrospectively assessed by different informants and questionnaires, and hair cortisol concentrations. *Sci Rep*. 2019;9(1):1-12. doi:10.1038/s41598-018-37526-2

7. Aggarwal NT, Wilson RS, Beck TL, et al. Perceived stress and change in cognitive function among adults 65 years and older. *Psychosom Med*. 2014;76(1):80-85. doi:10.1097/PSY.0000000000000016

8. Korten NCM, Comijs HC, Penninx BWJH, Deeg DJH. Perceived stress and cognitive function in older adults: Which aspect of perceived stress is important? *Int J Geriatr Psychiatry*. 2017;32(4):439-445. doi:10.1002/gps.4486

9. McLennan SN, Ihle A, Steudte-Schmiedgen S, Kirschbaum C, Kliegel M. Hair cortisol and cognitive performance in working age adults. *Psychoneuroendocrinology*. 2016;67:100-103. doi:10.1016/j.psyneuen.2016.01.029

10. Lazarus RS, Folkman S. *Stress, Appraisal, and Coping*. first. ed. Springer Publishing Company; 1984.

11. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983;24(4):385-396. doi:10.2307/2136404

12. Feeney J, O'Sullivan M, Kenny RA, Robertson IH. Change in perceived stress and 2-year change in cognitive function among older adults: The Irish Longitudinal Study on Ageing. *Stress Heal*. 2018;34(3):403-410. doi:10.1002/smi.2799

13. Katz MJ, Derby CA, Wang C, et al. Influence of perceived stress on incident amnestic mild cognitive impairment: Results from the Einstein Aging Study. *Alzheimer Dis Assoc Disord*. 2016;30(2):93-98. doi:10.1097/WAD.0000000000000125

14. Munoz E, Sliwinski MJ, Scott SB, Hofer S. Supplemental material for global perceived stress predicts cognitive change among older adults. *Psychol Aging*. 2015;30(3):487-499. doi:10.1037/pag0000036.supp

15. Turner AD, James BD, Capuano AW, Aggarwal NT, Barnes LL. Perceived stress and cognitive decline in different cognitive domains in a cohort of older African Americans. *Am J Geriatr Psychiatry*. 2017;25(1):25-34. doi:10.1016/j.jagp.2016.10.003

16. Ezzati A, Jiang J, Katz MJ, Sliwinski MJ, Zimmerman ME, Lipton RB. Validation of the Perceived Stress Scale in a community sample of older adults. *Int J Geriatr Psychiatry*. 2014;29(6):645-652. doi:10.1002/gps.4049

17. Jiang JM, Seng EK, Zimmerman ME, Kim M, Lipton RB. Positively worded subscale score of the Perceived Stress Scale is associated with cognitive domain function. *J Behav Brain Sci*. 2017;7:311-324. https://doi.org/10.4236/jbbs.2017.77023

18. Lupien SJ, Juster R-P, Raymond C, Marin MF. The effects of chronic stress on the human brain: From neurotoxicity, to vulnerability, to opportunity. *Front Neuroendocrinol*. 2018;49:91-105. doi:10.1016/j.yfrne.2018.02.001

19. Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain Cogn*. 2007;65(3):209-237. doi:10.1016/j.bandc.2007.02.007

20. Gerritsen L, Comijs HC, Deeg DJH, Penninx BWJH, Geerlings MI. Salivary cortisol, APOE-ε4 allele and cognitive decline in a prospective study of older persons. *Neurobiol Aging*. 2011;32(9):1615-1625. doi:10.1016/j.neurobiolaging.2009.09.007

21. Kalmijn S, Launer LJ, Stolk RP, et al. A prospective study on cortisol, dehydroepiandrosterone sulfate, and cognitive function in the elderly. *J Clin Endocrinol Metab*. 1998;83(10):3487-3492. doi:10.1210/jcem.83.10.5164

22. Karlamangla AS, Singer BH, Chodosh J, McEwen BS, Seeman TE. Urinary cortisol excretion as a predictor of incident cognitive impairment. *Neurobiol Aging*. 2005;26(SUPPL.):80-84. doi:10.1016/j.neurobiolaging.2005.09.037

23. Lee BK, Glass TA, McAtee MJ, et al. Associations of salivary cortisol with cognitive function in the Baltimore Memory Study. *Arch Gen Psychiatry*. 2007;64(7):810-818. doi:10.1001/archpsyc.64.7.810

24. Potvin O, Forget H, Préville M, Berbiche D, Chagnon YC, Hudon C. Relationship between cortisol level and prevalent/incident cognitive impairment and its moderating factors in older adults. *Int Psychogeriatrics*. 2013;25(2):252-262. doi:10.1017/S1041610212001706

25. Schrijvers EMC, Direk N, Koudstaal PJ, et al. Associations of serum cortisol with cognitive function and dementia: The Rotterdam Study. *J Alzheimer's Dis*. 2011;25(4):671-677. doi:10.3233/JAD-2011-110224

26. Russell E, Koren G, Rieder M, Van Uum S. Hair cortisol as a biological marker of chronic stress: Current status, future directions and unanswered questions. *Psychoneuroendocrinology*. 2012;37(5):589-601. doi:10.1016/j.psyneuen.2011.09.009

27. Feeney J, O'Halloran AM, Kenny RA. The association between hair cortisol, hair cortisone, and cognitive function in a population-based cohort of older adults: Results from the Irish Longitudinal Study on Ageing. *Journals Gerontol Ser A gly258*. 2020;75(2):257-265. https://doi.org/10.1093/gerona/gly258

28. Jansen P, Dahmen-Zimmer K, Kudielka BM, Schulz A. Effects of karate training versus mindfulness training on emotional well-being and cognitive performance in later life. *Res Aging*. 2017;39(10):1118-1144. doi:10.1177/0164027516669987

29. Pulopulos MM, Hidalgo V, Almela M, Puig-Perez S, Villada C, Salvador A. Hair cortisol and cognitive performance in healthy older people. *Psychoneuroendocrinology*. 2014;44:100-111. doi:10.1016/j.psyneuen.2014.03.002

30. Stalder T, Steudte-Schmiedgen S, Alexander N, et al. Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. *Psychoneuroendocrinology*. 2017;77(16):261-274. doi:10.1016/j.psyneuen.2016.12.017

31. Feller S, Vigl M, Bergmann MM, Boeing H, Kirschbaum C, Stalder T. Predictors of hair cortisol concentrations in older adults. *Psychoneuroendocrinology*. 2014;39(1):132-140. doi:10.1016/j.psyneuen.2013.10.007

32. Gidlow CJ, Randall J, Gillman J, Silk S, Jones M V. Hair cortisol and self-reported stress in healthy, working adults. *Psychoneuroendocrinology*. 2016;63:163-169. doi:10.1016/j.psyneuen.2015.09.022

33. Michaud DS, Feder K, Keith SE, et al. Self-reported and measured stress related responses associated with exposure to wind turbine noise. *J Acoust Soc Am*. 2016;139(3):1467-1479. doi:10.1121/1.4942402

34. O'Brien KM, Tronick EZ, Moore CL. Relationship between hair cortisol and perceived chronic stress in a diverse sample. *Stress Heal*. 2013;29(4):337-344. doi:10.1002/smi.2475

35. Wells S, Tremblay PF, Flynn A, et al. Associations of hair cortisol concentration with self-reported measures of stress and mental health-related factors in a pooled database of diverse community samples. *Stress*. 2014;17(4):334-342. doi:10.3109/10253890.2014.930432

36. Breteler MMB, Wolf H. The Rhineland Study: A novel platform for epidemiologic research into Alzheimer disease and related disorders. *Alzheimer's Dement*. 2014;10(4):P520. doi:10.1016/j.jalz.2014.05.810

37. Cohen S, Williamson GM. Perceived Stress in a probability sample of the United States. In: Spacapan S, Oskamp S, eds. *The Social Psychology of Health*. Sage; 1988:31-67.

38. Wennig R. Potential problems with the interpretation of hair analysis results. *Forensic Sci Int*. 2000;107:5-12. https://doi.org/10.1016/S0379-0738(99)00146-2

39. Gao W, Stalder T, Foley P, Rauh M, Deng H, Kirschbaum C. Quantitative analysis of steroid hormones in human hair using a column-switching LC-APCI-MS/MS assay. *J Chromatogr B*. 2013;928:1-8. doi:10.1016/j.jchromb.2013.03.008

40. Lehrl S. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. Spitta Verlag GmbH & Co. KG; 2005.

41. Rey A. Mémorisation d'une série de 15 mots en 5 répetitions. In: Rey A, ed. *L'examen Clinique En Psychologie*. 3rd ed. Presses Universitaries de France; 1970:141-193.

42. Mueller ST, Piper BJ. The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *J Neurosci Methods*. 2014;222:250-259. doi:10.1016/j.jneumeth.2013.10.024

43. R Core Team. R: A language and environment for statistical computing. Published online 2017. https://www.r-project.org/

44. RStudio Team. RStudio: Integrated development environment for R. Published online 2015. http://www.rstudio.com/

45. Revelle W. psych: Procedures for psychological, psychometric, and personality research. Published online 2018.

46. Snyder HR. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychol Bull*. 2013;139(1):81-132. doi:10.1037/a0028727

47. Luszcs M, Giles L, Eckermann S, et al. *The Australian Longitudinal Study of Ageing: 15 Years of Ageing in South Australia*. (Luszcs M, Giles L, Eckermann S, et al., eds.).; 2006.

48. Cohen S, Murphy MLM, Prather AA. Ten surprising facts about stressful life events and disease risk. *Annu Rev Psychol*. 2019;70(1):577-597. doi:10.1146/annurev-psych-010418-102857

49. Conklin AI, Guo SXR, Yao CA, Tam ACT, Richardson CG. Stressful life events, gender and obesity: A prospective, population-based study of adolescents in British Columbia. *Int J Pediatr Adolesc Med*. 2019;6(2):41-46. doi:10.1016/j.ijpam.2019.03.001

5 General discussion and outlook

MEMORY AND OTHER COGNITIVE FUNCTIONS AS MARKERS FOR NEURODEGENERATION

With increasing age, memory and most other cognitive functions have been found to decline except for general knowledge/crystallized intelligence, verbal performance and some numerical abilities.[1,2] Advancing age is also the single most important risk factor for dementia, a disorder characterized by loss of multiple cognitive functions. However, age is unlikely the cause of cognitive decline, but rather a marker of the pathologic burden associated with increasing age.[3,4] Disentangling the association between pathological stress and cognitive decline and identifying pathologies that cause cognitive decline remains a challenge.

Across age, cognitive decline is not homogenous across cognitive domains but shows distinctive patterns. For example, AD patients typically suffer from episodic memory impairment whereas patients suffering from frontotemporal dementia display more deficits in executive functioning.[5] In addition, quantitative and qualitative cognitive differences can be used to differentiate between dementia patients and healthy contemporaries.[1] In AD, these differences can not only be detected by the classic AD screening tests (Mini-Mental-State Examination (MMSE)[6] or Montreal Cognitive Assessment (MoCA)[7]), but also by other brief individual cognitive tests that show high sensitivity and specificity for identifying and distinguishing between individuals suffering from AD and individuals with suspected cognitive impairment (delayed list-recall memory test and semantic (category) fluency language tests).[8] However, most cognitive tests fail to distinguish between normal, age-associated and pathological cognitive changes in very early stages of the disease.[1,8]

One exception is the delayed memory score assessed by the AVLT, which has been found to be a reliable predictor of dementia about ten years before clinical diagnosis.[9,10] However, to infer whether this and other tests might be valid and reliable cognitive marker of pathological changes, it is important to assess their use to screen for normal and abnormal cognitive trajectories across adulthood. In addition, it is important to examine whether the AVLT may provide additional benefits in combination with other potential biomarkers.

Prospective cohort studies, such as the Rhineland Study, which examine participants longitudinally and over a wide age range using deep phenotyping, are especially well suited to examine cognitive trajectories in a cohort and allow the examination of potential biomarkers and risk factors of cognitive decline. However, selecting memory and other cognitive tests for such a study is challenging.

CHALLENGES OF SELECTING MEMORY AND OTHER COGNITIVE MEASUREMENTS FOR PROSPECTIVE COHORT STUDIES

Cognitive functions are traditionally classified into so-called domains.[11] The large number of domains poses a challenge for prospective cohort studies, which have to be mindful of the allocation of their financial resources as well as the examination time. Previous work reported that age-associated cognitive decline occurs across most domains, namely memory, processing speed, reasoning and executive functions, but verbal abilities, some numerical abilities and general/world knowledge are less affected.[1,12] Assessing all of these cognitive domains requires a large investment of time and resources by both the study and participants. Moreover, many cognitive tests capture several cognitive domains simultaneously, can be confounded by varying perceptual performance of the participants, and lack specificity, which renders their interpretation difficult. To prevent undesired task-specific factors to affect cognitive performance, it is also advised to measure one domain using multiple tasks. In addition, memory tests should allow for repeated testing without obscuring age-related cognitive decline. Memory effects caused by multiple measurements should therefore be prevented as much as possible.

Within the Rhineland Study we had a time restriction of 50 minutes for the whole cognitive examination, which led to narrowing down our test battery to only a small number of tests. We aimed to cover the larger domains for which an age-associated decrease in cognitive performance have been observed (episodic verbal memory, working memory, processing speed, executive function). To maximize test interpretability, we preferred tests measuring domains as distinctively as possible. We included different sensory information to ensure participants' performance was independent of their perceptual skills and focused on previously validated tests that were used in other cohort studies to be able to compare results. We also included a test on crystallized intelligence as a skill that is not expected

to decrease with age and to estimate participants' best cognitive ability. To assess fluid intelligence, we preferably chose tests that have been previously found to predict or to be associated with dementia (e.g. AVLT, semantic category fluency, eye movement examination etc.).[10,13–15]

As our main memory measurement, I selected the AVLT to assess learning and long-term memory. The AVLT is a time-efficient test, which provides an in-depth assessment of learning and memory function in young as well as in older adults. It can also be administered in memory-impaired participants, and has been found to predict dementia years before disease onset.[9,10,16–19] Therefore, it has the potential to serve as a marker to examine the biological basis of neurodegenerative diseases. However, for previous versions we found no thoroughly tested German adaptation with multiple list versions, which allows repeated testing without introducing bias due to memory effects.[20,21] Therefore, I developed a new German version of the AVLT including ten comparable versions based on prior published word lists in different languages to achieve comparable memorability and difficulty levels. While nine out of ten presented list versions were highly comparable, the tenth list proved slightly more difficult. We also showed that the level of difficulty was comparable to other test versions[22,23] and gave detailed references for age and sex effects on a large number of results.

These new lists are a valuable resource for clinical and scientific testing. Although the comparability of the lists could have been of no concern for the baseline assessments of the Rhineland Study if we had used only one list for all participants, we varied word lists across participants. This was to avoid learning effects as participants discussed the assessments with each other.

We also included an eye movement examination as an (more) experimental condition in our cognitive assessment battery. Eye movement examinations require usually only little instructions, are not language and education dependent and are therefore considered as culture-fair measurements.[24] We included it as objective and reproducible measures for executive functions and processing speed.[25,26] Control of eye movements is also biologically well understood[27] and impaired control of eye movements has been found in many neurodegenerative diseases including dementia.[14] This makes eye movement

examinations an interesting tool to explore as potential biomarkers, but also as behavioral outcomes for brain changes.

An alternative way of assessing memory and other cognitive functions is in combination with fMRI measurement. Task fMRI allows the assessment of cognitive functions as well as the assessment of brain activation evoked by a cognitive task. This can give valuable information on activity changes over age for specific cognitive functions like memory. Task-fMRI is not often applied in prospective cohort studies as it is not standardized and most tasks are time-consuming which makes them expensive. Most of the existing tasks are also specifically designed to answer a specific hypothesis. As mentioned in the introduction, prospective cohort studies on the other hand are often not driven by a specific hypothesis. Therefore, I developed a fMRI task measuring sensory-specific (auditory: vocal and environmental, and visual: face and scene) brain activity as well as sensory-specific and -unspecific memory-encoding-related brain activity within only ten minutes of fMRI acquisition time. The task provides a recognition memory testing and a versatile number of contrasts. Our findings replicated previously observed sensory-specific brain activity as well as sensory-specific memory brain activity (auditory and visual cortex).[28–35] In addition, we were also able to detect sensory-unspecific memory-related brain activity (hippocampus, default mode network structures, like the precuneus and the angular gyrus) supporting the central memory network.[36–39] These results proved the feasibility of our memory-encoding task in a young sample and showed the potential of providing information on changes of memory-encoding-associated brain activity in longitudinal studies within just ten minutes of fMRI acquisition time. Further, in the supplementary analyses, we showed that behavioral memory performance was slightly worse in a small older sample of participants compared to the younger sample, but hit-rates were comparable at 50 % which would allow the creation of stable contrasts. Although we did not examine whether these results also extend to older participants individuals with dementia, we assume that the task is also applicable when testing participants with mild AD, since Sperling et al.[40] successfully performed an even more difficult face-name association task. We were also unable to examine how much the task performance may be affected by memory bias in case of repeated measurements.

Although participants did not learn the stimuli in depth, it is necessary to investigate how long memory effects persist. Nevertheless, the task provides results on behavioral recognition memory performance for face, scene, vocal and environmental stimuli. This could be an interesting neuropsychological result in itself, complementing our neuropsychological memory measures. In addition, examining sensory-specific brain activity in visual and auditory regions, as well as the sensory-specific and -unspecific memory brain activity may help to identify markers of age-related changes.

The task fMRI measurement could not be integrated into the current examination protocol of the Rhineland study due to time constrains and the preference for the more general resting-state fMRI. Resting-state fMRI was preferred over task fMRI as it is practical, reproducible, standardized, allows data pooling, can be used for longitudinal examinations and shows consistent results in line with structural connectivity derived by diffusion MRI.[41] It can be used in clinical conditions and has been found useful in characterizing abnormal brain connectivity in clinical conditions.[42] However, it is also affected by explicit mental activities that occur while lying in the scanner which might make it an uncontrolled task of thoughts at rest.[43] It is also affected by age differences in factors not associated with cognitive function including movement, tendency to fall asleep, or vascular health.[43,44] Task fMRI, on the other hand, helps understanding the mechanisms of brain activity and changes using task-associated activity and has been shown to be an excellent tool for studying changes in brain activity related to AD risk.[43,45] It enables to map cognitive functions directly and not via indirect associations and is more “entertaining” for participants, who in turn are less likely to fall asleep or move compared to resting-state fMRI.[41,43,44] However, it only measures task-specific activity. Many tasks are not standardized, have incomplete descriptions or are affected by language and cultural background.[41]

Weighing the advantages and disadvantages of resting-state and task fMRI is difficult as they complement each other.[41] Resting-state allows to pose a broader range of hypotheses whereas the fMRI task paradigm may especially be of use to cohort studies interested in cognitive function and their biological basis. As it has been found that under certain conditions results of task-fMRI can be predicted by resting-state fMRI,[46] resting-state fMRI may be the more traditional approach for prospective cohort studies without

clear hypotheses. Nonetheless, I would like to encourage future studies to examine this task further. The variability in results as well as its time efficiency makes it a valuable addition to both fMRI measurement and cognitive memory examination. If the results appear as expected or even improve, I would advocate to include this task into the general protocol of a prospective cohort study or to conduct a sub-study to identify more targeted approaches to examine brain regions showing effects of cognitive decline more specifically.

USE OF COGNITIVE ASSESSMENTS IN BIOMARKER AND RISK FACTOR ASSESSMENT

We used the baseline data collection of the Rhineland Study to examine the use of our memory and cognitive assessment battery as a tool to examine biomarkers and risk factors of neurodegenerative diseases. We also aimed to assess whether cognitive functions reflect similar or maybe distinctive neuropsychological changes as potential biomarkers and risk factors. We selected retinal layers as potential biomarker and chronic stress as potential risk factor.

Retinal layers are potential biomarkers of neurodegenerative diseases as the retinal tissue shares a lot of similarities with brain tissue.[47] Therefore, neurodegenerative processes, such as thinning and loss of cells in brain tissue, might also be reflected in the retinal layers. Retinal layer assessment is non-invasive and straightforward using imaging techniques (e.g. SD-OCT), which makes it a very attractive potential biomarker. However, in contrast to previous studies, we found no association between pRNFL and any cognitive domain, including memory.[48,49] On the other hand, mGCL showed small, linear associations with global cognitive function, processing speed and episodic verbal memory.[50] The effect between lower mGCL volume and lower verbal episodic memory performance strengthened with age. These results underline the sensitivity of our cognitive battery and are in line with previous studies finding processing speed and verbal episodic memory as very early markers for cognitive decline.[9,10,51] Interestingly, we found the strength of the association between mGCL and cognitive function varied across smoking status – current smokers showed a diminished effect, whereas the effect was strongest in non-smokers. As mentioned earlier, this could be due to degeneration caused

by reduced axonal flow leading to thickening or axonal swelling.[52] It could also be due to the effects of smoking on cognitive functions, which is reported to be initially positive due to nicotine exposure,[53] but detrimental over time.[54] Overall, our results suggest that mGCL but not pRNFL may be a marker of memory and cognitive performance in our non-demented sample. However, further longitudinal studies are needed to assess whether retinal layers can capture early neuropathological changes and whether the retinal assessment reflects the same neuropathological processes captured by cognitive tests.

While examining whether chronic stress may be a modifiable risk factor of decreased memory (Chapter 4.2) we observed striking differences between our subjective and objective chronic stress measures. As mentioned previously, the difficulty in assessing chronic stress lies in the heterogeneity of the construct. We observed that HCC and PSS were not associated with another and HCC did not show associations with memory and other cognitive functions, whereas PSS scores did. In our population-based sample, which had only low to moderate PSS scores, stress levels might have been too low to show effects. This supports the stress-intensity hypothesis, which states that cortisol levels only increase above a certain stress threshold.[55] Different results in the literature exploring HCC can therefore largely be explained by sample differences.[56–59] In our project, analysis we were not able to disentangle the biological basis between perceived stress and cognitive differences due to a low number of participants reporting high perceived stress levels. Therefore, it would be of interest to repeat this analysis in a larger sample and to focus on a subset of participants reporting higher perceived stress levels. Further studies could also examine whether perceived stress levels may be associated with changes in brain regions like the limbic system and prefrontal cortex areas.[60] In addition, few studies have examined whether markers of accelerated biological aging including, telomere length and oxidative stress moderate the relation between cognitive functions and perceived stress, but further research is warranted.[61,62] In conclusion and in line with previous studies, we observed that participants reporting higher perceived stress show lower performance in all cognitive domains, especially working memory and executive functions.[63,64] This effect was strongest in participants with lower perceived self-efficacy. However, the biological basis needs to be explored in more depth.

As mentioned earlier, a major limitation of the reported studies is the cross-sectional nature of the data. Longitudinal analyses could provide essential information about normal and pathological cognitive development across adulthood and elucidate the relation of memory and cognitive function with potential biomarkers and risk factors in the adult population. Furthermore, the Rhineland Study sample is not representative of the general German population. Compared to the German population, our participants have higher socioeconomic status, higher education and lower disease prevalence on average.[65–67] Thus, the conclusions drawn so far are limited to this population.

While the observed effect sizes were smaller than expected, one has to bear in mind that our participants were on average very healthy and well-educated. Given that high levels of education are known to be associated with higher cognitive reserve,[68] it is likely that we underestimated the strength of the association between cognitive performance, retinal layer thickness and stress. Moreover, it highlights the sensitivity of our cognitive test battery which successfully captures cognitive changes across age in a non-demented sample. To further increase the validity of our cognitive domains inclusion of additional tests could be considered.

Finally, I would like to note that in the Rhineland study we were only able to start with a very short standard test battery, so there is of course room for improvement in retrospect. In the planned follow-up study, I would like to add a dementia screening assessment, especially for very old participants or for participants where we observe difficulties in the cognitive examination. We found that some of our standard tests were difficult specifically for some very old participants, resulting in a small number of missing or dubious results. Knowing whether these are participants already have dementia would be very informative, also in terms of improving our cognitive test battery. In addition, we can also consider combining our cognitive scores and adding tests where appropriate to create a score that can measure early signs of cognitive decline based on known pathology. One example is the Alzheimer's Disease Cooperative Study Preclinical Alzheimer's Cognitive Composite (ADCS-PACC) score, which has been validated based on Aβ-pathology and has been shown to reliably measure early signs of cognitive decline.[69]

CONCLUSION AND OUTLOOK

This thesis provides an overview of the use of memory and cognitive assessments as potential markers of neurodegeneration, and discusses the challenges of measuring memory and cognitive function in prospective cohort studies. This is exemplified by the neuropsychological test battery of the Rhineland Study, which aims to longitudinally assess normal and pathological changes of cognitive development across adulthood in a time-efficient manner. The cognitive battery covers episodic verbal memory, working memory, processing speed, executive function and crystallized intelligence and especially memory tests allow for repeated testing. I developed and validated two cognitive tests examining learning and memory performance and their neuronal correlates, which meet the demands of prospective cohort studies and may not only be of relevance in the research but also in the clinical setting. In addition, I found the performance on the cognitive test battery to show an association with another known biomarker and risk factor of neurodegeneration supporting the utility of the test battery as potential cognitive marker of neuropathological changes.

While the described findings were based on cross-sectional baseline data of a very healthy sample, the Rhineland Study aims to conduct multiple follow-up assessments during the next decades. This will allow to monitor healthy and pathological adult cognitive development longitudinally and assess the utility of our cognitive battery to detect neuropathological changes at an early stage. For the follow-up assessments, I would advocate to also include a dementia screening test and to consider including additional tests which have been suggested as early markers of pathological cognitive decline and differential marker of dementia subtypes.

REFERENCES

1. Deary IJ, Corley J, Gow AJ, et al. Age-associated cognitive decline. *Br Med Bull*. 2009;92(1):135-152. doi:10.1093/bmb/ldp033

2. Park DC, Reuter-Lorenz P. The adaptive brain: Aging and neurocognitive scaffolding. *Annu Rev Psychol*. 2009;60:173-196. doi:10.1146/annurev.psych.59.103006.093656.

3. Wilson RS, Wang T, Yu L, Bennett DA, Boyle PA. Normative cognitive decline in old age. *Ann Neurol*. 2020;87(6):816-829. doi:10.1002/ana.25711

4. Buckley R, Pascual-Leone A. Age-related cognitive decline is indicative of neuropathology. *Ann Neurol*. 2020;87(6):813-815. doi:10.1002/ana.25733

5. Park LQ, Harvey D, Johnson J, Farias ST. Deficits in everyday function differ in AD and FTD. *Alzheimer Dis Assoc Disord*. 2015;29(4):301-306. doi:10.1097/WAD.0000000000000081

6. Folstein MF, Folstein SE, McHugh PR. “Mini-Mental State” - A practical method for grading the cognitive state of patients for the clinician. *J Psychiat Res*. 1975;12:189-198. doi:10.1016/0022-3956(75)90026-6

7. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699. doi:10.1111/j.1532-5415.2005.53221.x

8. Hemmy LS, Linskens EJ, Silverman PC, et al. Brief cognitive tests for distinguishing clinical Alzheimer-type dementia from mild cognitive impairment or normal cognition in older adults with suspected cognitive impairment. *Ann Intern Med*. 2020;172(10):678-687. doi:10.7326/M19-3889

9. Boraxbekk C-J, Lundquist A, Nordin A, Nyberg L, Nilsson L-G, Adolfsson R. Free recall episodic memory performance predicts dementia ten years prior to clinical diagnosis: Findings from the Betula Longitudinal Study. *Dement Geriatr Cogn Dis Extra*. 2015;5(2):191-202. doi:10.1159/000381535

10. Ding X, Charnigo RJ, Schmitt FA, Kryscio RJ, Abner EL. Evaluating trajectories of episodic memory in normal cognition and mild cognitive impairment: Results from ADNI. *PLoS One*. 2019;14(2):e0212435. doi:10.1371/journal.pone.0212435

11. Harvey PD. Domains of cognition and their assessment. *Dialogues Clin Neurosci*. 2019;21(3):227-237. doi:10.31887/DCNS.2019.21.3/pharvey

12. Tucker-Drob EM. Global and domain-specific changes in cognition throughout adulthood. *Dev Psychol*. 2011;2(47):331-343. doi:10.1037/a0021361

13. Crossley M, D'Arcy C, Rawson NSB. Letter and category fluency in community-dwelling Canadian seniors: A comparison of normal participants to those with dementia of the Alzheimer or vascular type. *J Clin Exp Neuropsychol*. 1997;19(1):52-62. doi:10.1080/01688639708403836

14. Anderson TJ, MacAskill MR. Eye movements in patients with neurodegenerative disorders. *Nat Rev Neurol*. 2013;9(2):74-85. doi:10.1038/nrneurol.2012.273

15. MacPherson SE, Allerhand M, Cox SR, Deary IJ. Individual differences in cognitive processes underlying Trail Making Test-B performance in old age: The Lothian Birth Cohort 1936. *Intelligence*. 2019;75(March):23-32. doi:10.1016/j.intell.2019.04.001

16. Chen P, DeKosky ST, Ganguli M. Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*. 2000;55(12):1847-1853. doi:10.1212/wnl.55.12.1847

17. Estévez-González A, Kulisevsky J, Boltes A, Otermín P, García-Sánchez C. Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer's disease: Comparison with mild cognitive impairment and normal aging. *Int J Geriatr Psychiatry*. 2003;18(11):1021-1028. doi:10.1002/gps.1010

18. Landau SM, Harvey D, Madison CM, et al. Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology*. 2010;75(3):230-238. doi:10.1212/WNL.0b013e3181e8e8b8

19. Schoenberg MR, Dawson KA, Duff K, Patton D, Scott JG, Adams RL. Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch Clin Neuropsychol*. 2006;21(7):693-703. doi:10.1016/j.acn.2006.06.010

20. Lezak MD. Memory I: Tests. In: *Neuropsychological Assessment*. Oxford University Press; 1995:429-489.

21. Crawford JR, Stewart LE, Moore JW. Demonstration of savings on the AVLT and development of a parallel form. *J Clin Exp Neuropsychol*. 1989;11(6):975-981. doi:10.1080/01688638908400950

22. Blachstein H, Vakil E. Verbal learning across the lifespan: an analysis of the components of the learning curve. *Aging, Neuropsychol Cogn*. 2015;23(2):133-153. doi:10.1080/13825585.2015.1063579

23. Tuokko H, Griffith LE, Simard M, Taler V. Cognitive measures in the Canadian Longitudinal Study on Aging. *Clin Neuropsychol*. 2016;31(1):233-250. doi:10.1080/13854046.2016.1254279

24. Ettinger, U., Klein C. Eye Movements. In: Reuter M, Montag C, eds. *Neuroeconomics*. Springer-Verlag; 2016:481-502.

25. Mirsky JB, Heuer HW, Jafari A, et al. Anti-saccade performance predicts executive function and brain structure in normal elders. *Cogn Behav Neurol*. 2011;24(2):50-58. doi:10.1097/WNN.0b013e318223f6c6

26. Hutton SB. Cognitive control of saccadic eye movements. *Brain Cogn*. 2008;68(3):327-340. doi:10.1016/j.bandc.2008.08.021

27. Luna B, Velanova K, Geier CF. Development of eye-movement control. *Brain Cogn*. 2008;68(3):293-308. doi:10.1016/j.bandc.2008.08.019

28. Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. Voice-selective areas in human auditory cortex. *Nature*. 2000;403:309-312. doi:10.1038/35002078

29. Pernet CR, McAleer P, Latinus M, et al. The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage*. 2015;119:164-174. doi:10.1016/j.neuroimage.2015.06.050

30. Agus TR, Paquette S, Suied C, Pressnitzer D, Belin P. Voice selectivity in the temporal voice area despite matched low-level acoustic cues. *Sci Rep*. 2017;7(1):1-7. doi:10.1038/s41598-017-11684-1

31. Young AW, Frühholz S, Schweinberger SR. Face and voice perception: understanding commonalities and differences. *Trends Cogn Sci*. 2020;24(5):398-410. doi:10.1016/j.tics.2020.02.001

32. Kanwisher N, McDermott J, Chun MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci*. 1997;17(11):4302-4311. doi:10.1523/JNEUROSCI.17-11-04302.1997

33. Epstein R, Kanwisher N. A cortical representation of the local visual environment. *Nature*. 1998;392(6676):598-601. doi:10.1038/33402

34. Gazzaley A, Cooney JW, Rissman J, D'Esposito M. Top-down suppression deficit underlies working memory impairment in normal aging. *Nat Neurosci*. 2005;8(10):1298-1300. doi:10.1038/nn1543

35. Collins JA, Olson IR. Beyond the FFA: The role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*. 2014;61(1):65-79. doi:10.1016/j.neuropsychologia.2014.06.005

36. Raichle ME, MacLeod AM, Snyder AZ, Powers WJ, Gusnard DA, Shulman GL. A default mode of brain function. *PNAS*. 2001;98(2):676-682. doi:10.1073/pnas.98.2.676

37. Daselaar SM, Prince SE, Cabeza R. When less means more: deactivations during encoding that predict subsequent memory. *Neuroimage*. 2004;23(3):921-927. doi:10.1016/j.neuroimage.2004.07.031

38. Huijbers W, Schultz AP, Vannini P, et al. The encoding/retrieval flip: Interactions between memory performance and memory stage and relationship to intrinsic cortical networks. *J Cogn Neurosci*. 2013;25(7):1163-1179. doi:10.1162/jocn_a_00366

39. Krieger-Redwood K, Jefferies E, Karapanagiotidis T, et al. Down but not out in posterior cingulate cortex: Deactivation yet functional coupling with prefrontal cortex during demanding semantic cognition. *Neuroimage*. 2016;141:366-377. doi:10.1016/j.neuroimage.2016.07.060

40. Sperling RA, Bates JF, Chua EF, et al. fMRI studies of associative encoding in young and elderly controls and mild Alzheimer's disease. *J Neurol Neurosurg Psychiatry*. 2003;74(1):44-50. doi:10.1136/jnnp.74.1.44

41. Damoiseaux JS, Huijbers W. The complementary value of task-evoked and resting-state functional imaging: a commentary on Campbell and Schacter (2016). *Lang Cogn Neurosci*. 2017;32(6):678-680. doi:10.1080/23273798.2016.1248452

42. Lv H, Wang Z, Tong E, et al. Resting-state functional MRI: Everything that nonexperts have always wanted to know. *Am J Neuroradiol*. 2018;39(8):1390-1399. doi:10.3174/ajnr.A5527

43. Campbell KL, Schacter DL. Ageing and the resting state: is cognition obsolete? *Lang Cogn Neurosci*. 2017;32(6):661-668. doi:10.1080/23273798.2016.1227858

44. Huijbers W, Van Dijk KRA, Boenniger MM, Stirnberg R, Breteler MMB. Less head motion during MRI under task than resting-state conditions. *Neuroimage*. 2017;147:111-120. doi:10.1016/j.neuroimage.2016.12.002

45. McDonough IM, Festini SB, Wood MM. Risk for Alzheimer's disease: A review of long-term episodic memory encoding and retrieval fMRI studies. *Ageing Res Rev*. 2020;62(May):101133. doi:10.1016/j.arr.2020.101133

46. Tavor I, Parker Jones O, Mars RB, Smith SM, Behrens TE, Jbabdi S. Task-free MRI predicts individual differences in brain activity during task performance. *Science (80-)*. 2016;352(6282):216-220. doi:10.1126/science.aad8127

47. den Haan J, Verbraak FD, Visser PJ, Bouwman FH. Retinal thickness in Alzheimer's disease: A systematic review and meta-analysis. *Alzheimer's Dement Diagnosis, Assess Dis Monit*. 2017;6:162-170. doi:10.1016/j.dadm.2016.12.014

48. Ko F, Muthy ZA, Gallacher J, et al. Association of retinal nerve fiber layer thinning with current and future cognitive decline: A study using optical coherence tomography. *JAMA Neurol*. 2018;75(10):1198-1205. doi:10.1001/jamaneurol.2018.1578

49. Khawaja AP, Chan MPY, Yip JLY, et al. Retinal nerve fiber layer measures and cognitive function in the EPIC-Norfolk cohort study. *Investig Ophthalmol Vis Sci*. 2016;57(4):1921-1926. doi:10.1167/iovs.16-19067

50. Mutlu U, Colijn JM, Ikram MA, et al. Association of retinal neurodegeneration on optical coherence tomography with dementia: A population-based study. *JAMA Neurol*. 2018;75(10):1256-1263. doi:10.1001/jamaneurol.2018.1563

51. Clay OJ, Edwards JD, Ross LA, et al. Visual function and cognitive speed of processing mediate age-related decline in memory span and fluid intelligence. *J Aging Health*. 2009;21(4):547-566. doi:10.1177/0898264309333326

52. Mauschitz MM, Bonnemaijer PWM, Diers K, et al. Systemic and ocular determinants of peripapillary retinal nerve fiber layer thickness measurements in the European Eye Epidemiology (E3) population. *Ophthalmology*. 2018;125(10):1526-1536. doi:10.1016/j.ophtha.2018.03.026

53. Levin ED, McClernon FJ, Rezvani AH. Nicotinic effects on cognitive function: Behavioral characterization, pharmacological specification, and anatomic localization. *Psychopharmacology (Berl)*. 2006;184(3-4):523-539. doi:10.1007/s00213-005-0164-7

54. Sabia S, Elbaz A, Dugravot A, et al. Impact of smoking on cognitive decline in early old age: The Whitehall II cohort study. *Arch Gen Psychiatry*. 2012;69(6):627-635. doi:10.1001/archgenpsychiatry.2011.2016

55. Stalder T, Steudte-Schmiedgen S, Alexander N, et al. Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. *Psychoneuroendocrinology*. 2017;77(16):261-274. doi:10.1016/j.psyneuen.2016.12.017

56. Jansen P, Dahmen-Zimmer K, Kudielka BM, Schulz A. Effects of karate training versus mindfulness training on emotional well-being and cognitive performance in later life. *Res Aging*. 2017;39(10):1118-1144. doi:10.1177/0164027516669987

57. McLennan SN, Ihle A, Steudte-Schmiedgen S, Kirschbaum C, Kliegel M. Hair cortisol and cognitive performance in working age adults. *Psychoneuroendocrinology*. 2016;67:100-103. doi:10.1016/j.psyneuen.2016.01.029

58. Pulopulos MM, Hidalgo V, Almela M, Puig-Perez S, Villada C, Salvador A. Hair cortisol and cognitive performance in healthy older people. *Psychoneuroendocrinology*. 2014;44:100-111. doi:10.1016/j.psyneuen.2014.03.002

59. Feeney J, O'Halloran AM, Kenny RA. The association between hair cortisol, hair cortisone, and cognitive function in a population-based cohort of older adults: Results from the Irish Longitudinal Study on Ageing. *Journals Gerontol Ser A gly258*. 2020;75(2):257-265. https://doi.org/10.1093/gerona/gly258

60. Boccardi M, Boccardi V. Psychological wellbeing and healthy aging: Focus on telomeres. *Geriatr*. 2019;4(1):1-10. doi:10.3390/geriatrics4010025

61. Aschbacher K, O'Donovan A, Wolkowitz OM, Dhabhar FS, Su Y, Epel E. Good stress, bad stress and oxidative stress: Insights from anticipatory cortisol reactivity. *Psychoneuroendocrinology*. 2013;38(9):1698-1708. doi:10.1016/j.psyneuen.2013.02.004

62. Epel ES, Blackburn EH, Lin J, et al. Accelerated telomere shortening in response to life stress. *Proc Natl Acad Sci U S A*. 2004;101(49):17312-17315. doi:10.1073/pnas.0407162101

63. Lupien SJ, Maheu F, Tu M, Fiocco A, Schramek TE. The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain Cogn*. 2007;65(3):209-237. doi:10.1016/j.bandc.2007.02.007

64. Korten NCM, Comijs HC, Penninx BWJH, Deeg DJH. Perceived stress and cognitive function in older adults: Which aspect of perceived stress is important? *Int J Geriatr Psychiatry*. 2017;32(4):439-445. doi:10.1002/gps.4486

65. Mauschitz MM, Holz FG, Finger RP, Breteler MMB. Determinants of macular layers and optic disc characteristics on SD-OCT: The Rhineland Study. *Transl Vis Sci Technol*. 2019;8(3). doi:10.1167/tvst.8.3.34

66. Oumohand SE, Ward DD, Boenniger MM, Merten N, Kirschbaum C, Breteler MMB. Perceived stress but not hair cortisol concentration is related to adult cognitive performance. *Psychoneuroendocrinology*. 2020;121:104810. doi:10.1016/j.psyneuen.2020.104810

67. Ward DD, Mauschitz MM, Bönniger MM, Merten N, Finger RP, Breteler MMB. Peripapillary retinal nerve fiber layer thickness and cognitive function across the adult lifespan in the Rhineland Study. *Rev Epidemiol Sante Publique*. 2018;66(Supplement 5):S238. https://doi.org/10.1016/j.respe.2018.05.018

68. Wilson RS, Yu L, Lamar M, Schneider JA, Boyle PA, Bennett DA. Education and cognitive reserve in old age. *Neurology*. 2019;92(1-10):10.1212/WNL.0000000000007036. doi:10.1212/wnl.0000000000007036

69. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurol*. 2014;71(8):961-970. doi:10.1001/jamaneurol.2014.803

6 Supplementary material

6.1 Supplementary material to “A functional MRI paradigm for efficient mapping of memory encoding across sensory conditions”

Supplementary material online available under:
https://www.frontiersin.org/articles/10.3389/fnhum.2020.591721/full#supplementary-material or https://www.rheinland-studie.de/data-code/boenniger2020/

Task scripts available under:
https://www.rheinland-studie.de/data-code/boenniger2020/

(Alternative direct link to the supplementary material: https://www.rheinland-studie.de/fileadmin/Dateiliste/Dokumente/Boenniger2021-MSES-Supplement.pdf

Alternative direct link to the task scripts: https://www.rheinland-studie.de/fileadmin/Dateiliste/Dokumente/Boenniger-2020.zip)

6.2 Supplementary material to “Ten German versions of Rey’s Auditory Verbal Learning Test: Age and sex effects in 4,000 adults of the Rhineland Study”

Supplementary material online available under:
https://www.rheinland-studie.de/data-code/boenniger2020/

or under the direct link: https://www.rheinland-studie.de/fileadmin/Dateiliste/Dokumente/Boenniger2021-AVLT-Supplement.pdf

6.3 Appendix to "Association of retinal layer measurements and adult cognitive function - A population-based study"

Appendix. Authors

Name	Location	Contribution
David D. Ward, PhD	German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany	Designed and conceptualized study, analyzed and interpreted the data, performed the statistical analysis, drafted the manuscript
Matthias M. Mauschitz, MD, PhD	German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany	Designed and conceptualized study, interpreted the data, revised the manuscript
Meta M. Bönniger, MSc	German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany	Interpreted the data, revised the manuscript
Natascha Merten, PhD	Department of Population Health Sciences, School of Medicine and Public Health, University of Wisconsin–Madison	Interpreted the data, revised the manuscript
Robert P. Finger, MD, PhD	Department of Ophthalmology, Faculty of Medicine, University of Bonn, Bonn, Germany	Designed and conceptualized study, revised the manuscript
Monique M.B. Breteler, MD, PhD	German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany	Designed and conceptualized study, interpreted the data, revised the manuscript, obtained funding, supervised the study

Table A.1. Effects of subjective and physiological measures of chronic stress on cognitive performance

	Crystallized intelligence		Executive functioning		Processing speed		Verbal episodic memory		Working memory	
	β (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*
HCC	0.034 (-.020, .088)	.212	-0.029 (-.077, .020)	.245	-0.007 (-.052, .038)	.749	-0.032 (-.079, .015)	.176	-0.035 (-.085, .014)	.160
PSS	-0.072 (-.126, -.019)	.008	-0.101 (-.221, -.067)	<.001	-0.054 (-.137, .010)	.018	-0.060 (-.122, .030)	.011	-0.129 (-.097, .063)	<.001
PSS-H	-0.049 (-.127, .050)	.078	-0.074 (-.201, -.045)	.003	-0.039 (-.118, .030)	.087	-0.056 (-.103, .049)	.019	-0.120 (-.082, .078)	<.001
PSS-SE	0.093 (-.056, .117)	.001	0.120 (.044, .195)	<.001	0.065 (-.004, .140)	.005	0.052 (-.018, .131)	.029	0.111 (-.054, .103)	<.001

Standardized associations and 95 % confidence intervals of chronic stress measures and cognitive domain performance as determined by multiple linear regressions models. Values represent change in cognitive domain z-score per one standard deviation increase in stress measure after adjusting for age, sex and education. All models were based on n = 1,098 participants who provided a hair sample.
Abbreviations: CI: confidence interval, HCC: log hair cortisol concentration, PSS: Perceived stress, PSS-H: perceived helplessness, PSS-SE: perceived self-efficacy

Table A.2. Interaction effects of subjective and physiological chronic stress measures with age and sex on cognitive performance

	Crystallized intelligence		**Executive functioning**		**Processing speed**		**Verbal episodic memory**		**Working memory**	
Interaction	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*
HCC*sex	-0.033 (-0.172, 0.107)	.645	0.052 (-0.073, 0.177)	.418	0.056 (-0.060, 0.173)	.344	0.038 (-0.083, 0.159)	.535	0.059 (-0.069, 0.187)	.364
PSS*sex	0.021 (-0.110, 0.151)	.754	-0.146 (-0.262, -0.030)	.013	-0.096 (-0.204, 0.013)	.085	-0.067 (-0.180, 0.046)	.244	0.038 (-0.080, 0.157)	.524
PSS-H*sex	-0.003 (-0.137, 0.132)	.969	-0.144 (-0.264, -0.024)	.018	-0.066 (-0.178, 0.046)	.248	-0.054 (-0.170, 0.062)	.360	0.040 (-0.082, 0.162)	.517
PSS-SE*sex	-0.059 (-0.183, 0.066)	.354	0.101 (-0.010, 0.212)	.075	0.108 (0.004, 0.212)	.042	0.066 (-0.042, 0.174)	.230	-0.031 (-0.144, 0.083)	.595

Continuation Table A.2.

	Crystallized intelligence		Executive functioning		Processing speed		Verbal episodic memory		Working memory	
Interaction	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*	*β* (95 % CI)	*p*
HCC*age	-0.018 (-0.058, 0.022)	.378	0.016 (-0.020, 0.052)	.388	-0.011 (-0.044, 0.023)	.530	0.001 (-0.034, 0.035)	.964	0.010 (-0.026, 0.047)	.577
PSS*age	0.017 (-0.021, 0.056)	.380	0.012 (-0.023, 0.047)	.492	-0.017 (-0.050, 0.015)	.303	-0.025 (-0.059, 0.009)	.143	-0.038 (-0.073, -0.002)	.037
PSS-H*age	0.029 (-0.010, 0.068)	.146	0.034 (-0.001, 0.069)	.055	-0.010 (-0.042, 0.023)	.553	-0.012 (-0.045, 0.022)	.493	-0.020 (-0.056, 0.015)	.258
PSS-SE*age	-0.002 (-0.040, 0.036)	.923	0.015 (-0.019, 0.049)	.372	0.017 (-0.015, 0.049)	.291	0.032 (-0.001, 0.065)	.056	0.039 (0.005, 0.074)	.026

Standardized associations and 95 % confidence intervals of chronic stress measures and cognitive domain performance as determined by multiple linear regressions models. For interactions with sex, values represent the difference in the association of stress measure and cognitive domain z-score between women and men (reference = women). For interactions with age, values represent the change in the association of chronic stress measure and cognitive domain z-score per 10 years of age. All associations were adjusted for age, sex, and education. All models were based on n = 1,098 participants who provided a hair sample.

Abbreviations: CI: confidence interval, HCC: log hair cortisol concentration, PSS: Perceived stress, PSS-H: perceived helplessness, PSS-SE: perceived self-efficacy

7 Acknowledgement

A thesis is never the work of just one person alone. I would therefore like to thank a few people who made this thesis possible and who helped me to keep going, even when times were difficult.

First and foremost, I would like to thank my supervisors and members of the Thesis Committee. In particular, Monique M.B. Breteler for her supervision and challenging and constructive feedback. Michael Wagner, Carole Dufouil and Sterling C. Johnson, I would like to thank for their time, valuable comments and suggestions both inside and outside the Thesis Committee meetings. I would also like to thank Ulrich Ettinger and Ahmad Aziz, who always gave me valuable advice and patiently responded to my questions. Special thanks also go to Willem Huijbers, who supported me from the Master's thesis onwards and encouraged me to start the PhD in the first place.

Furthermore, I would like to thank all my friends and colleagues in the Rhineland Study team. Setting up and conducting such a large study only works with excellent teamwork.

Thanks to the study assistants, especially the cognition team with whom I worked the most, for the pleasant and constructive exchange and your great work in collecting the data. Your work and that of the other study assistants made my projects possible to a large extent. Special thanks here to Sabine Neuhaus and Anja Sperling.

Thanks to our IT team, especially Christof Meigen, André Medek and Thomas H. Schmidt, who gave me special support in processing and cleaning the data and were always there to help and advise me.

Thanks to my post-doctoral and PhD colleagues, who have been very supportive over the years with advice both privately and professionally. Special thanks to Sadia E. Oumohand Fabienne Fox and Valerie Lohner for their friendship, their continuous support and their feedback on the thesis. David D. Ward, Nersi Alaeddin, Elvire Landstra, Matthias M. Mauschitz and Gökhan Pehlivan for their friendship.

Finally, and most especially, I would like to thank my friends and family. Your continuous and unconditional support and belief in me has kept me going.

Special thanks also to Frederik, who was the sunshine during my lunch breaks at the home office.

8 List of publications and conference presentations

Peer-reviewed publications

Boenniger, M.M., Staerk C., Coors, A., Huijbers, W., Ettinger, U., Breteler, M.M.B. (2021). Ten German versions of Rey's Auditory Verbal Learning Test: Age and sex effects in 4,000 adults of the Rhineland Study. *Journal of Clinical and Experimental Neuropsychology* (accepted)

Merten, N., **Boenniger, M.M.**, Herholz, S.C., Breteler, M.M.B. (2021). The associations of hearing sensitivity and different cognitive functions with perception of speech-in-noise. *Ear and Hearing* (accepted)

Coors, A., Imtiaz, M.-A., **Boenniger, M.M.**, Aziz, A., Breteler, M.M.B, Ettinger, U. (2021). Polygenic risk scores for schizophrenia are associated with oculomotor endophenotypes. *Psychological Medicine*, 1-9. https://doi.org/10.1017/S0033291721003251

Boenniger, M.M., Diers, K., Herholz, S., Shahid, M., Stöcker, T., Breteler, M.M.B., Huijbers, W. (2021). A functional MRI paradigm for efficient mapping of memory encoding across sensory conditions. *Frontiers in Human Neuroscience*, 14(591721). https://doi.org/10.3389/fnhum.2020.591721

Oumohand, S.E., Ward, D.D., **Boenniger, M.M.**, Merten, N., Kirschbaum, C., Breteler, M.M.B. (2020). Perceived stress but not hair cortisol concentration is related to adult cognitive performance. *Psychoneuroendocrinology*, 121, 104810. https://doi.org/10.1016/j.psyneuen.2020.104810

Ward, D.D., Mauschitz, M.M., **Bönniger, M.M.**, Merten, N., Finger, R.P., Breteler, M.M.B. (2020). Association of retinal layer measurements and adult cognitive function: a population-based study. *Neurology*, 95(9), e1144-e1152. https://doi.org/10.1212/WNL.0000000000010146

Huijbers, W., Dijk, K.R., **Boenniger, M.M.**, Stirnberg, R., & Breteler, M.M.B. (2017). Less head motion during MRI under task than resting-state conditions. *NeuroImage, 147*, 111-120. https://doi.org/10.1016/j.neuroimage.2016.12.002

Conference presentations

09/2018 **13. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi), Bremen, Germany** (talk)
Bönniger, M.M., Merten, N., Ward, D.D., Breteler, M.M.B. (2018). Sex differences in cognitive performance across the adult lifespan in the Rhineland Study. V-02-M-03

07/2018 **European Congress of Epidemiology, Lyon, France** (poster)
Boenniger, M.M., Merten, N., Ward, D.D., Breteler, M.M.B. (2018). Sex differences across age in cognitive tests in the Rhineland Study. Revue d'Épidémiologie et de Santé Publique, 66(Supplement 5), S281. https://doi.org/10.1016/j.respe.2018.05.119

07/2017 **Alzheimer's Association International Conference (AAIC) – Alzheimer's Imaging Consortium (AIC) Preconference, London, UK** (poster)
Boenniger, M., Herholz, S., Stoecker, T., Breteler, M., & Huijbers, W. (2017a). A parallel mixed design for mapping multisensory and memory-related fMRI activity in a population study. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 13(7), P1080–P1081. https://doi.org/10.1016/j.jalz.2017.06.1549

07/2017 **Alzheimer's Association International Conference (AAIC), London, UK** (poster)
Boenniger, M., Herholz, S., Stoecker, T., Breteler, M., & Huijbers, W. (2017b). A parallel mixed design for mapping multisensory and memory-related fMRI activity in a population study. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 13(7), P123–P124. https://doi.org/10.1016/j.jalz.2017.06.1549