

# AEROBIC EXERCISE FOR THE PROMOTION OF HEALTHY AGING: CHANGES IN BRAIN STRUCTURE ASSESSED WITH NEW METHODS

## **A Dissertation**

Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Doctor rerum naturalium  
(Dr. rer. nat.)

to the Department of Education and Psychology  
of Freie Universität Berlin



by

**Sarah Elisabeth Polk**

M.Sc. in Social, Cognitive, and Affective Neuroscience  
B.A. in Psychology

Berlin, 2022



International Max Planck  
Research School  
on the Life Course

**MAX PLANCK INSTITUTE**  
FOR HUMAN DEVELOPMENT



**First reviewer:**

Prof. Dr. Ulman Lindenberger  
Max Planck Institute for Human Development, Berlin

**Second reviewer:**

Univ.-Prof. Dr. med. Felix Blankenburg  
Freie Universität Berlin

**Dissertation defense:** October 17<sup>th</sup>, 2022

## **Declaration of authorship / Eidesstattliche Erklärung**

I hereby declare in lieu of oath

- that I have written this dissertation independently and without unauthorized assistance,
- that I have not submitted this dissertation to any other university and that I do not hold a doctoral degree in the subject of psychology, and
- that I am aware of the doctoral regulations for the degree of Dr. rer. nat./Ph. D. in the Department of Education and Psychology at the Freie Universität Berlin dated August 8<sup>th</sup>, 2016 (official gazette of the Freie Universität Berlin 35/2016).

Hiermit erkläre ich an Eides statt,

- dass ich die vorliegende Arbeit selbstständig und ohne unerlaubte Hilfe verfasst habe,
- dass ich die Dissertation an keiner anderen Universität eingereicht habe und keinen Doktorgrad in dem Promotionsfach Psychologie besitze und,
- dass mir die Promotionsordnung zum Dr. rer. nat./Ph. D. des Fachbereichs Erziehungswissenschaft und Psychologie der Freien Universität Berlin vom 8. August 2016 (Amtsblatt der Freien Universität Berlin 35/2016) bekannt ist.

Berlin, 2022

Sarah E. Polk



## Acknowledgements

This dissertation was completed with data from the Mechanisms and Sequential Progression of Plasticity project in the Center for Lifespan Psychology (LIP) at the Max Planck Institute for Human Development, Berlin. During my time working on this dissertation, I was a pre-doctoral fellow in the International Max Planck Research School on the Life Course (LIFE).

This dissertation could not have been completed without a great deal of support from those around me. In particular, I would like to thank Ulman Lindenberger for the opportunity to expand my scientific knowledge and practical skills through my experience as a pre-doctoral researcher on the AKTIV project with its invaluable rich dataset, as well as for his constructive feedback and comprehensive perspective on all my work leading up to this dissertation. I would also like to thank my day-to-day supervisor, Sandra Düzel, for all her support and guidance throughout the entire process, from the preparation of the initial proposal all the way up to this final product. Thank you for the frequent brainstorming sessions and the occasional coffee and cake, and I wish you all the best in your next steps.

I would also like to thank the AKTIV team, without whom this project would not have been possible. My deepest gratitude goes to Elisabeth Wenger for her admirable leadership and close guidance on this project. Thank you for your advice on all things science and academia, and for giving me the opportunity to work with the AKTIV data. Thank you to Maike Kleemeyer for the countless hours spent preprocessing MRI data and for taking the time to walk me through numerous data processing pipelines. Thank you to Ylva Köhncke and Andreas Brandmaier for all the guidance regarding the statistics and modeling. Thank you to Nils Bodammer for the feedback on all things neuroimaging. Thank you to Simone Kühn for the constructive and practical feedback and for serving on my thesis advisory committee. And a big thank you goes out to everyone else who was involved in the AKTIV Study, including Carola Misgeld, Johanna Porst, and Bernd Wolfarth at the Charité, Michael Krause and Sebastian Schröder for the computational and technical support, Kirsten Becker and Anke Schepers-Klingebiel for the organizational support, Sonali Beckmann, Nadine Taube, Thomas Feg, and Davide Santoro on the MRI team, and all the other student assistants in the gone (but not forgotten) Room 12.

A special thanks goes to Steve Boker and Timo von Oertzen for inviting the LIFE fellows to participate in their courses, which was integral to my understanding of statistics and especially structural equation modeling.

Thank you to all the other PhD students in LIFE and LIP for your companionship and solidarity, both during the first year of my dissertation, when in-person coffee breaks and lunches were a highlight of daily office life, but also during the last two years, full of Zoom meetings and the occasional outdoor meet-up.

Finally, this dissertation would not have been possible without the support of those closest to me. Thank you to my parents for your unlimited support, thank you to Rachael, Lydia, and Aaron for your unconditional sibling love, and thank you to Antonin for your patience and support.



## List of included works

This doctoral dissertation is based on the following original works:

1. Wenger, E., **Polk, S. E.**, Kleemeyer, M. M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for MT and PD but attenuated for R1 and R2\* in healthy young adults. *Human Brain Mapping*. <https://doi.org/10.1002/hbm.25870>
2. **Polk, S. E.**, Kleemeyer, M. M., Köhncke, Y., Brandmaier, A. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Wenger, E., & Düzel, S. (2022). Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2022.852737>
3. **Polk, S. E.**, Kleemeyer, M. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Düzel, S., & Wenger, E. (2022). *Aerobic exercise is associated with region-specific changes in volumetric, tensor-based, and fixel-based measures of white matter integrity in healthy older adults* [Manuscript under revision].





## List of abbreviations

ACG	Active control group
AKTIV	Aktives Altern für Körper und Geist [active aging for body and mind]
ASL	Arterial spin labeling
CBF	Cerebral blood flow
CFI	Comparative fit index
CoV	Coefficient of variation
CPET	Cardiopulmonary exercise testing
CSD	Constrained spherical decomposition
DSST	Digit Symbol Substitution task
DW(I)	Diffusion weighted (imaging)
EG	Exercise group
EPI	Echo planar imaging
FA	Fractional anisotropy
FC	Fiber cross-section
FD	Fiber density
FDC	Fiber density and cross-section
FDR	False discovery rate
FIML	Full information maximum likelihood
FLASH	Fast low angle shot
FOD	Fiber orientation distribution
FWHM	Full-width half-maximum
GM	Gray matter
GMV	Gray matter volume
GRE	Gradient echo
ICC	Intra-class correlation
ICED	Intra-class effect decomposition
LCSM	Latent change score model
LRT	Likelihood-ratio test
MD	Mean diffusivity
MPM	Multiparameter map(ping)
MPRAGE	Magnetization prepared gradient echo
MR(I)	Magnetic resonance (imaging)

MT	Magnetization transfer (saturation)
MTMM	Multi-trait multi-method
MWT-A	Mehrfachwahl-Wortschatz-Test A [multiple-choice vocabulary test A]
PD	Proton density
RF	Radio frequency
RMSEA	Root mean square error of approximation
ROI	Region of interest
R1	Longitudinal relaxation rate
R2*	Effective transverse relaxation rate
SEM	Structural equation model(ing)
T1/T2/T3	Time point 1/2/3
Var D	Day-specific variance
Var E	Residual-error variance
Var <sub>eff</sub>	Effective variance
Var S	Session-specific variance
Var T	True-score variance
VBM	Voxel-based morphometry
VO <sub>2</sub> peak	Peak oxygen uptake (mL/kg/min)
WM	White matter
WMV	White matter volume

## Summary

As the proportion of older individuals in the population increases, so does the scientific concern surrounding age-related deterioration of brain tissue and related cognitive decline. One modifiable lifestyle factor of interest in the pursuit to slow or even reverse age-related brain atrophy is aerobic exercise. A number of studies have already demonstrated that aerobic exercise in older age can induce maintenance (i.e., reduction of loss) of both gray and white matter volume, particularly in the frontal regions of the brain, which are vulnerable to shrinkage in older age. Other magnetic resonance imaging (MRI)-based techniques, such as quantitative MRI and diffusion-weighted MRI, have been used to measure age-related deterioration of gray and white matter integrity in both voxel-wise analyses as well as on the latent level, but whether these negative changes can be ameliorated through exercise has yet to be shown. The current dissertation includes three papers which used a number of both established and novel MRI-based metrics to quantify changes in brain tissue integrity resulting from aging, as well as to investigate whether these changes can be ameliorated through aerobic exercise.

In Paper I (Wenger, Polk et al., 2022), we tested the reliability of quantitative MRI measures, namely longitudinal relaxation rate, effective transverse relaxation rate, proton density, and magnetization transfer saturation, by measuring them in a two-day, four-session design with repositioning in the scanner. Using the intra-class effect decomposition model, we found that magnetization transfer saturation could reliably detect individual differences, validating its use to investigate changes in brain structure longitudinally, as well as correlations with other variables of interest, such as change in cardiovascular fitness.

In Paper II (Polk, Kleemeyer, Köhncke et al., 2022), we tested the effects of aerobic exercise on a latent factor of gray-matter structural integrity, comprising observed measures of gray-matter volume, magnetization transfer saturation, and mean diffusivity, in regions of interest that have previously shown volumetric effects of aerobic exercise. We found that gray-matter structural integrity was maintained in frontal and midline regions, and that change in gray-matter structural integrity in the right anterior cingulate cortex was positively correlated with change in cardiovascular fitness within exercising participants. These results suggest a causal relationship between aerobic exercise, cardiovascular fitness, and gray-matter structural integrity in this region.

In Paper III (Polk, Kleemeyer, Bodammer et al., 2022), we tested the effects of aerobic exercise on white matter integrity, measured with both established and recently developed metrics. We were able to replicate findings from a previous study on the effects of aerobic exercise on white matter volume, and we also found change-change correlations between white matter volume and cardiovascular fitness as well as between white matter volume and performance on a test of perceptual speed. We also found unexpected exercise-induced changes in the diffusion weighted imaging-derived metrics of fractional anisotropy, mean diffusivity, fiber density, and fiber density and cross-section. Specifically, we found increases (or decreases in the case of mean diffusivity) within control participants and decreases (or increases in mean diffusivity) in exercisers. Furthermore, we found that

percent change in fiber density and fiber density and cross-section correlated negatively with percent change in both cardiovascular fitness and cognitive performance. This casts doubt on the functional interpretation of these measures and suggests that the “more is better” principle may not be universally applicable when investigating age-related and exercise-induced changes in white matter integrity.

In sum, this dissertation showed that regular at-home aerobic exercise, which may be more accessible for older individuals than supervised exercise, can be an effective tool to ameliorate age-related decreases in a latent measure of gray-matter structural integrity as well as white matter volume. It also illuminated potential limitations of other measures of white matter integrity in the context of aging and aerobic exercise, and calls for further research into these novel measures, especially when considering functional outcomes such as cognitive performance.

## Zusammenfassung

Mit dem zunehmenden Anteil älterer Menschen an der Bevölkerung wächst auch die wissenschaftliche Besorgnis über den altersbedingten Abbau von Hirngewebe und dem damit verbundenen kognitiven Abbau. Ein modifizierbarer Lifestyle-Faktor, der in dem Bestreben, die altersbedingte Hirnatrophie zu verlangsamen oder sogar umzukehren, von Interesse ist, ist aerobes Training. Eine Reihe von Studien hat bereits gezeigt, dass aerobes Fitnessstraining im Alter den Erhalt (d. h. die Verringerung des Verlusts) des Volumens sowohl der grauen als auch der weißen Substanz bewirken kann, insbesondere in den Frontalregionen des Gehirns, die anfällig für altersbedingten Abbau sind. Andere auf der Magnetresonanztomographie (MRT) basierte Techniken, wie die quantitative MRT und die diffusionsgewichtete MRT, wurden eingesetzt, um die altersbedingte Verschlechterung der Integrität der grauen und weißen Substanz sowohl in voxelbasierten Analysen als auch auf latenter Ebene zu messen, aber es wurde noch nicht festgestellt, ob diese negativen Veränderungen durch Training gemildert werden können. Die vorliegende Dissertation umfasst drei Arbeiten, die darauf abzielen, die Auswirkungen eines häuslichen Fitnessstrainings bei älteren Erwachsenen zu untersuchen, wobei eine Reihe etablierter und neuartiger MRT-basierter Metriken zur Quantifizierung altersbedingter Veränderungen der Integrität des Hirngewebes eingesetzt werden, um zu untersuchen, ob diese Veränderungen durch aerobes Fitnessstraining gemildert werden können.

In Paper I (Wenger, Polk et al., 2022) wurde die Zuverlässigkeit quantitativer MRT-Messungen getestet, nämlich die longitudinale Relaxationsrate, die effektive transversale Relaxationsrate, die Protonendichte und die Magnetisierungstransfersättigung. Wir verwendeten ein Studiendesign mit insgesamt vier Sitzungen, jeweils zwei an zwei aufeinanderfolgenden Tagen, mit dem wir die Reliabilität dieser Maße während Repositionierung untersuchten. Unter Verwendung des Intra-Class Effect Decomposition-Modells wurde festgestellt, dass die Magnetisierungstransfersättigung zuverlässig individuelle Unterschiede erkennen kann, was ihre Verwendung zur Untersuchung von Veränderungen der Hirnstruktur im Längsschnitt sowie von Korrelationen mit anderen Variablen von Interesse, wie z. B. Veränderungen der kardiovaskulären Fitness, bestätigt.

In Paper II (Polk, Kleemeyer, Köhncke et al., 2022) testeten wir die Effekte von aerobem Fitnessstraining auf einen latenten Faktor der strukturellen Integrität der grauen Substanz, der gemessene Variablen des Volumens der grauen Substanz, der Magnetisierungstransfersättigung und der mittleren Diffusivität umfasst, in einer Reihe von Regionen, die zuvor volumetrische Auswirkungen von aerobem Fitnessstraining gezeigt haben. Es wurde festgestellt, dass die strukturelle Integrität der grauen Substanz in den Frontal- und Mittellinienregionen erhalten blieb und dass bei den trainierenden Teilnehmern die Veränderung der strukturellen Integrität der grauen Substanz im rechten anterioren Cinguli-Cortex positiv mit der Veränderung der kardiovaskulären Fitness im Verlauf der sechsmonatigen Intervention korreliert war, was auf einen kausalen Zusammenhang zwischen aerobem Fitnessstraining, kardiovaskulärer Fitness und struktureller Integrität der grauen Substanz in dieser Region hindeutet.

In Paper III (Polk, Kleemeyer, Bodammer et al., 2022) wurden die Auswirkungen von aerobem Fitnesstraining auf die Integrität der weißen Substanz untersucht, die mit einer Reihe etablierter und neu entwickelter Messmethoden gemessen wurde. Wir konnten die Ergebnisse einer früheren Studie über die Auswirkungen des aeroben Fitnesstrainings auf das Volumen der weißen Substanz replizieren und fanden außerdem Korrelationen zwischen der Veränderung des Volumens der weißen Substanz und der kardiovaskulären Fitness sowie zwischen dem Volumen der weißen Substanz und der Leistung in einem Test zur Wahrnehmungsgeschwindigkeit. Bei den aus der diffusionsgewichteten Bildgebung errechneten Metriken der fraktionierten Anisotropie, der mittleren Diffusivität, der Faserdichte sowie der Faserdichte und des Faserquerschnitts fanden wir unerwartete trainingsinduzierte Veränderungen, nämlich Zunahmen (oder Abnahmen im Fall der mittleren Diffusivität) bei den Kontrollteilnehmern und Abnahmen (oder Zunahmen der mittleren Diffusivität) bei den Trainierenden. Darüber hinaus wurde festgestellt, dass die prozentuale Veränderung der Faserdichte und der Faserdichte und des Querschnitts negativ mit der prozentualen Veränderung der kardiovaskulären Fitness und der kognitiven Leistung korreliert. Dies lässt Zweifel an der funktionellen Interpretation dieser Messwerte aufkommen und deutet darauf hin, dass das Prinzip "mehr ist besser" bei der Untersuchung altersbedingter und durch Fitnesstraining hervorgerufener Veränderungen der Integrität der weißen Substanz möglicherweise nicht universell anwendbar ist.

Zusammenfassend konnte in dieser Dissertation gezeigt werden, dass regelmäßiges häusliches Fitnesstraining, das für ältere Menschen möglicherweise zugänglicher ist als ein betreutes Fitnesstraining, ein wirksames Mittel sein kann, um altersbedingte Abnahmen eines latenten Maßes für die strukturelle Integrität der grauen Substanz sowie des Volumens der weißen Substanz zu verbessern. Die Studie zeigt auch mögliche Limitationen für die Anwendung neuerer Messmethoden der Integrität der weißen Substanz im Rahmen von fitness basierten Interventionen im höheren Altern auf und fordert weitere Studien zu diesen neuen Messgrößen, um insbesondere funktionelle Ergebnisse wie die kognitive Leistungsfähigkeit zu untersuchen.

# Contents

<b>Declaration of authorship / Eidesstattliche Erklärung</b> .....	<b>iii</b>
<b>Acknowledgements</b> .....	<b>v</b>
<b>List of included works</b> .....	<b>vii</b>
<b>List of abbreviations</b> .....	<b>ix</b>
<b>Summary</b> .....	<b>xi</b>
<b>Zusammenfassung</b> .....	<b>xiii</b>
<b>1. General theoretical and empirical background</b> .....	<b>1</b>
1.1 Effects of aging on the brain.....	1
1.2 Effects of aerobic exercise and physical fitness on the aging brain.....	3
1.3 Measuring changes in brain structure.....	5
1.4 Goals of the current dissertation.....	7
<b>2. Research questions and hypotheses</b> .....	<b>9</b>
<b>3. General methodology</b> .....	<b>11</b>
3.1 Young adult sample.....	11
3.2 Older adult sample.....	16
<b>4. Overview of individual works</b> .....	<b>27</b>
4.1 Paper I.....	27
4.2 Paper II.....	28
4.3 Paper III.....	30
<b>5. General discussion</b> .....	<b>33</b>
5.1 Discussion of research questions and hypotheses.....	33
5.2 Limitations.....	41
5.3 Future directions.....	43
5.4 Conclusion.....	45
<b>References</b> .....	<b>47</b>
<b>Appendix</b> .....	<b>63</b>
A. Individual works.....	63
B. Declaration of researcher contributions.....	147





## **GENERAL THEORETICAL AND EMPIRICAL BACKGROUND**

The proportion of older individuals in the global population is increasing. By 2050, it is projected that 16% of the population, that is one in six people, will be 65 years of age or older (United Nations, Department of Economic and Social Affairs, Population Division, 2019). With this shift in population age comes growing scientific concern related to the changes that accompany advancing age. Brain health and cognitive function are of particular interest, as the risk of mild cognitive impairment and dementia increases with age, and because even healthy aging is associated with declines in cognitive performance that can compromise individuals' quality of life.

### **1.1 Effects of aging on the brain**

The link between senescence and decline in the structural integrity of both gray matter (GM) and white matter (WM) in the human brain has been well established. Both cross-sectional and longitudinal studies using various metrics of brain tissue integrity have found associations between advancing age and declines in GM and WM integrity, generally following an anterior-to-posterior gradient of decline with the greatest deterioration occurring in frontal regions.

Regarding GM changes, cross-sectional studies have shown reduced GM volume (GMV) in older adults as compared to middle-aged adults (e.g., Ramanoël et al., 2018), as well as negative associations between age and GMV (Good et al., 2001; Raz et al., 2004; Van Petten et al., 2004) and cortical thickness (Lemaitre et al., 2012). One study using magnetization-transfer (MT) imaging, a quantitative magnetic resonance imaging (MRI) technique used to detect changes in microstructural tissue changes, found significantly lower values of MT ratio, an indicator of tissue integrity, in the GM of older versus younger participants, with decreases in MT ratio starting around age 40 (Ge et al., 2002). Another study using MT ratio to index GM integrity found negative correlations between age and MT ratio (Benedetti et al., 2006). Yet another study using a multivariate approach to measure GM integrity as a latent factor comprising GMV, MT ratio, and mean diffusivity (MD), a diffusion tensor-derived measure of tissue integrity, also found cross-sectional associations with age, with lower values of integrity in the medial temporal and prefrontal areas associated with

greater age (Köhncke et al., 2021). Longitudinal studies have further corroborated the cross-sectional results, with findings that GMV declines with age within individuals, and have also shown accelerated decline in older to old age (Raz et al., 2005; Taki et al., 2011; Tisserand et al., 2004). Inter-individual changes in other metrics of GM integrity, such as MT, have not been widely investigated in healthy older adults, although significant declines in MT ratio have been reported in patients with Alzheimer's disease (Ropele et al., 2012).

Similar reductions of both macrostructural WM integrity, indexed by WM volume (WMV), and microstructural WM integrity, measured with diffusion-weighted imaging (DWI), have been observed as well. In cross-sectional studies, negative associations between age and WMV in numerous brain regions have been reported, with the strongest associations again found in frontal and prefrontal regions (Jernigan et al., 2001; Raz et al., 1997, 2004). Cross-sectional associations between age and diffusion tensor-derived metrics, namely fractional anisotropy (FA) and MD, have also been widely reported, with lower FA values and higher MD values seen in older individuals (see review by Bennett & Madden, 2014). More recently, associations between age and fixel-based metrics of WM integrity, fiber density (FD), fiber cross-section (FC), and a combined measure, fiber density and cross-section (FDC), have been investigated cross-sectionally (Choy et al., 2020; Kelley et al., 2021). These fixel-based metrics were developed in order to combat the weakness of diffusion tensor-based metrics in the face of crossing fibers within voxels (Raffelt et al., 2012, 2015, 2017). Kelley and colleagues (2021) compared a group of older adults to a group of young adults and found widespread areas of lower FD, FC, and FDC in the older adults, including in anterior WM and the corpus callosum. Choy and colleagues (2020) also found negative associations between age and fixel-based metrics in a sample that included a wide range of ages (21 to 86 years). Longitudinal studies of WM integrity have also reported negative associations with age. Raz and colleagues (2005) found within-person decreases in WMV in prefrontal WM over five years, and also found that shrinkage accelerated with age. Another longitudinal study that investigated change in diffusion tensor-derived metrics found associations between age and FA in the negative direction and MD in the positive direction, as well as accelerated change in both with increasing age (Sexton et al., 2014).

Changes in brain structure are important to investigate as they may have implications for functional outcomes such as cognitive performance. The brain maintenance hypothesis posits that individual differences in the degree to which brain structure, as well as other brain-based factors, is maintained into old age predicts individual differences in cognitive performance later in life (Cabeza et al., 2018; Johansson et al., 2022; Lindenberger, 2014;

Nyberg et al., 2012; Nyberg & Lindenberger, 2020; Nyberg & Pudas, 2019). That is, individuals who exhibit less deterioration of the structural integrity of brain tissue with age may also perform better on cognitive tasks in older age. Indeed, GMV in frontal areas has been shown to predict performance on tasks of attention and executive function cross-sectionally (Zimmerman et al., 2006), and a longitudinal association has been found between atrophy in the hippocampus and decline in episodic memory performance (Gorbach et al., 2017). Köhncke and colleagues (2021) found that GM integrity in those areas showing negative cross-sectional associations with age also showed positive associations with a latent factor of episodic memory, which comprised scores on four different tasks measuring verbal, spatial, and associative memory. Relative WMV in frontal areas has also been cross-sectionally associated with performance on tasks of memory and executive function, and was also shown to partially mediate the relationship between age and declarative memory (Brickman et al., 2006). The diffusion tensor-derived measures of FA and MD have also shown cross-sectional associations with performance on cognitive tasks; O'Sullivan and colleagues (2001) reported correlations between FA in middle WM (between the genu and splenium of the corpus callosum) and verbal memory performance, as well as between MD in anterior WM (anterior to the genu of the corpus callosum) and a task of attention and executive function. Longitudinally, FA in the genu of the corpus callosum at baseline has been shown to predict annual decline over four years in a marker of global cognition (Raghavan et al., 2020).

## **1.2 Effects of aerobic exercise and physical fitness on the aging brain**

In light of these extensive age-related declines in brain tissue integrity and cognition, as well as the hypothesized link between brain structure deterioration and decline in cognitive function, efforts have been made to find ways to slow or even prevent this deterioration. Regular physical exercise training (i.e., structured physical activity aimed at increasing cardiovascular fitness), particularly aerobic exercise, has gained popularity as a modifiable lifestyle factor with the potential to evoke maintenance of brain tissue integrity in older adults.

A number of cross-sectional studies have reported correlations between GMV and levels of both physical activity and cardiovascular fitness (see Erickson et al., 2014 for a review), although there have been reports of no significant association with physical activity as well (Ho et al., 2011; J. C. Smith et al., 2011). Similarly, a number of studies have found positive associations between physical activity/fitness and WMV, as well as FA and MD (see

Sexton et al., 2016 for a meta-analytic review). However, other studies investigating the link between cardiovascular fitness and diffusion tensor-derived metrics have reported no significant associations (Burzynska et al., 2014; Marks et al., 2011; Tian, Erickson, et al., 2014; Tian, Simonsick, et al., 2014). Longitudinally, physical activity has been shown to predict GMV, with greater amounts of physical activity at baseline, quantified as distance walked in a week, predicting greater GMV nine years later (Erickson et al., 2010). Other longitudinal studies implementing intervention designs with exercise and control groups have found exercise-induced maintenance of GMV in frontal areas, including the anterior cingulate cortex, inferior frontal gyrus, and supplementary motor area, as well as WMV in the genu of the corpus callosum (Colcombe et al., 2006), and volumetric increases in the hippocampus (Erickson et al., 2011). In terms of exercise-induced changes in diffusion tensor-derived metrics, Voss and colleagues (2013) reported no significant group differences in change in FA after one year, but found significant positive correlations between percent change in frontal, parietal, and temporal FA and percent change in cardiovascular fitness within an aerobic exercise group, but not within a stretching and toning control group; they also found that the aerobic exercise group showed a greater percent change in fitness than the control group. Clark and colleagues (2019) reported decreases in FA and increases in MD in a group of older adults who exercised for six months, although they did not compare this group with a control group, and they also found no significant association between percent change in FA or MD and percent change in cardiovascular fitness.

Findings regarding the association between cognitive outcomes and aerobic exercise and cardiovascular fitness have been mixed. Older meta-analyses found evidence for a positive link between cardiovascular fitness and cognitive function across a number of domains, including attention, processing speed, memory, and executive function (Colcombe & Kramer, 2003; P. J. Smith et al., 2010). A systematic review by van Uffelen and colleagues (2008) also found evidence for improvements in memory and processing speed in healthy older adults who participated in aerobic exercise interventions. More recently, a meta-analysis of randomized controlled trials reported an overall positive association between physical activity and protection against cognitive decline in aging (Blondell et al., 2014), however, two other meta-analyses also including randomized controlled trials found no benefit of exercise on cognitive function in older adults (Kelly et al., 2014; Young et al., 2015).

In sum, there is evidence substantiating the claim that aerobic exercise is beneficial for brain tissue integrity in old age, with exercise ameliorating the declines in GM and WM

integrity that occur in undiseased aging, but these findings are juxtaposed with a number of studies reporting no significant exercise-related effects on brain structure. There is also weak evidence that aerobic exercise may induce similar maintenance in cognitive function, but a number of null findings have also been reported, especially in recent years.

### **1.3 Measuring changes in brain structure**

As outlined in Section 1.1, changes in the structural integrity of brain tissue have been quantified in a number of different ways in the context of aging. Not all of these metrics have been investigated in regards to aerobic exercise-induced changes in brain structure, however (see Section 1.2). Therefore, this dissertation aimed to quantify changes in brain tissue integrity resulting from engagement in aerobic exercise using more recently developed MRI-based methods, as well as to replicate findings from earlier studies using more established methods.

#### ***1.3.1 Voxel-based morphometry***

Early studies exploring exercise effects on brain tissue used measures of GMV and WMV. One of the first studies to use an exercise vs. control intervention design, Colcombe et al. (2006), used voxel-based morphometry (VBM; Ashburner & Friston, 2000) to investigate changes in the voxel-wise concentrations of brain tissue types, GM or WM, that differed between groups. GMV and WMV changes assessed with VBM are thought to capture macrostructural changes (i.e., general atrophy) without any explicit reference to the anatomy of the brain. Comparisons with visual and manual evaluations of certain structures suggests that the results of VBM analyses are typically biologically valid and consistent with manually assessed atrophy (Davies et al., 2009; Giuliani et al., 2005; Good et al., 2002; Whitwell et al., 2005). Analyses using VBM in the current dissertation aimed to replicate findings from previous studies using the same measures.

#### ***1.3.2 Quantitative multiparameter mapping***

Recently, a comprehensive quantitative MRI approach using multiparameter mapping (MPM) was developed to measure various microstructural properties of brain tissue, namely longitudinal relaxation rate ( $R1 = 1/T_1$ ), effective transverse relaxation rate ( $R2^* = 1/T_2^*$ ), proton density (PD), and MT saturation (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008; Weiskopf et al., 2011), the last of which is an improvement on the measure of the above-mentioned MT ratio, which is affected by  $T_1$  relaxation and flip angle inhomogeneities (Helms, Dathe, Kallenberg, et al., 2008). In the following, I will refer to MT

saturation as MT, as MT ratio will not be discussed further. Quantitative MRI aims to provide absolute measurements of physical parameters related to water proton spins. As water molecules are influenced by the microstructure of the surrounding tissue, inferences can be made about the properties of that surrounding tissue; R1 and MT can be used as indices for tissue myelination, and R2\* as a marker of iron concentration, while PD is thought to reflect the content of free water in the tissue (Draganski et al., 2011; Weiskopf et al., 2021). Analysis using MPMs in the current dissertation aimed to validate their use in longitudinal studies, as well as test whether they are suitable for measuring correlational relationships with other variables of interest, such as cardiovascular fitness, and MT was also used in a latent factor that was built to investigate exercise effects on general GM structural integrity (see Section 1.3.4).

### ***1.3.3 Diffusion-tensor modeling***

Other measures of microstructural integrity can be derived from DW images using diffusion-tensor modeling. FA and MD are commonly used to index the diffusion of water molecules in tissue, with higher values of FA corresponding to greater anisotropic diffusion, indicating greater organization of WM tracts directing diffusion, and lower values of MD corresponding to less movement of molecules, indicating the presence of more barriers restricting diffusion (Basser & Pierpaoli, 1996; Beaulieu, 2002; Pierpaoli & Basser, 1996). Analysis using diffusion tensor-derived metrics in the current dissertation aimed to test the effects of a moderate, at-home exercise intervention, as reports on these effects have so far been inconclusive. MD was also used in the latent factor of GM structural integrity (see Section 1.3.4).

### ***1.3.4 Latent gray-matter structural integrity***

To take a step back from individual measures of tissue integrity that reflect specific properties of macro- or microstructure in the brain and to instead capture GM integrity on a more generalized level, Köhncke and colleagues (2021) developed a multi-trait multi-method (MTMM) model to measure latent GM structural integrity using structural equation modeling (SEM), building upon work by Kühn and colleagues (2017). In general, latent constructs may afford more valid and reliable conclusions than those drawn from analyses using observed variables, as latent variables capture only the shared variance across multiple indicators, parsing out specific residual variance (Wansbeek & Meijer, 2003). Thus, to measure GM structural integrity using various measures that tap into different aspects of macro- and microstructural integrity in GM in a factor without random measurement error, Köhncke and

colleagues (2021) combined GMV, MT, and MD in a cross-sectional model. The analyses in the current dissertation aimed to extend this model of latent GM structural integrity for use in a longitudinal context, and to explore exercise-induced changes in this factor in regions that have previously shown volumetric effects of exercise.

### **1.3.5 Fixel based analysis**

Although diffusion tensor-derived metrics have been widely used to investigate WM changes in older age, as well as exercise effects in aging, their use has recently been challenged, as the diffusion tensor cannot model more than a single fiber population in a given voxel, even though an estimated 60–90% of voxels contain crossing fibers (Jeurissen et al., 2013). Therefore, a new method using “fixels” was developed, which refer to specific fiber bundles within specific voxels, and which can account for voxels containing multiple fiber bundles in various directions (Raffelt et al., 2017). To do this, the distribution of fiber orientations within each voxel, termed the fiber orientation distribution (FOD), is estimated using constrained spherical decomposition (CSD; Dhollander et al., 2016; Jeurissen et al., 2014). These FODs are then used to calculate FD and FC, the latter of which is then transformed into  $\log(\text{FC})$ , as well as the product of FD and FC, FDC. Changes in FD may represent a microstructural difference in within-voxel density of a fiber bundle, and changes in FC may represent a macrostructural difference in the voxels occupied by the cross-section of a fiber bundle, while changes in FDC may represent a combination of the two, all of which would theoretically affect the ability of the fiber bundle to “relay information” (Raffelt et al., 2017). The biological validity of these measures is at least partially supported by findings from a combined histological and *ex vivo* MRI study conducted in rats, which reported that FD was indeed tightly correlated with axonal density in the optic nerve and optic chiasm (Rojas-Vite et al., 2019). The effect of exercise on FD, FC, and FDC has not yet been examined, thus the analyses including these fixel-based metrics in the current dissertation aimed to investigate whether these metrics would show exercise-induced changes in aging.

### **1.4 Goals of the current dissertation**

In sum, various MRI-based measures of brain tissue integrity have been used to demonstrate structural changes that the brain undergoes as humans age. Some of these measures have also been used to investigate the effects that aerobic exercise may have on the aging brain, however open questions remain regarding latent GM structural integrity, the diffusion tensor-derived measures of FA and MD, and the novel fixel-based metrics of FD, FC, and FDC. Therefore, this dissertation aimed to answer some of these open questions

using two studies: In Paper I, we used data from a sample of young adults to test the reliability of the quantitative metrics assessed with MPM, and in Papers II and III, we used data from a sample of older adults to test the effects of a moderate, at-home exercise intervention on brain tissue integrity using the metrics described in Section 1.3.



*Chapter 2.*

## **RESEARCH QUESTIONS AND HYPOTHESES**

The overall aim of this dissertation was to use novel metrics of GM and WM integrity in the brain to investigate the effects of six months of moderate, at-home aerobic exercise on the structural integrity of the aging brain.

First, in order to investigate how well recently developed quantitative MR measures can capture individual differences, a study with a small sample of young adults ( $N = 15$ , mean age = 27.3 years) was conducted with a two-day, four-session design. With this design, we aimed to answer the following questions:

1. What are the proportions of variance that can be attributed to repositioning a participant in the scanner either on the same day or across days, to random error, and to between-person differences?
2. Is this reliability consistent or does it vary across different regions in the brain?

Then, we adapted an existing MTMM model of latent GM structural integrity to look at exercise-induced changes in GM structural integrity. These models were fit with data from a sample of older adults ( $N = 75$ , mean age = 70.3 years) to address the following questions:

3. Can moderate, at-home exercise induce an increase in cardiovascular fitness in healthy, previously sedentary older adults?
4. Can moderate, at-home exercise induce positive changes and/or maintenance in GM structural integrity in areas of the brain that have previously shown volumetric effects of exercise?
5. Is change in cardiovascular fitness positively correlated with changes in GM structural integrity?

Regarding Question 3, we expected to see gains in cardiovascular fitness within a group of exercising participants after six months of at-home aerobic exercise, consistent with studies looking at whether home-based exercise interventions can improve fitness in older adults (King, 1991; Salvetti et al., 2008). Regarding Question 4, we expected to see positive

changes in GM structural integrity, consistent with gains in hippocampal volume seen in Erickson et al. (2011), or maintenance of GM structural integrity, consistent with maintenance in frontal regions seen in Colcombe et al. (2006). Regarding Question 5, we expected to see a positive relationship between change in cardiovascular fitness and changes in regional GM structural integrity, consistent with the studies discussed in Erickson et al. (2014).

Finally, we used both established as well as newly developed metrics of WM integrity to investigate the effects of aerobic exercise on WM structure in the brain. Using data from the same sample of older adults, we aimed to address the following questions:

6. Can moderate, at-home exercise induce changes in metrics of WM integrity, namely WMV, diffusion tensor-derived FA and MD, and fixel-based FD, log(FC), and FDC?
7. Are changes in WM integrity correlated with change in cardiovascular fitness or change in cognitive performance?

Regarding Question 6, we expected to see maintenance of WMV in anterior regions of the brain, consistent with the findings reported in Colcombe et al. (2006). We also expected to see increases in FA and decreases in MD in exercisers as compared to controls, indicating an amelioration of age-related declines, no effects of exercise on either FA or MD, consistent with findings reported by Voss and colleagues (2013), or decreases in FA and increases in MD in exercisers, consistent with findings reported by Clark and colleagues (2019). Based on reports of aging effects on FD, log(FC), and FDC (e.g., Choy et al., 2020; Kelley et al., 2021), we expected to see decreases in all three metrics in control participants, with potential amelioration of these decreases in exercise participants, manifesting either as positive changes or no significant change (i.e., maintenance). Regarding Question 7, we expected to see correlations between change in cardiovascular fitness and changes in WMV (positive), FA (positive), and MD (negative), consistent with the studies discussed in Sexton et al. (2016), as well as positive correlations with changes in FD, log(FC), and FDC. We also expected to see correlations between change in cognitive performance and changes in WMV (positive), consistent with Brickman et al. (2006), FA (positive) and MD (negative), consistent with O'Sullivan et al. (2001) and Raghavan et al. (2020), and FD, log(FC), and FDC (positive).

## **GENERAL METHODOLOGY**

This dissertation comprises a paper detailing results from a short two-day, four-session MRI study run on a sample of healthy young adults (22–31 years) and two papers reporting findings of a multivariate intervention study that included a sample of healthy older adults (63–78 years).

### **3.1 Young adult sample**

#### ***3.1.1 Participants***

To investigate the reliability of individual differences in four quantities indexed by MRI-based MPMs, a sample of 15 healthy young adults (53% females, mean age =  $27.3 \pm 3.34$ ) were scanned. All participants had normal or corrected-to-normal vision, normal hearing, no history of a psychiatric or neurological disorder, and met MRI compatibility requirements. Participants also showed comparable cognitive performance to other samples of young adults (e.g., Hoyer et al., 2004; Lövdén et al., 2012), with a mean score of 65 points (standard deviation [SD] = 6.9) on the Digit Symbol Substitution task (DSST) and 31 points (SD = 2.5) on a multiple-choice vocabulary test (see Section 3.1.5 for details).

#### ***3.1.2 Study design***

Participants were scanned four times over two consecutive days (days 1 and 2). On the first day, participants underwent two measurements in succession with no repositioning between scans (sessions 1 and 2). On the second day, participants were removed from the scanner after the first measurement (session 3) and briefly got up and walked around while the head coil was removed for a short period before undergoing the second measurement (session 4).

#### ***3.1.3 MRI acquisition***

MR images were acquired on a Siemens Tim Trio 3 Tesla MRI Scanner (Erlangen, VB17a software version) with a standard radio-frequency (RF) 32-channel receive head coil and an RF transmit body coil. The MPM acquisition protocol and parameters were designed and chosen following previous work (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008; Lutti et al., 2010; Weiskopf et al., 2011, 2013). One static magnetic

( $B_0$ ) gradient echo (GRE)-field map, one RF transmit ( $B_{1+}$ ) field map, and three multi-echo 3D fast low angle shot (FLASH) sequences were acquired, predominantly weighting  $T_1$ , PD, and MT, respectively. See Table 1 for details regarding acquisition parameters.

### 3.1.4 MRI (pre)processing

Pre-processing and processing were performed in SPM12<sup>1</sup> (Institute of Neurology) in MATLAB (version: 2017b; The MathWorks Inc., Natick, MA, USA) using the hMRI toolbox<sup>2</sup> (Tabelow et al., 2019). Quantitative and semi-quantitative estimates of  $R_1$ ,  $R_2^*$ , PD, and MT were computed using the *Create hMRI maps* module.

Table 1

*Acquisition parameters for MR sequences applied in the young adult sample*

Sequence	Parameters	Acquisition time
$B_0$ GRE-field map	64 transverse slices, slice thickness = 2 mm with a 50% distance factor, TR = 1020 ms, $TE_1/TE_2 = 10/12.46$ ms, $\alpha = 90^\circ$ , matrix = $64 \times 64$ , FOV = $192 \times 192$ mm, right-left PE direction, BW = 260 Hz/Px, flow compensation	2:14 min
$B_{1+}$ field map	4 mm isotropic resolution, TR = 500 ms, $TE_{SE/STE}/\text{mixing time} = 39.06/33.80$ ms, $\alpha = 115^\circ$ to $65^\circ$ in steps of $-5^\circ$ (see Akoka et al., 1993), matrix = $64 \times 48 \times 48$ , FOV = $256 \times 192 \times 192$ mm, parallel imaging using GRAPPA factor $2 \times 2$ in PE and partition directions	3:00 min
Multi-echo 3D FLASH, general parameters	1 mm isotropic resolution, 176 slices per slab, FOV = $256 \times 240$ mm, TR = 24.5 ms, BW = 465 Hz/pixel, gradient spoilers and RF spoiling with phase increment = $50^\circ$ after each TR, parallel imaging with GRAPPA factor 2 in PE direction and 6/8 partial Fourier acquisitions in partition direction (inner/fast PE loop)	
$T_1$ -weighted specifics	$TE_{1-6}$ with alternating readout polarity = 2.34/4.68/7.02/9.36/11.70/14.04 ms, $\alpha = 21^\circ$	7:03 min
PD-weighted specifics	$TE_{1-8}$ with alternating readout polarity = 2.34/4.68/7.02/9.36/11.70/14.04/16.38/18.72 ms, $\alpha = 6^\circ$	7:03 min
MT-weighted specifics	$TE_{1-6}$ with alternating readout polarity = 2.34/4.68/7.02/9.36/11.70/14.04 ms, $\alpha = 6^\circ$ , off-resonance Gaussian-shaped RF pulse before excitation (4 ms, $220^\circ$ nominal $\alpha$ , 2 kHz frequency offset from water resonance)	7:03 min

*Note.* TR = repetition time; TE = echo time;  $\alpha$  = flip angle; FOV = field of view; PE = phase encoding; BW = bandwidth; GRAPPA = Generalized autocalibrating partial parallel acquisition; SE = spin echo; STE = stimulated echo.

<sup>1</sup> [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)

<sup>2</sup> <https://hmri-group.github.io/hMRI-toolbox/>

First,  $R2^*$  ( $1/T_2^*$ ) maps were estimated using ESTATICS (Weiskopf et al., 2014), applying a joint log-linear fit using ordinary least squares to the  $T_1$ -weighted, PD-weighted, and MT-weighted contrasts. Then, uncorrected MT, PD, and R1 maps were calculated from the  $T_1$ -weighted, PD-weighted, and MT-weighted measurements (extrapolated to echo time = 0 to minimize dependency on  $R2^*$ ) by applying the Ernst equation (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008).  $T_1$  maps were corrected for imperfect RF spoiling following the approach by Preibisch and Deischmann (2009). R1 ( $1/T_1$ ) maps were corrected for local RF transmit field inhomogeneities using the  $B_{1+}$  field maps, which were corrected for echo planar imaging (EPI)-specific distortions using the  $B_0$  GRE-field maps. MT maps were also corrected for residual local RF transmit field inhomogeneities following Rowley et al. (2021) and Weiskopf et al. (2013). PD maps were estimated from the signal amplitude maps, adjusting for global and local receive sensitivity inhomogeneities using the “unified segmentation” approach (Ashburner & Friston, 2005), and calibrating mean WM PD to 69 percent units. For more details regarding the estimation of MPMs, see Wenger, Polk et al. (2022).

A longitudinal processing pipeline was used to improve within-subject coregistration of R1,  $R2^*$ , PD, and MT maps. First, within-subject templates were created from the MT and PD maps. Thesholding was applied to MT (between 0 and 5) and PD maps (between 0 and 200) to improve segmentation performance. Multichannel segmentations were conducted on each pair of thresholded MT and PD maps for each session and each subject separately to create GM- and WM-segmented images. Unbiased within-subject templates were created from these segmented images using SHOOT, a diffeomorphic registration tool with geodesic shooting and Gauss-Newton optimization in SPM12 (Ashburner & Friston, 2011). To create the between-subject template, the deformation fields that resulted from the previous step were applied to the raw MT and PD maps, and within-subject median MT and PD maps were computed. These median maps were also segmented using multichannel segmentation, resulting in median GM- and WM-segmented images, which were then used to create a between-subject template with SHOOT, again producing deformation fields. The within- and between-subject deformation fields were combined and all R1,  $R2^*$ , PD, and MT maps were normalized to MNI space with these combined deformation fields using SHOOT. Median GM- and WM-segmented images were also spatially normalized to MNI space, applying Jacobian modulation. These normalized tissue-segmented images and the four parameter maps were used to compute “smoothed tissue specific MPMs,” following the voxel-based

quantification approach by Draganski and colleagues (2011), applying a 6-mm full-width half-maximum (FWHM) smoothing kernel and weighted averaging.

Within-session, within-subject, across-voxel means were extracted using the tools from the FSL (version 6.0; Jenkinson et al., 2012; S. M. Smith et al., 2004; Woolrich et al., 2009) from a number of predefined regions of interest (ROIs) from the Harvard-Oxford cortical and subcortical structural atlases<sup>3</sup> (Desikan et al., 2006) with a threshold of .25. The ROIs discussed in Wenger, Polk et al. (2022) may be of interest to researchers interested in longitudinal changes in brain structure, particularly those that might exhibit intervention-induced changes in response to cognitive or physical training: whole-brain GM, whole-brain WM, inferior frontal gyrus – pars triangularis, inferior frontal gyrus – pars opercularis, orbitofrontal cortex, anterior cingulate cortex, precuneus, middle temporal gyrus, caudate nucleus, putamen, and pallidum. Additionally, in a supplementary analysis, reliability was also estimated within all other Harvard-Oxford ROIs, and whole-brain voxel-wise analyses were also performed, estimating reliability for each voxel in GM and WM segmented images separately (available at <https://osf.io/6p9bf/>).

### **3.1.5 Cognitive testing**

To ensure that the sample of young adults included in Wenger, Polk et al. (2022) was comparable to other samples of young adults, the DSST (Wechsler, 1981) and a multiple choice German vocabulary test (Mehrfachwahl-Wortschatz-Test A [MWT-A]; Lehrl et al., 1991) were administered with paper and pencil.

The DSST is a digit-symbol matching task in which participants are provided a key code of the digits 1 through 9, each paired with a unique symbol. Beneath the key code, participants are provided several rows of double boxes, each containing a digit in the top box and a blank space in the bottom box. Participants are asked to fill in the empty boxes with the symbol that corresponds to the digit in the respective upper box, completing as many pairs as possible within 90 seconds. The DSST is a widely used test that can index general cognitive acuity and taps into attention and perceptual speed, as well as associative learning and working memory (Jaeger, 2018).

The MWT-A comprises 37 lines of five words each, one of which is real and four of which are nonsense words. Participants are asked to select the real word (in the German language) on each line. Score on the MWT-A is used to index crystallized intelligence.

---

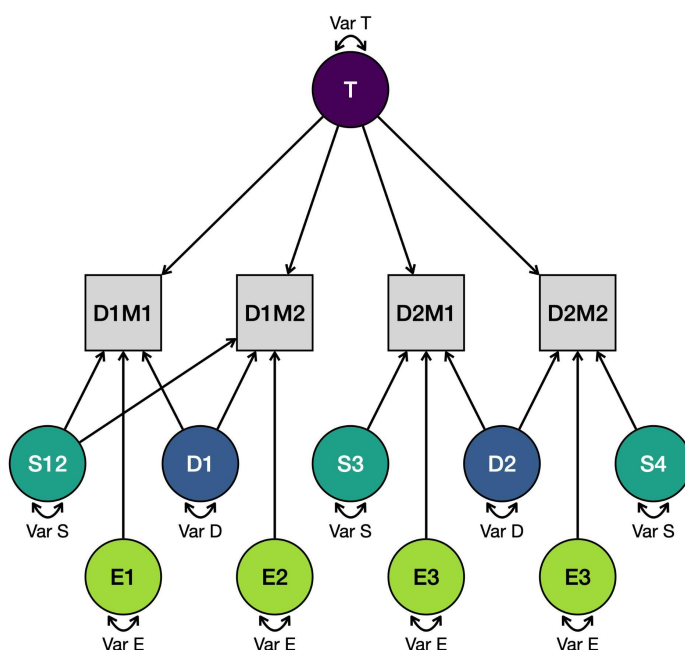
<sup>3</sup> <https://neurovault.org/collections/262/>

### 3.1.6 Statistical analyses

**Paper I.** Model specification and estimation were conducted in  $\Omega$ nyx (von Oertzen et al., 2015), and R (R Core Team, 2021) in RStudio (RStudio Team, 2021).

Reliability of each parameter was estimated with the intra-class effect decomposition (ICED) model (Brandmaier et al., 2018) using SEM with the lavaan R package (Rosseel, 2012). This model allows for the decomposition of measurement error into orthogonal sources. With the implemented two-day, four-session design, effects of day, session, and residual variance on measurement error could be estimated, and therefore the total variance could be parsed into true-score variance (Var T; representing true inter-individual differences), day-specific variance (Var D), session-specific variance (Var S), and residual-error variance (Var E). Full information maximum likelihood (FIML) was used to account for missing data without the need for case-wise deletion.

The full model including all four variance parameters (Var T, Var D, Var S, Var E), bounded to  $\geq 0.0001$ , was estimated (see Figure 1 for an exemplary path diagram), as well as three null models in which one variance parameter, Var T, Var D, or Var S, was set to zero, in order to conduct likelihood-ratio tests (LRTs). An LRT is a test in which a model in which the parameter(s) of interest is freely estimated is compared to a nested model in which this parameter is fixed (to zero in this case). The likelihood ratio, that is, the difference in  $\chi^2$  between the two models, indicates the difference in fit, and, if significant, that the null hypothesis (the models fit equally well) can be rejected (Kline, 2016). Wald tests were used to test Var E against zero, as a model with no orthogonal error structure cannot be estimated.



**Figure 1 | Exemplary path diagram of the ICED model with generic indicators.**

*Note.* Unlabeled paths are fixed to 1, labeled paths are freely estimated, with identically named paths constrained to be equal.

T = true score; D = day; S = session; E = residual error; Var = variance; D1M1 = D1, measurement 1 (S1); D1M2 = D1, measurement 2 (S2); D2M1 = D2, measurement 1 (S3); D2M2 = D2, measurement 2 (S4).

Adapted with permission from Wenger, Polk et al. (2022).

© 2022 Wenger and colleagues.

Intra-class correlation (ICC) was then calculated for each parameter in each ROI from the estimated variance components, and was defined as the ratio of inter-individual variance to total variance:

$$ICC = \frac{Var T}{Var T + Var D + Var S + Var E} .$$

In this way, ICC describes the test-retest reliability of a single measurement, and indicates how well a measurement at any single time point can capture the underlying quantitative value, with higher values indicating better suitability of the parameter to measure individual differences in a given ROI.  $ICC_2$  was also calculated for each parameter in each ROI, defined as the ratio of true score variance to total effective variance ( $Var_{eff}$ ), where effective variance is a single residual error term calculated from the variance components other than the one of interest:

$$ICC_2 = \frac{Var T}{Var T + Var_{eff}} ,$$

where in the current study design

$$Var_{eff} = \frac{(2 Var D + Var S + Var E)(2 Var D + 2 Var S + Var E)}{8 Var D + 3 Var S + 4 Var E} ,$$

as calculated with the algorithm from von Oertzen (2010). In this way,  $ICC_2$  describes test-retest reliability within the implemented two-day, four-session design, and indicates how well the underlying quantitative value is measured within this specific design. For ICC and  $ICC_2$  values in each model, 95% confidence intervals were estimated by bootstrapping over 1000 iterations with the *boot.ci* command in the boot R package (Canty & Ripley, 2021; Davison & Hinkley, 1997). Finally, coefficient of variation (CoV) was calculated for each parameter in each ROI, defined as the ratio of SD across the four sessions, normalized over the number of measurements minus one to avoid bias, to overall mean across the four sessions:

$$CoV_4 = \frac{SD_{4-1}}{Mean_4} .$$

In this way, CoV describes the precision of measurement, with larger values indicating greater imprecision.

The data collected from this sample of young adults are openly available in OSF at <https://osf.io/6p9bf/>, as are the scripts written for the processing of MRI data and all statistical analyses.

### 3.2 Older adult sample

The older adult sample included in Papers II (Polk, Kleemeyer, Köhncke et al., 2022) and III (Polk, Kleemeyer, Bodammer et al., 2022) of this dissertation was drawn from the



“Aktives Altern für Körper und Geist” [active aging for body and mind] (AKTIV) study, the full protocol of which is detailed in Wenger, Düzel et al. (2022). In brief, the larger study aimed to investigate the effects of cognitive (Spanish learning) and physical training (aerobic exercise), as well as any additive or multiplicative effects of a combined training, on the brain and cognition in older adults. This dissertation focused on structural brain changes induced by aerobic exercise, and thus compared a group of sedentary individuals in an active control condition to a group of individuals who regularly engaged in aerobic exercise.

### **3.2.1 Participants**

Healthy older adults were recruited via newspaper advertisements and previous participation in studies at the Max Planck Institute for Human Development, Berlin. Volunteers were included in the study if they met the following criteria: no MRI contraindications; able to meet the time requirements of the study (ca. 45 min a day for six months); right-handed; between 63 and 78 years old at the start of the study; no history of head injury, or medical (e.g., heart attack, cancer), neurological (e.g., epilepsy), or psychological disorder (e.g., depression); not on medication that could affect memory function. Additionally, to control for baseline levels of language ability and physical activity, individuals were only included if they did not speak a romance language and were not proficient in more than two languages including German (e.g., participants who spoke German and English were included), and if they did not exercise more than once every two weeks. Two hundred and one individuals qualified for participation in the study.

Participants were pseudo-randomly assigned to one of the four intervention groups: (1) an active control group (ACG; included in this dissertation), (2) a language group (not included in this dissertation), (3) an exercise group (EG; included in this dissertation), or (4) a combined language and exercise group (not included in this dissertation). Participants were blinded to the group design of the study.

### **3.2.2 Study design**

At time point one (T1), before starting any training, participants underwent a physical examination with cardiopulmonary exercise testing (CPET) and an MRI session, and also completed a comprehensive cognitive battery. Twenty-three participants were excluded at this stage due to pre-existing medical conditions determined at the physical examination ( $n = 22$ ) or claustrophobia in the scanner ( $n = 1$ ), and 19 dropped out before starting the interventions due to disinterest. After three months of training, at time point two (T2), MRI and cognitive sessions were repeated. Finally, after a total of six months of training, at time point three

(T3), physical examinations, MRI sessions, and cognitive sessions were repeated. Seventeen participants dropped out over the course of the six months due to health complaints, disinterest, time constraints, or unspecified reasons, resulting in a total of  $N = 142$  participants who completed the intervention. See Figure 2 for a schematic diagram of participant recruitment, group assignment, and the study design.

### 3.2.3 Interventions

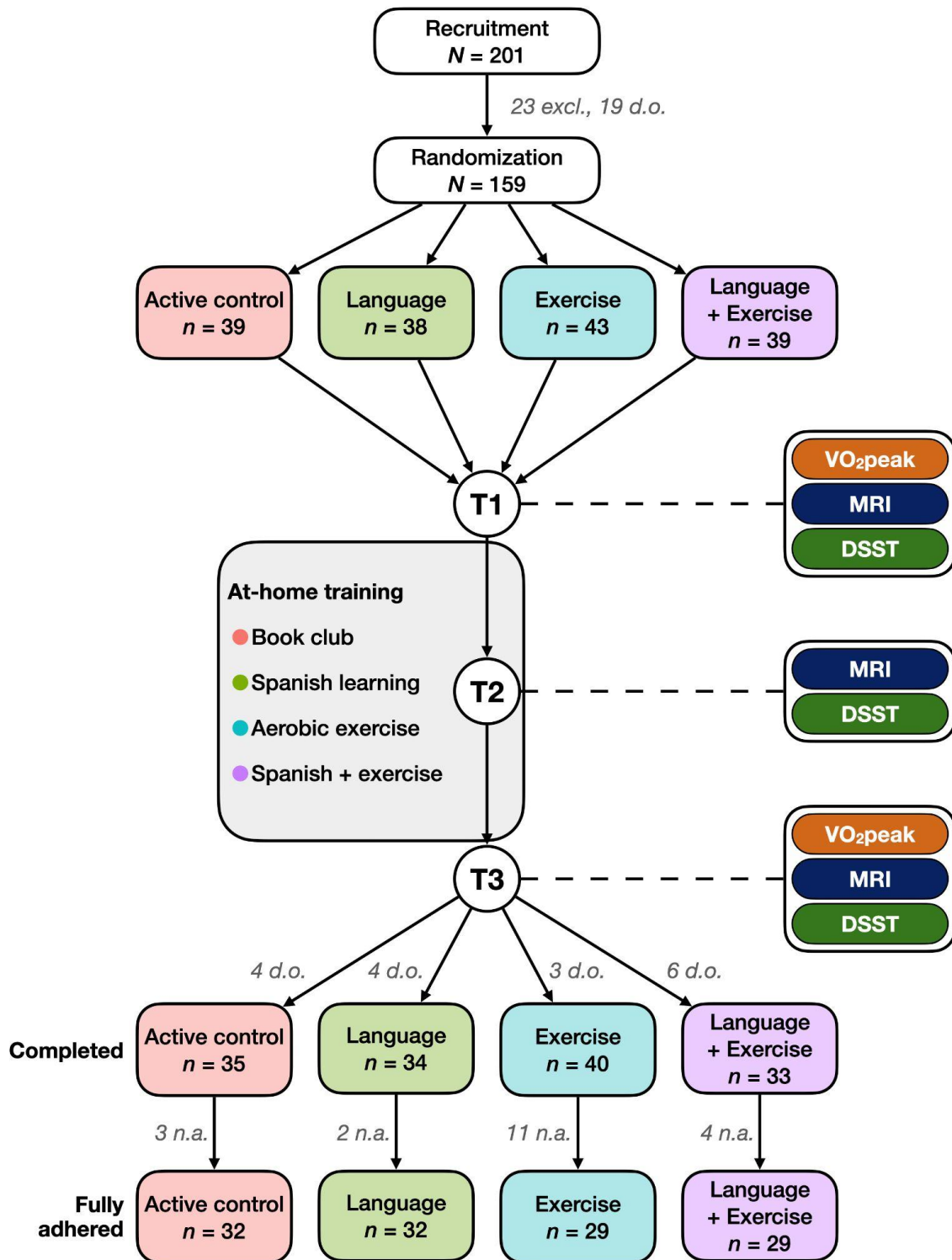
All participants were asked to complete their respective training at home at their convenience on six days per week for approximately 45 minutes per day starting immediately after measurements were taken at T1 and continuing until their T3 appointments. Additionally, participants came to intervention-specific group sessions (5–10 individuals per group) at the Max Planck Institute for Human Development each week. Only the ACG and EG are considered in this dissertation.

**Active control group.** Individuals in the ACG ( $n = 35$ , 40% females, mean age =  $70.7 \pm 3.81$  years) participated in a book club as an active control condition. They were asked to read a selection of German literature at a slow pace on a tablet (Lenovo TB2-X30L TAB) at least six days a week for 45 minutes each day. ACG participants also took part in group discussion sessions once a week, which were hosted by a facilitator from an external party<sup>4</sup>. Full adherence to the active control condition was defined as reading at least 1890 minutes over the course of the study, corresponding to circa 90 minutes of reading per week for at least 21 weeks, with no pauses of longer than two weeks. Three participants did not meet this criterion, resulting in 32 fully compliant participants in the ACG with an average total time spent reading of 4772 minutes (SD = 1819.6 min). Adherence was assumed to be independent from any baseline measurements.

**Exercise group.** Individuals in the EG ( $n = 40$ , 50% females, mean age =  $69.8 \pm 3.49$  years) exercised at home at a moderate intensity three to four days per week on a stationary bicycle (DKN Ergometer AM-50) using an interval training regime that was personalized to each individual's fitness measured at the physical examination at T1. The training regime was programmed onto a tablet, and the tablet and bicycle were synced via Bluetooth. At baseline, participants trained for 30 minutes per session at their individualized intensity; every two to three weeks, the difficulty of the training was raised by increasing the duration of the intervals by three minutes and the resistance of the bicycle by three to four Watts. Following

---

<sup>4</sup> <http://shared-reading.de/>



**Figure 2 | Schematic diagram of participant recruitment, randomization into groups, and study design in the AKTIV Study.**

*Note.* Excl. = excluded: individuals excluded from participation due to pre-existing health conditions or claustrophobia; d.o. = drop-outs: participants who dropped out to disinterest, physical complaints, time constraints, or unspecified reasons; n.a. = non-adhering: participants who did not meet the criteria for full adherence to the intervention. Participants were asked to start training at home immediately following measurements at T1 and to continue until measurements at T3. Only the participants from the active control and exercise conditions are considered in the current dissertation.

each session, participants rated their subjective exhaustion using the Borg Rating of Perceived Exertion Scale, which is a 15-point rating scale where a rating of 6 corresponds to no exertion at all and a rating of 20 to maximal exertion. If participants indicated that the training was too easy (Borg rating of  $< 12$ ) or too difficult (Borg rating of  $> 15$ ) for multiple sessions in a row, the intensity of the training could be adjusted accordingly. EG participants were also asked to read the same German literature as the ACG; on those days when participants completed an exercise session, they were asked to read for an additional 15 minutes and on those days when they did not complete an exercise session, they were asked to read for 45 minutes. In this way, the EG completed approximately 45 minutes of study-related activity (i.e., reading and/or exercise) six days a week. EG participants also took part in group sessions consisting of toning and stretching exercises led by an external instructor. Adherence to the aerobic exercise condition was defined as exercising at least 1890 minutes over the course of the study, corresponding to circa 90 minutes of exercise per week for at least 21 weeks, with no pauses of longer than two weeks, and a slight steady increase in training intensity from T1 to T3, as was automatically implemented in the exercise regime. Ten participants did not meet these criteria, and an additional participant experienced prohibitive technical difficulties and thus was also categorized as non-adherent, resulting in 29 fully compliant participants in the EG who spent an average of 3381 minutes reading (SD = 1143.3 min) and an average of 3173 minutes exercising (SD = 410.4 min) in total over the course of the intervention. Again, adherence was assumed to be independent from any baseline measurements.

#### ***3.2.4 Physical assessments***

The physical examinations with CPET took place at the Charité – University Medicine Berlin under the supervision of a sports medicine physician. Indices of cardiovascular fitness were assessed at T1 and T3 to include in the analyses as measures of interest, as well as to determine the individual baseline training intensities for participants in the EG. The main cardiovascular measure of interest in this dissertation is peak oxygen uptake ( $VO_{2peak}$ ), relativized by body weight in kilograms (measured with the Seca 285 DP measuring station, Seca GmbH, Hamburg, Deutschland).  $VO_{2peak}$  was assessed with bicycle ergospirometry (Ergoselect 100k, Ergoline GmbH, Bitz, Germany) using the Quark Clinical-based Metabolic Cart with the standard Breath-by-Breath setup, and the V2 Mask (Hans Rudolph, Inc.). The ergospirometry protocol included a three-minute rest phase, an exertion phase starting at 20 Watts of resistance and increased by an additional 20 Watts every three

minutes until participants reported subjective maximum exertion (which may not necessarily reflect absolute maximum exertion possible, especially in older individuals), and a five-minute recovery phase with no resistance. During the entire protocol, participants were instructed to hold a constant pedaling rate of 60–70 rotations per minute.

### 3.2.5 MRI acquisition

The older adult sample was scanned in the same 3 Tesla MRI scanner with the same head and body coils as the young adult sample. Of interest in this dissertation are the  $T_1$ -weighted structural images, acquired with a 3D  $T_1$ -weighted magnetization prepared gradient-echo (MPRAGE), DWI, acquired with a single-shot DW spin-echo-refocused EPI sequence with a two-shell scheme plus six additional images inverting the phase encoding direction without diffusion weighting, and MPMs, acquired with the same protocol as the one described in Section 3.1.3. Identical sequences were obtained at T1, T2, and T3. See Table 2 for details regarding acquisition parameters.

Table 2

*Acquisition parameters for MR sequences applied in the older adult sample*

Sequence	Parameters	Acquisition time
3D $T_1$ -weighted MPRAGE	1 mm isotropic resolution, TR = 2500 ms, TE = 4.77 ms, TI = 1100 ms, $\alpha = 7^\circ$ , matrix = $256 \times 256 \times 192$ , with prescan normalize option and 3D distortion correction for non-linear gradients	9:20 min
DW EPI	62 slices, 2 mm isotropic resolution, TR = 9700 ms, TE = 120 ms, FOV = $224 \times 224$ mm, $b$ -values = 710/2850 $s/mm^2$ , directions = 30/60, 10 images with no DW, anterior-to-posterior PE direction, parallel imaging with GRAPPA factor 2 in PE direction	16:41 min
PE-inverted non-DW EPI	6 images with no DW, posterior-to-anterior PE direction	1:29 min
$B_0$ GRE-field map	See Table 1	2:14 min
$B_{1+}$ field map	See Table 1	3:00 min
Multi-echo 3D FLASH	See Table 1	21:09 min

*Note.* TR = repetition time; TE = echo time; TI = inversion time;  $\alpha$  = flip angle; FOV = field of view; PE = phase encoding; BW = bandwidth; GRAPPA = Generalized autocalibrating partial parallel acquisition.

### 3.2.6 MRI (pre)processing

MR images were processed for use in an ROI analysis approach for Paper II (Polk, Kleemeyer, Köhncke et al., 2022) and a whole-brain analysis approach for Paper III (Polk, Kleemeyer, Bodammer et al., 2022). For Paper II, GMV, MD, and MT were evaluated from structural T<sub>1</sub>-weighted images, DW images, and MPMs, respectively. Within-subject, within-ROI mean values were extracted from ROIs that were selected based on findings from previous studies that implemented aerobic exercise interventions (Colcombe et al., 2006; Erickson et al., 2011). The Harvard-Oxford cortical and subcortical atlases (Desikan et al., 2006) with a threshold of .25 were used to define the left and right hippocampus, anterior cingulate cortex, posterior cingulate cortex, precentral gyrus, juxtapositional lobule cortex (terminology from the Harvard-Oxford cortical atlas; previously termed the supplementary motor cortex), and inferior frontal gyrus (combined pars triangularis and pars opercularis). For Paper III, whole-brain WMV maps were calculated from structural T<sub>1</sub>-weighted images, while FA and MD maps as well as voxel-wise FD, log(FC), and FDC maps were calculated from the DW images.

Structural T<sub>1</sub>-weighted images were preprocessed in CAT12 (Structural Brain Mapping group, Jena University Hospital; Gaser & Dahnke, 2016) in SPM12. The longitudinal pipeline was used with default parameters. To realign images within each participant, inverse-consistent rigid registration with intra-subject bias correction was applied. These realigned images then underwent tissue segmentation. Deformation fields were then estimated using non-linear Dartel normalization and averaged across time points to create mean intra-subject deformation fields. These mean deformation fields were applied to the segmented images from all participants at all time points and nonlinear-only modulation was applied. The resulting segmented GM and WM images were smoothed with an 8-mm FWHM standard Gaussian kernel. These smoothed, segmented images were later used to extract within-subject, within-ROI mean GM probability values with FSL, which were used in the longitudinal GM structural integrity models in Paper II (Polk, Kleemeyer, Köhncke et al., 2022), and to conduct VBM on whole-brain WMV in Paper III (Polk, Kleemeyer, Bodammer et al., 2022).

DW images were preprocessed using commands from MRtrix (version 3.0\_RC3; Tournier et al., 2019), FSL (version 6.0.2; Jenkinson et al., 2012; Smith et al., 2004; Woolrich et al., 2009), and ANTS (version 2.2.0; Avants et al., 2010, 2011), following recommendations from the Basic and Advanced Tractography with MRtrix for All Neurophiles (B.A.T.M.A.N.) tutorial (Tahedl, 2018). Images were first denoised with a  $5 \times 5$

× 5 patch using *dwidenoise* (MRtrix) and Gibb's ringing artifacts were removed using *mrdegibbs* (MRtrix). Corrections for EPI distortion and B<sub>0</sub>-field inhomogeneity were applied with *topup* (FSL) and corrections for eddy-current and movement distortion were applied with *eddy\_cuda* (FSL), using a quadratic spatial model for *eddy*, replacing outliers and saving contrast-to-noise ratio maps and residual maps. Binary brain maps were generated from each individual's mean b<sub>0</sub> image with *bet* (FSL) and visually inspected, manually adjusting the maps when necessary.

Diffusion tensor maps were fit to each dataset using the lower *b*-value (710 s/mm<sup>2</sup>) with the MRtrix command *dwi2tensor*. At this point, MD values that were later included in the GM structural integrity models in Paper II (Polk, Kleemeyer, Köhncke et al., 2022) were derived via *tensor2metric* (MRtrix). To derive FA and MD maps for the whole-brain analyses described in Paper III (Polk, Kleemeyer, Bodammer et al., 2022), FSL commands were used following the TBSS User Guide from FSL<sup>5</sup> (Smith et al., 2004, 2006). FA images were slightly eroded and end slices were zeroed with *tbss\_1\_preproc*. Nonlinear registration was applied to align FA images to the JHU ICBM FA 1mm atlas<sup>6</sup> from FSL with *tbss\_2\_reg*. The generated nonlinear transforms were then applied to all images with *tbss\_3\_postreg* to bring them into standard space. MD maps were generated in the same manner, applying the nonlinear transforms generated in the previous step using *tbss\_non\_FA*. FA and MD images were finally smoothed in SPM12 using a 4-mm FWHM standard Gaussian kernel.

The FD, log(FC), and FDC maps used in Paper III (Polk, Kleemeyer, Bodammer et al., 2022) were generated from the preprocessed DW images with MRtrix commands following the Fibre density and cross-section – Multi-tissue CSD tutorial from the MRtrix3 documentation<sup>7</sup> (Tournier et al., 2019). Average tissue response functions were computed using *dwi2response* and *responsemean*. Multi-shell multi-tissue CSD was performed to estimate the FODs using *dwi2fod* with the *msmt\_csd* option. Joint bias field correction was performed with *mtnormalise*. Intra-subject templates were created from the FOD images from all time points using *population\_template*, to which all images within each subject were registered with *mrregister*. An unbiased, study-specific FOD template was then created using 40 randomly selected participants (ten from each intervention group) with *population\_template*, to which each within-subject FOD template was registered with *mrregister*. A template mask of voxels containing data in all images was created using

<sup>5</sup> <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS/UserGuide>

<sup>6</sup> <https://identifiers.org/neurovault.image:1402>

<sup>7</sup> [https://mrtrix.readthedocs.io/en/latest/fixel\\_based\\_analysis/mt\\_fibre\\_density\\_cross-section.html](https://mrtrix.readthedocs.io/en/latest/fixel_based_analysis/mt_fibre_density_cross-section.html)

*mrtransform* and *mrmath*, on which fixel segmentation was performed with *fod2fixel* and the default threshold of 0.06 to compute a WM template fixel mask. Intra-subject and study-specific warps were then combined using *transformcompose*, and were subsequently applied to the individual unwarped FOD images with *mrtransform* to bring them into common space without FOD reorientation. FOD images were then segmented to identify the number and direction of fixels in each voxel using *fod2fixel*. Fixels of each image were reoriented in template space based on the combined warps using *fixelreorient* and assigned to template fixels using *fixelcorrespondence* to establish correspondence of fixels across images. This resulted in FD images in template space for each subject at each time point. FC was then calculated for each image using *warp2metric*, and  $\log(\text{FC})$  and FDC (the product of FD and FC) were calculated using *mrcalc*. Voxel-wise total FD (sum across directions), mean  $\log(\text{FC})$  (mean value across directions weighted by FD), and total FDC (sum across directions) were extracted using *fixel2voxel* and the resulting images were converted to Nifti format using *mrconvert*. Finally, FD,  $\log(\text{FC})$ , and FDC maps were smoothed in SPM12 with a 10-mm FWHM standard Gaussian kernel.

The pre-processing and estimation pipeline of MPMs that was applied in the young adult sample was also applied to the data from the older adult sample, resulting in smoothed R1, R2\*, PD, and MT maps. Within-subject, within-ROI mean values were extracted from MT maps using FSL for use in the GM structural integrity models in Paper II (Polk, Kleemeyer, Köhncke et al., 2022).

### 3.2.7 Cognitive testing

Out of the cognitive battery administered to the older adult sample, the DSST (described in Section 3.1.5) is of interest in this dissertation. Identical versions were administered to the older adult sample at T1, T2, and T3.

### 3.2.8 Statistical analysis

**Paper II.** SEM was used to investigate group differences in mean change in  $\text{VO}_2\text{peak}$  and GM structural integrity, as well as the correlation between changes in fitness and changes in integrity. Statistical analyses were conducted using R in RStudio using the OpenMx R package (version 2.19.8; Boker et al., 2021; Hunter, 2018; Neale et al., 2016; Pritikin et al., 2015) to specify and estimate models. Univariate outliers ( $\pm 4$  SD away from the mean) and multivariate outliers (detected using the classical product-moment method, criterion = .001, with the *robustMD* command from the *faoutlier* R package; Chalmers & Flora, 2015) were removed from the data. FIML was used for estimation, meaning missing data could be

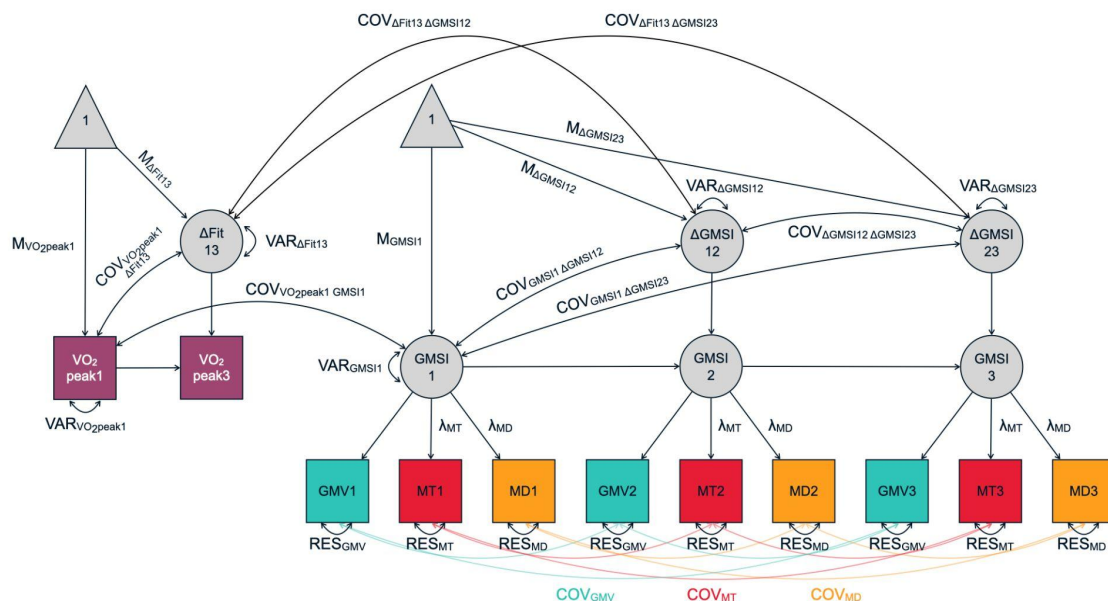


accounted for without the need for case-wise deletion. Manifest variables were scaled longitudinally (i.e., data from all time points were first stacked then rescaled to mean = 0, SD = 1) to preserve the relative mean relationships between the same variables measured at different time points. To evaluate whether models showed acceptable fit, thresholds of root mean square error of approximation (RMSEA) < .08 and comparative fit index (CFI) > .90 were used as rough guidelines (Schermelleh-Engel et al., 2003). The statistical significance of parameters of interest was evaluated using LRTs, as described in Section 3.1.6.

To test for group differences in change in VO<sub>2</sub>peak and GM structural integrity, multigroup latent change score models (LCSM) were built following the tutorial by Kievit and colleagues (2018). A univariate LCSM with a pseudo-latent factor capturing change in VO<sub>2</sub>peak was used to evaluate the difference in change in cardiovascular fitness from T1 to T3 between groups. A latent factor of GM structural integrity at T1, T2 and T3 capturing the shared variance between GMV, MT, and MD in the pre-selected ROIs was established, and multivariate LCSMs with latent variables capturing change in GM structural integrity from T1 to T2 and from T2 to T3 were run through a sequence of validation steps. Those models that survived testing for factorial invariance across groups and across time points were then tested for group differences in GM structural integrity change. Finally, if significant individual differences in change could be established (see Ghisletta & Lindenberger, 2004), the covariances between change in cardiovascular fitness and change in GM structural integrity were tested using bivariate LCSMs (see Figure 3 for an exemplary path diagram).

**Paper III.** Flexible factorial models, which can account for dependent data points (e.g., images from the same individual at different time points), were used to evaluate changes in WM integrity over time, looking for voxel-wise differences in change between groups as indicated by a significant time-by-group interaction effect. All statistical analyses were performed in SPM12 and CAT12. Smoothed WMV maps, smoothed FA and MD maps, and smoothed voxel-wise FD, log(FC), and FDC maps were entered into flexible factorial models with subject as a within-subject factor, time point as a within-subject factor (T1, T2, T3), and group as a between-subject factor (ACG, EG). Age, sex, and level of education were entered into the model as covariates of no interest. First, a conservative threshold of  $p < .050$  with correction for false discovery rate (FDR) at the peak-level, correction for non-isotropic smoothness, and minimum cluster size of  $k = 100$  was applied. If no clusters showing a significant time-by-group interaction were revealed at this threshold, a more liberal threshold of  $p < .001$ , only correcting for non-isotropic smoothness, and a minimum cluster size of  $k =$

100 was applied. Missing data were handled case-wise; four cases were excluded from the diffusion tensor-derived and fixel-based metrics due to missing DW images.



**Figure 3 | Exemplary path diagram of the bivariate LCSM measuring the correlation between change in cardiovascular fitness from T1 to T3 and change in GM structural integrity from T1 to T2 and from T2 to T3.**

*Note.* Fit = cardiovascular fitness; GMSI = gray-matter structural integrity; 1 = time point 1; 2 = time point 2; 3 = time point 3; Δ = change; M = mean; COV = covariance; VAR = variance; λ = loading; RES = residual. Covariances between manifest variables of integrity were set to be equal within each modality, denoted by color-coding; these are only labeled once for visual clarity. Other unlabeled paths are fixed to 1.

From Polk, Kleemeyer, Köhncke et al. (2022). © 2022 Polk and colleagues.

## OVERVIEW OF INDIVIDUAL WORKS

Papers I and II sought to validate a longitudinal multivariate model of GM structural integrity. In Paper I, we investigated the reliability of the quantitative MR measure of MT to measure between-person differences in a longitudinal context. In Paper II, we used this quantitative measure, along with the well-established measures of GMV (Hedges et al., 2020) and MD (Luque Laguna et al., 2020), in a multivariate model of GM structural integrity to estimate the change in a latent factor capturing the shared variance between these metrics of macro- and microstructural integrity in the brain, and to test for group differences in mean change in this latent factor of GM structural integrity. Paper III investigated the effects of aerobic exercise on WM integrity in older adults using a whole-brain approach with both established and novel metrics of WM integrity.

### 4.1 Paper I

Wenger, E., **Polk, S. E.**, Kleemeyer, M. M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for MT and PD but attenuated for R1 and R2\* in healthy young adults. *Human Brain Mapping*. <https://doi.org/10.1002/hbm.25870>

#### 4.1.1 Objective

In Paper I, in order to test how reliably the recently developed comprehensive quantitative MPMs can measure individual differences over time, we used a two-day, four-session design and the ICED model to parse different sources of error and to isolate the variance attributed to repositioning in the scanner and scanning on different days, to random error, and to the variance resulting from true between-person differences. In this way, we sought to validate the use of these quantitative metrics, including MT, in longitudinal study designs interested in the correlation between change in these metrics and other variables of interest, such as indices of cognitive performance. If a relatively large proportion of variance could be attributed to true between-person differences, we concluded that the respective quantity could reliably measure individual differences, and that any subsequent correlational analyses between the metric of interest and other variables would also be reliable.

### **4.1.2 Main findings**

In the current sample, most variance of MT in whole-brain GM and WM could be attributed to true between-person differences. No significant day- or session-specific variance was detected, indicating robustness against repositioning in the scanner both across days and across same-day sessions. This was also true for most localized ROIs, except for the orbitofrontal cortex and caudate. Similarly, most variance of PD in whole-brain GM and WM could be attributed to true between-person differences, with no significant day- or session-specific variance. This was true for most cortical ROIs as well, except for the orbitofrontal cortex, but not for subcortical ROIs. Estimated true score variances indicated that measures of R1 were less reliable, but were still robust against repositioning. Indeed, for R1, whole-brain GM and WM as well as all localized ROIs showed estimated true score variance of less than 75% of the total variance, with large proportions of variance up to over 50% being explained by random error. Most variance of R2\* in whole-brain GM and WM could be attributed to true between-person differences, with no significant day- or session-specific variance. This was also true for the putamen, pallidum, and the anterior cingulate cortex, but estimated true score variance was relatively low in the other ROIs, and was not statistically distinguishable from zero in the precuneus.

In sum, using the ICED model, we could establish good reliability of the quantitative MT, PD, and R2\* measures in terms of capturing individual differences in whole-brain GM and WM, but region-specific reliability varied. Importantly in the context of this dissertation, this finding indicates that the quantitative measure of MT can reliably detect individual differences and is robust against effects of repositioning, which validates its use to investigate changes in brain structure longitudinally, as well as correlations with other variables of interest, such as change in physical fitness as a result of aerobic exercise.

## 4.2 Paper II

**Polk, S. E.,** Kleemeyer, M. M., Köhncke, Y., Brandmaier, A. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Wenger, E., & Düzel, S. (2022). Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2022.852737>

### 4.2.1 Objective

In Paper II, we aimed to validate a longitudinal model of change in a latent factor of GM structural integrity, and to estimate group differences in mean change in GM structural integrity. We also investigated correlations between exercise-induced changes in cardiovascular fitness and in GM structural integrity.

### 4.2.2 Main findings

Older adults who exercised regularly at home at a moderate and highly personalized intensity exhibited increases in  $VO_2$ peak, an index of cardiovascular fitness, while older adults who did not exercise and instead participated in a book club showed no change in  $VO_2$ peak. The difference in change between groups, the time-by-group interaction effect, was also statistically significant, indicating that the increase in cardiovascular fitness was indeed driven by engaging in aerobic exercise. We were able to establish factorial invariance across groups and time points in eight models of GM structural integrity: right and left hippocampus, right and left anterior cingulate cortex, right posterior cingulate cortex, right and left juxtapositional lobule cortex, and right inferior frontal gyrus. Significant time-by-group interaction effects were found for GM structural integrity in the right and left anterior cingulate cortex, right posterior cingulate cortex, and left juxtapositional lobule cortex. In all of these ROIs, the interaction effect was driven by negative changes in control participants, while exercisers showed no significant increases, indicating a maintenance effect of aerobic exercise on GM structural integrity. Significant individual differences in change were able to be established in the right anterior cingulate cortex and left juxtapositional lobule cortex, allowing us to investigate change-change correlations with cardiovascular fitness. We found significant positive correlations between change in cardiovascular fitness and change in GM structural integrity in both regions. Furthermore, within-group correlation coefficients were significantly different in the right anterior cingulate cortex, with the control group showing no significant correlation and the exercisers showing a significant positive correlation.

In sum, results reported in Paper II indicated that moderate, at-home aerobic exercise was sufficient to boost cardiovascular fitness in healthy, previously sedentary older adults. Additionally, engaging in aerobic exercise resulted in maintenance of GM structural integrity in areas of the brain that were previously shown to exhibit exercise-induced maintenance of GMV. Finally, the positive association between change in cardiovascular fitness and change in GM structural integrity in the right anterior cingulate cortex in combination with the significant time-by-group interaction effects suggests that physical fitness and GM structural integrity in this region change concurrently in older adults in response to aerobic exercise.

### 4.3 Paper III

**Polk, S. E.,** Kleemeyer, M. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Düzel, S., & Wenger, E. (2022). *Aerobic exercise is associated with region-specific changes in volumetric, tensor-based, and fixel-based measures of white matter integrity in healthy older adults* [Manuscript under revision].

#### 4.3.1 Objective

In Paper III, we investigated the effect of aerobic exercise on WM integrity using a number of established and novel metrics of WM integrity. We sought to replicate findings previously reporting maintenance of WMV in anterior brain regions, as well as to investigate the effects of exercise on FA and MD in the current sample, as findings regarding FA and MD are currently mixed. We also used newly established fixel-based metrics of WM integrity to evaluate exercise effects on the density and diameter of WM fibers in a voxel-wise manner. Finally, we investigated associations between percent change in  $VO_{2peak}$  and DSST score with percent change in WM metrics where we saw significant time-by-group interaction effects.

#### 4.3.2 Main findings

We were able to replicate previous reports of exercise-induced maintenance of WMV in the current sample of older adults. Specifically, we saw voxel clusters showing significant time-by-group interactions driven by maintenance (i.e., no mean change in exercisers and significant decreases in controls) in the right anterior corona radiata extending into the genu of the corpus callosum as well as in the splenium of the corpus callosum. Percent change in WMV in the splenium from T1 to T3 was positively correlated with percent change in  $VO_{2peak}$ , which increased more in exercisers versus controls, and percent change in WMV in the right anterior corona radiata/genu was positively correlated with percent change in DSST score, which showed a non-significant trend toward greater improvement within exercisers than controls. One cluster of significant group differences in change in FA was found in the left part of the genu of the corpus callosum, and clusters of group differences in change in MD were found in the right posterior corona radiata extending into the splenium of the corpus callosum and the right superior longitudinal fasciculus. However, changes in these clusters were in the opposite direction than that which we hypothesized; namely, exercisers showed decreases in FA and increases in MD while controls showed increases in FA and decreases in MD. Similarly, clusters of voxels showing significant time-by-group interaction

effects in FD and FDC were seen in the dorsomedial and dorsolateral prefrontal cortex in superficial WM, but with within-group changes occurring in the unexpected direction. In both clusters, exercisers showed decreases in total FD and FDC while controls showed increases. However, percent change in FD and FDC in the dorsomedial prefrontal cluster was negatively correlated with percent change in  $VO_{2peak}$ , indicating that greater fitness gains were associated with more negative changes in these metrics of WM integrity. Furthermore, weak negative correlations were found between percent change in FD and FDC in the dorsolateral prefrontal cluster and percent change in DSST score. This correlation was significantly different between groups, with exercisers showing a significant negative correlation and controls showing no significant association. This indicates that the exercise-induced decrease in these WM metrics observed in this sample was associated with better cognitive outcomes.

In sum, the findings reported in Paper III corroborated previous reports that aerobic exercise can induce maintenance of WMV in the corpus callosum, and that these changes in WMV are associated with both changes in cardiovascular fitness as well as a cognitive outcome indexing general cognitive acuity. Our results also support the existing skepticism of the interpretation that greater increases in FA and decreases in MD are invariably beneficial in aging. Similarly, findings of decreased FD and FDC in exercisers, as well as the negative associations between change in these metrics and change in  $VO_{2peak}$  and DSST score suggest that interpreting higher FD and FDC as universally beneficial in aging may not be precise enough, and that more nuanced, region-specific investigation (e.g., between deep and superficial WM) into age-related changes in these voxel-based metrics may be necessary.



*Chapter 5.*  
**GENERAL DISCUSSION**

This dissertation aimed to use novel models and metrics of brain tissue integrity to investigate the effects of moderate, at-home aerobic exercise in older adults, focusing on the questions outlined in Chapter 2.

### **5.1 Discussion of research questions and hypotheses**

Questions 1 and 2 were concerned with the reliability of quantitative MPMs. This was investigated within a small sample of young adults ( $N = 15$ , mean age = 27.3 years) who underwent four measurement sessions across two days. Questions 3 through 7 were concerned with the effects of aerobic exercise on brain tissue integrity in older adults. This was investigated with a sample of older adults ( $N = 75$ , mean age = 70.3 years), who participated in an at-home intervention study for six months.

#### **5.1.1 Question 1**

*What are the proportions of variance that can be attributed to repositioning a participant in the scanner either on the same day or across days, to random error, and to between-person differences?* We found that, within a sample of young adults, reliability in whole-brain GM and WM was acceptable (i.e., estimated value of  $\text{Var } T \geq 75\%$ ) for MT, PD, and R2\*, but not for R1, with between 78% and 88% of the total variance attributable to true between-person differences for MT, PD, and R2\*, but only 73% of variance attributable to true between-person differences for R1. We also found that all measures in whole-brain ROIs were robust against effects of repositioning, both across days as well as on the same day across sessions, with negligible proportions of variance of 0% to 9% (not significantly different from zero) attributable to day- and session-specific effects. The rest of the total variance was attributed to random measurement error, with between 6% and 17% of total variance attributable to random error for MT, PD, and R2\*, and 24% to 27% attributable to random error for R1.

Regarding the attenuated reliability of R1 to capture true individual differences in whole-brain GM and WM, R1 maps have an increased sensitivity to myelination and are thus widely used for myelin mapping in both in vivo and post-mortem studies (see Weiskopf et al., 2021). It could be, however, that within healthy, neurotypical young adults, individual differences in whole-brain myelination are not particularly pronounced. Reliability of R1 (as

well as the other quantitative maps) should be investigated within populations with known myelin abnormalities, such as in patients with schizophrenia, bipolar disorder, or autism, or within aging participants or patients with Alzheimer's disease (T. W. Chapman & Hill, 2020).

### 5.1.2 Question 2

*Is this reliability consistent or does it vary across different regions in the brain?* We found that, within a sample of young adults, reliability varied notably across regions of the brain. The reliability of MT and PD was acceptable in all cortical ROIs except for the orbitofrontal cortex. In subcortical regions, reliability was also acceptable in the putamen and pallidum for MT, but not in the caudate. For PD, reliability was unacceptable in all subcortical regions. The reliability of R1 was unacceptable in all ROIs. For R2\*, reliability was acceptable only in the anterior cingulate cortex, putamen, and pallidum, and attenuated in the other ROIs. Researchers should use caution when interpreting inter-individual differences in those areas where measures showed attenuated reliability, especially with for MT in the caudate and R1 in the anterior cingulate cortex, orbitofrontal cortex, and caudate, where estimates of random error variance exceeded those of true score variance.

Similar to R1, R2\*, which can be used as a marker of iron concentration (Weiskopf et al., 2021), may only show measurable individual differences in certain ROIs later in life. Indeed, in a sample of 80 individuals aged one to 80 years old, the standard deviation of R2\* measurements within the ten participants in their twenties was descriptively lower for frontal WM than for regions in the basal ganglia except for the caudate, while participants in their sixties and seventies showed greater standard deviations in in all basal ganglia ROIs as well as frontal WM (Aquino et al., 2009). Again, investigations of reliability within aging and diseased populations should be conducted in the future to understand the reliability of this parameter in other samples.

As discussed in Section 4.1.2, the finding that MT can reliably measure individual differences longitudinally in a number of cortical ROIs is particularly important in the context of this dissertation, as we used MT in a latent factor of GM structural integrity in Paper II (Polk, Kleemeyer, Köhncke et al., 2022) to investigate the effects of aerobic exercise on tissue integrity in GM in similar ROIs (e.g., anterior cingulate cortex, inferior frontal gyrus), as well as the correlations between change in GM structural integrity in these regions and change in cardiovascular fitness.

### 5.1.3 Question 3

*Can moderate, at-home exercise induce an increase in cardiovascular fitness in healthy, previously sedentary older adults?* We found that cardiovascular fitness, measured with  $\text{VO}_2\text{peak}$ , improved in older adults who exercised regularly for six months, with a mean change of 2.6 mL/kg/min (standard error [SE] = 0.45; mean percent change = 12.8%), while it did not change for those who did not exercise, with a mean percent change of 0.7 mL/kg/min (SE = 0.54; mean percent change = 3.7%). Furthermore, we found a significant time-by-group interaction, indicating that the change within the EG was significantly greater than the change within the ACG. This confirmed our hypothesis that moderate, at-home exercise can induce positive change in cardiovascular fitness in older adults who had not previously exercised regularly, consistent with reports by King (1991) and Salvetti and colleagues (2008). This finding also serves as a proof of concept that our at-home intervention design was fundamentally effective, even though it differed from other exercise interventions that invited participants to the lab multiple times a week to participate in supervised exercise sessions (e.g., Colcombe et al., 2006; Erickson et al., 2011; Kleemeyer et al., 2016; Maass et al., 2015; Ruscheweyh et al., 2011; Voss et al., 2013).

Parenthetically, although a difference in baseline  $\text{VO}_2\text{peak}$  was observed between males and females, consistent with aerobic fitness norms in older adults (Shvartz & Reibold, 1990), with means of 25.5 mL/kg/min (SD = 5.99 mL/kg/min) and 20.7 mL/kg/min (SD = 4.46 mL/kg/min), respectively, there was no significant effect of sex on change in  $\text{VO}_2\text{peak}$ . Within the EG,  $\text{VO}_2\text{peak}$  in both males and females improved, with a mean change of 2.8 mL/kg/min (SE = 0.65; mean percent change = 11.3%) and 3.4 mL/kg/min (SE = 0.63; mean percent change = 13.7%), respectively, while in the ACG,  $\text{VO}_2\text{peak}$  did not improve in either males or females, with a mean change of 1.2 mL/kg/min (SE = 0.79; mean percent change = 5.6%) and 0.0 mL/kg/min (SE = 0.64; mean percent change = 0.9%), respectively. This suggests that an at-home exercise program for older adults is effective for both males and females.

### 5.1.4 Question 4

*Can moderate, at-home exercise induce positive changes and/or maintenance in GM structural integrity in areas of the brain that have previously shown volumetric effects of exercise?* We found that GM structural integrity in the right and left anterior cingulate cortices, right posterior cingulate cortex, and left juxtapositional lobule cortex (also known as the supplementary motor area) was maintained in exercisers, that is, a significant time-by-

group interaction effect was detected whereby exercisers showed no significant increases in these areas but control participants did show significant decreases. This confirmed our hypothesis that moderate, at-home can induce maintenance of GM structural integrity, consistent with volumetric maintenance effects seen in Colcombe et al. (2006). We did not find evidence that aerobic exercise can boost GM structural integrity, as Erickson and colleagues (2011) found in hippocampal volume.

Notably, it seemed that change in the latent factor of GM structural integrity in all of the areas showing significant time-by-group interactions was not mostly driven by change in GMV. Indeed, latent GM structural integrity had the greatest effect on observed MD, which had an average standardized loading of  $-.771$  across the four ROIs showing significant interaction effects, followed by MT, which had an average loading of  $.496$ , and GMV, which had an average loading of  $.462$ . Therefore, we can infer that latent GM structural integrity is not simply a latent variable with a high correlation to GMV, in which case the changes seen in response to exercise could have been attributed to changes in GMV as seen in Colcombe et al. (2006), but rather that latent GM structural integrity indeed captures a more general latent measure of integrity that influences the three indicators of macro- and microstructural integrity used here: GMV, MT, and MD.

A potential mechanism underlying these exercise effects on GM structural integrity could be the increase in cerebral blood flow (CBF) induced by bouts of aerobic exercise. The hypothesized causative factor underlying this increase in CBF is the increased brain metabolism during exercise, but a number of other factors may also play a modulating role, including partial pressure of carbon dioxide, muscle mechanoreceptors, and arterial blood pressure (see Ogoh & Ainslie, 2009; Querido & Sheel, 2007). These acute increases in CBF would regularly supply the surrounding tissue with oxygen and other nutrients carried in the blood, improving the maintenance of brain cells in exercisers, which in turn could be what is reflected in the maintenance of latent GM structural integrity seen here. Indeed, short-term intervention studies of eight weeks to four months in older adults found that CBF was increased even at rest in the anterior cingulate cortex within exercisers versus controls (Burdette et al., 2010; S. B. Chapman et al., 2013; Kleinloog et al., 2019), and increases in CBF in other regions of the brain, such as the hippocampus, have also been found to be related to exercise-induced cardiovascular fitness changes (Maass et al., 2015). Future research should therefore seek to understand the relationship between changes in CBF, either acute or long-lasting, resulting from aerobic exercise and changes in GM structural integrity in aging.

In addition to addressing the empirical questions posed in this dissertation, these findings also highlight some other important methodological and empirical points: First, we were able to establish factorial invariance in GM structural integrity across three time points in a number of ROIs, validating the use of this latent factor in a longitudinal model in the current sample. Additionally, we found declines in GM structural integrity over the course of six months in older control participants using this longitudinal model, corroborating the cross-sectional age associations found by Köhncke and colleagues (2021).

### 5.1.5 Question 5

*Is change in cardiovascular fitness positively correlated with changes in GM structural integrity?* We found that change in cardiovascular fitness, measured with  $VO_{2peak}$ , was positively correlated with change in GM structural integrity in both the right anterior cingulate cortex and left juxtapositional cortex (also known as the supplementary motor area). Furthermore, we found that the within-group correlation coefficients were significantly different for the correlation between cardiovascular change and GM structural integrity change in the right anterior cingulate cortex. This confirmed the hypothesis that exercise-induced changes in fitness and changes in GM structural integrity are positively associated, consistent with a number of studies discussed in Erickson et al. (2014).

The intervention-nonspecific finding that change in  $VO_{2peak}$  is correlated with change in GM structural integrity in the left juxtapositional lobule cortex suggests that on an individual level, changes in cardiovascular fitness are generally accompanied by maintenance of GM structural integrity in this region, regardless of exercise status. The exercise group-specific finding that change in  $VO_{2peak}$  is correlated with change in GM structural integrity in the right anterior cingulate cortex, in contrast, indicates that individuals who gained more in cardiovascular fitness as a result of engaging in aerobic exercise also showed better maintenance of GM structural integrity in this region, and suggests a causal relationship between aerobic exercise, cardiovascular fitness, and GM structural integrity in the right anterior cingulate cortex. Notably, as the relationship between changes in fitness and integrity here was simply correlational, we can make no claims about the causal relationship between these variables, and future research should be conducted to understand whether these changes are simply concurrent and are both caused by aerobic exercise, or if there is additional directionality within this relationship.

### 5.1.6 Question 6

*Can moderate, at-home exercise induce changes in metrics of WM integrity, namely WMV, diffusion tensor-derived FA and MD, and fixel-based FD, log(FC), and FDC? We found that WMV in the anterior and posterior corpus callosum was maintained in exercisers, demonstrated by a significant time-by-group interaction driven by decreases in control participants, consistent findings reported in Colcombe et al. (2006). We also found significant time-by-group effects on FA, MD, FD, and FDC in a number of voxel clusters, however, within-group effects in these clusters went in the opposite direction than that which was hypothesized. We saw decreases in FA in the genu of the corpus callosum and increases in MD in the right posterior corona radiata/splenium of the corpus callosum and right superior longitudinal fasciculus within the EG, consistent with the findings reported by Clark and colleagues (2019), but in those same regions, the ACG showed increases in FA and decreases in MD. Additionally, contrary to our hypotheses based on the age-related findings on the fixel-based metrics reported in Choy et al. (2020) and Kelley et al. (2021), we found decreases in FD and FDC in two clusters (overlapping across metrics) in the dorsomedial and dorsolateral prefrontal cortex within the EG, whereas the ACG showed increases in these clusters.*

Regarding the cluster of FA and MD in the context of aging, previous studies have also found results inconsistent with the “more is better” principle. For example, in a study comparing different metrics of WM integrity in young (19–29 years old) and older adults (65–87 years old), the older group showed greater values of FA in a number of brain regions than the younger group, including in areas similar to where the FA cluster found in Paper III (Polk, Kleemeyer, Bodammer et al., 2022) was localized (Kelley et al., 2021). This suggests that increases in FA in certain areas of the brain may actually be consistent with age-related declines, and that exercise-induced decrease in FA could be an amelioration of these declines. Additionally, the validity of FA and MD has also been questioned in recent years due to the inability of the diffusion-tensor model to accurately resolve diffusivity in voxels containing crossing fibers (Raffelt et al., 2012, 2015, 2017), which account for an estimated 60–90% of voxels in the brain (Jeurissen et al., 2013). Indeed, Kelley and colleagues (2021) also found widespread strong negative voxel-wise correlations between FA and a measure of the number of crossing fibers in each voxel. FA and MD should thus be interpreted cautiously; greater FA and reduced MD may not always be beneficial in the context of aging and aerobic exercise.

Because of the weakness of the diffusion-tensor model in areas of crossing fibers, new techniques were developed to model WM fibers in multiple directions per voxel, such as

multi-shell multi-tissue CSD, which was used in Paper III (Polk, Kleemeyer, Bodammer et al., 2022) to estimate voxel-wise FD, log(FC), and FDC. The results we found regarding FD and FDC were also surprising, however, as in both clusters, FD and FDC decreased within the EG but increased within the ACG. Notably, percent change in both FD and FDC correlated negatively with percent changes in cardiovascular fitness and cognitive performance; this finding is discussed in the following Section (5.1.7). The exercise effect on FD and FDC suggests that, similarly to FA and MD, more is not always better, and that age-related increases in these metrics in anterior WM may be ameliorated by aerobic exercise. Of note, clusters in both the dorsomedial and dorsolateral prefrontal cortex were found in areas bordering the cortex, that is, in superficial WM, which may be more susceptible to noise during MR acquisition than deep WM (Guevara et al., 2020; Jeurissen et al., 2014; Kirilina et al., 2020).

In order to more deeply understand the relationship between the novel fixel-based metrics and the more established measures of GMV, FA, and MD, we extracted FD, log(FC), and FDC values from each cluster showing significant interaction effects in the former metrics and calculated Pearson's correlation coefficients between metrics at baseline as well as change-change correlations (see Table 3). At baseline, metrics were generally correlated in the expected direction, with positive correlations seen between GMV and all fixel-based metrics in both clusters, and positive correlations seen between FA and all fixel-based metrics in the left genu cluster. MD values correlated negatively with FD and FDC in the right posterior corona radiata/splenium cluster and with FDC in the right superior longitudinal

Table 3

*Baseline and change-change correlation coefficients and uncorrected p-values (in parentheses) between white matter integrity metrics in clusters of voxels showing time-by-group interaction effects in GMV, FA, and MD*

	GMV in right ACR/genu	GMV in splenium	FA in left genu	MD in right PCR/splenium	MD in right SLF
Baseline correlations with fixel-based metrics in the same cluster					
FD	0.26 (0.050)	0.41 (0.001)	0.51 (0.000)	-0.51 (0.000)	-0.22 (0.091)
log(FC)	0.91 (0.000)	0.76 (0.000)	0.31 (0.018)	0.31 (0.016)	-0.25 (0.063)
FDC	0.76 (0.000)	0.73 (0.000)	0.51 (0.000)	-0.30 (0.022)	-0.32 (0.015)
Change-change correlations with fixel-based metrics in the same cluster					
FD	-0.01 (0.922)	0.04 (0.774)	0.10 (0.445)	0.11 (0.425)	0.00 (0.990)
log(FC)	-0.22 (0.108)	0.05 (0.685)	-0.10 (0.446)	0.15 (0.281)	0.05 (0.719)
FDC	-0.01 (0.931)	0.04 (0.765)	0.11 (0.425)	0.11 (0.435)	0.00 (0.972)

*Note.* ACR = anterior corona radiata; PCR = posterior corona radiata; SLF = superior longitudinal fasciculus. Insignificant correlation coefficients ( $p_{\text{uncorrected}} > .050$ ) are italicized.

fasciculus cluster, but positively with log(FC) in the right posterior corona radiata/splenium cluster. However, no change-change correlations were seen between GMV, FA, or MD and any of the fixel-based measures, indicating that these metrics do not seem to capture changes in the same tissue properties. More research is necessary to understand the longitudinal fixel-based metrics, especially in the context of age-related changes as well as aerobic exercise.

### 5.1.7 Question 7

*Are changes in WM integrity correlated with change in cardiovascular fitness or change in cognitive performance?* Regarding correlations with change in cardiovascular fitness, we saw a positive correlation between percent change in VO<sub>2</sub>peak and percent change in WMV in the splenium of the corpus callosum, confirming our hypothesis based on the studies discussed in Sexton et al. (2016). We also found negative correlations between percent change in VO<sub>2</sub>peak and percent change in both FD and FDC in a cluster of voxels in the dorsomedial prefrontal cortex. This finding was unexpected, but suggests that decreases in FD and FDC in this region are related to increases in cardiovascular fitness. As with the correlations between change in cardiovascular fitness and change in GM structural integrity, these change-change correlations indicate a concurrent effect of aerobic exercise on both cardiovascular fitness and WM integrity, but whether these two changes are causally related in some way cannot be inferred solely from correlational analyses.

Regarding correlations with change in cognitive performance, we saw a positive correlation between percent change in WMV in the anterior corona radiata/genu of the corpus callosum and percent change in DSST, supporting our hypothesis based on Brickman et al. (2006) and the brain maintenance hypothesis (Cabeza et al., 2018; Johansson et al., 2022; Lindenberger, 2014; Nyberg et al., 2012; Nyberg & Lindenberger, 2020; Nyberg & Pudas, 2019) that individual differences in WM integrity are related to individual differences in cognitive performance. This also corroborates existing cross-sectional evidence that volume in the anterior corpus callosum is related to DSST performance (Fling et al., 2011). Altogether, these findings suggest that aerobic exercise is causally related to WMV in the anterior corpus callosum, and that this change is associated with improvements in a cognitive task that recruits attention and perceptual speed. However, this finding should not be interpreted as evidence for a true transfer effect, as we did not find a significant difference in change between groups in DSST score.

We also found weak negative correlations between percent change in DSST and percent change in both FD and FDC in a cluster of voxels in the dorsolateral prefrontal



cortex. Furthermore, the correlation coefficients between change in DSST and change in FD and FDC were different across groups, with significant negative correlations within the EG but not within the ACG. This finding was also unexpected, but again, given the decreases in both FD and FDC in this cluster within the EG, suggests that individual differences in exercise-induced decreases in FD and FDC may be able to predict individual differences in DSST performance, also consistent with the brain maintenance hypothesis (Cabeza et al., 2018; Johansson et al., 2022; Lindenberger, 2014; Nyberg et al., 2012; Nyberg & Lindenberger, 2020; Nyberg & Pudas, 2019). Although surprising, this finding indicates that those individuals who showed greater exercise-induced decreases in FD and FDC showed greater improvements in performance on the DSST, and suggests a causal relationship between aerobic exercise, FD and FDC in the dorsolateral prefrontal cortex, and performance on the DSST. However, as we did not find a significant interaction effect on DSST score, this finding should again not be interpreted as a true transfer effect. More research should be conducted to understand the directionality of the relationship between these variables; according to Raffelt and colleagues (2012, 2015, 2017), FD and FDC should reflect WM fibers' "ability to relay information," with higher values indicating better information relay. However, this negative correlation between percent change in FD and FDC and percent change in DSST score, which can be considered a measure of perceptual speed, challenges this notion, and suggests that the directionality of the association between fixel-based metrics and information relay may be differentiated across regions (e.g., in deep vs. superficial WM).

## 5.2 Limitations

There are, of course, a certain number of limitations in this dissertation that should be discussed. Regarding Paper I (Wenger, Polk et al., 2022), the small sample consisted of healthy young adults, and generalizability is therefore limited. Furthermore, conclusions were derived from the estimated ICCs, which are standardized against between-person variance, and therefore necessarily contingent on the sample included in the estimation. Notably, in Paper II (Polk, Kleemeyer, Köhncke et al., 2022), we used the quantitative measure of MT in a model of latent GM structural integrity and were able to detect individual differences in change in this latent factor. However, the MPMs should still be evaluated regarding their reliability on the manifest level in other age groups, particularly R1 and R2\*, given their dependency on physiological parameters in which individual differences may only be apparent in older age.

Regarding the sample included in Papers II (Polk, Kleemeyer, Köhncke et al., 2022) and III (Polk, Kleemeyer, Bodammer et al., 2022), we included a moderately sized sample ( $N = 75$ ) due to the interventional nature of the study and the fact that we investigated the effects of aerobic exercise versus no exercise in a sub-sample of the full study, which included two further groups that participated in a cognitive training. This sample size is considered small for SEM, which could partially explain why we could only establish reliable individual differences in change in two of the regions showing group differences in change. Additionally, only those participants who fully adhered to the interventions were included in the analyses of WM integrity changes ( $n = 61$ ), as the flexible factorial model in SPM12 cannot account for data missing at random, in contrast to FIML. While this sample size is comparable to or even greater than sample sizes included in other exercise interventions (e.g.,  $N = 59$  in Colcombe et al., 2006;  $N = 40$  in Maass et al., 2015), including more participants could have increased our power to detect effects that would have survived correction for multiple comparisons. But given that certain voxel clusters in Paper III (Polk, Kleemeyer, Bodammer et al., 2022) did not survive FDR correction within the current sample, namely those of group differences in change in FA, MD, FD, and FDC, these results should be interpreted with caution.

Participants in both the young and older adult samples formed fairly homogeneous groups. As discussed in Sections 5.1.1 and 5.1.2, the young adult sample comprised healthy, neurotypical adults who may not show the same individual differences in certain physiological properties in brain tissue as older populations or patients. For the AKTIV study, we only recruited healthy, sedentary older adults in an area with a relatively high average socioeconomic status. These individuals may be equipped with other factors that are protective against an inactive lifestyle, which can lead to accelerated deterioration in other individuals. The generalizability of conclusions inferred from the papers included in this dissertation is therefore limited, and further research should be conducted to understand the reliability of quantitative MPMs, as well as the association between aging, aerobic exercise, and brain tissue integrity in less homogeneous populations.

Regarding the interventions implemented in the AKTIV study, it should be noted that the active control condition was conceptualized with the four-group design in mind. In addition to an exercise-only condition, the full study included also a language learning-only condition and a combined condition, comprising both aerobic exercise and language learning. In order to have a comparable control group for all conditions, the ACG did not engage in typical control activities for exercise intervention studies, such as stretching and toning (e.g.,

Colcombe et al., 2006; Erickson et al., 2011; Maass et al., 2015), but instead participated in a book club. Thus the comparisons made in the current dissertation do somewhat differ from those made in pure exercise interventions. However, given that we aimed to include an “active” control group, individuals in the control condition also participated in a significant amount of study-related activity and weekly group sessions, meaning that we can reasonably rule out effects of an overall change in lifestyle, such as a new daily routine, new social partners, or increased interaction with technology on changes in variables of interest in this dissertation.

Finally, although we found certain change-change correlations between WM integrity and performance the DSST, comparisons of control participants versus exercisers in the older sample did not reveal any significant effect of aerobic exercise on any cognitive task administered in the cognitive batteries at T1, T2, and T3. Therefore, we cannot make any inferences about so-called transfer effects. Perhaps if the intervention had taken place over a longer period of time (e.g., one year, as in Erickson et al., 2011), we would have been able to detect a significant difference in change between groups on DSST performance; in Paper III (Polk, Kleemeyer, Bodammer et al., 2022), we found a non-significant trend toward greater improvement in the EG than the ACG,  $F(2,102) = 2.696$ ,  $p = .072$ , Hedge's  $g = 0.010$ . Indeed, EG participants improved significantly from T1 to T3, while ACG participants showed no significant change. Assuming that either exercisers would continue to improve over a longer period of time or that age-related declines in performance would be detectable within control participants, a significant effect may have been found after a year. However, the lack of evidence for an effect of aerobic exercise on cognitive performance is also consistent with meta-analyses (Kelly et al., 2014; Young et al., 2015).

### 5.3 Future directions

Regarding Paper I (Wenger, Polk et al., 2022), as discussed in previous Sections (5.1.1, 5.1.2, 5.2), quantitative MPMs should be evaluated on their reliability within samples of populations other than healthy young adults. R1 and R2\*, for example, are dependent on physiological parameters which may be relatively homogeneous within a healthy, neurotypical sample, with between-person variance in these measures only developing later in life or found among populations with abnormalities in myelination or iron deposition.

Regarding Paper II (Polk, Kleemeyer, Köhncke et al., 2022), future analyses should be conducted to understand the mechanisms underlying changes in GM structural integrity in exercisers. For example, within the AKTIV study MRI protocol, we implemented arterial

spin labeling (ASL) sequences to measure perfusion in the brain (see Wenger, Düzel et al., 2022 for details), thus we could investigate whether perfusion improves with exercise, as found in previous interventions (e.g., Burdette et al., 2010; S. B. Chapman et al., 2013; Kleinloog et al., 2019), or if cardiovascular fitness improvement is related to perfusion, as found in Maass et al. (2015), as well as whether changes in perfusion are related to changes in GM integrity. Alternatively, if acute increases in CBF resulting from bouts of aerobic exercise underlie the maintenance of GM tissue found in the current dissertation, we may not be able to detect any longitudinal changes in cerebral perfusion.

It would also be possible to explore functional implications of the changes seen in brain structure in the current dissertation, as resting state functional MRI was also acquired in the AKTIV study (see Wenger, Düzel et al., 2022 for details). Research on aging populations has found overall reduced connectivity in older adults (Dennis & Thompson, 2014; Ferreira & Busatto, 2013). However, studies investigating the effects of aerobic exercise on connectivity in older participants have found some evidence that exercise can ameliorate age-related disconnectivity. For example, Burdette and colleagues (2010) found greater connectivity between the anterior cingulate cortex and the hippocampus at post-test in a small sample of older adults who participated in an exercise intervention versus controls, although they did not acquire MRI at pre-test and could therefore not directly investigate changes in connectivity. In a cross-sectional study comparing young and older adults, Voss and colleagues (2010) found that cardiovascular fitness was associated with improved connectivity in the older adults in networks that showed differences between the age groups. Given the longitudinal and interventional design of the AKTIV study, we would be equipped to investigate changes in connectivity, and also explore whether the structural changes reported in the current dissertation have functional implications on the level of brain connectivity.

Regarding Paper III (Polk, Kleemeyer, Bodammer et al., 2022), further research should be conducted to better understand the longitudinal relationship between WMV, diffusion tensor-derived metrics, and fixel-based metrics of WM integrity, as well as the underlying physiological properties of changes in each of these metrics. In the current dissertation, we found significant baseline correlations but no evidence for change-change correlations between WMV, FA, or MD and FD, log(FC), or FDC, indicating that these measures may not capture the same tissue properties in WM. Future studies may consider including additional MR sequences specifically targeting superficial WM to more deeply examine the FD and FDC findings in the current dissertation, or even including histology in

animal samples to examine the physiological changes underlying these changes. Additionally, studies that aim to investigate age-related changes in fixel-based metrics may consider using higher diffusion weightings than the ones used in the sequences acquired in current sample of older adults ( $b$ -values = 710 s/mm<sup>2</sup> and 2850 s/mm<sup>2</sup>), as a recent report of FD in children and adolescents showed that FD values estimated from multi-shell data with higher  $b$ -values appear to be more sensitive to age associations (Genc et al., 2020).

## 5.4 Conclusion

Altogether, the three works included in this cumulative dissertation add to the evidence that aerobic exercise can be an effective modifiable lifestyle factor to slow the age-related deterioration of brain tissue integrity. Specifically, we first showed that a recently established quantitative measure, MT, can reliably measure inter-individual differences in a number of brain regions that have shown age-related declines in GM integrity. Second, we used this quantitative measure alongside established MRI-based measures of GMV and MD to longitudinally model GM structural integrity on a latent level, parsing out the random measurement error to which observed variables are susceptible. Using this longitudinal model, we found that aerobic exercise can induce maintenance of latent GM structural integrity, and furthermore, that exercisers who showed greater changes in cardiovascular fitness also showed better maintenance of GM structural integrity in the right anterior cingulate cortex. Third, we investigated the effects of aerobic exercise on both established and newly developed metrics of WM integrity and found that WMV was maintained within exercisers versus controls. Our findings also add to the existing skepticism that the “more is better” principle is always applicable when interpreting age-related changes in the diffusion tensor-derived metrics of FA and MD. Furthermore, our results suggest that increases in the fixel-based metrics of FD and FDC may not necessarily be functionally beneficial, as we found negative correlations between percent change in these metrics in anterior WM and percent change in DSST performance. This dissertation suggests steps for future research as well, including gaining a better understanding of the mechanisms underlying exercise-induced structural integrity changes in aging, functional changes in the brain following aerobic exercise, and the relationship between WM metrics and cognitive outcomes in a region-specific manner.



## REFERENCES

- Akoka, S., Franconi, F., Seguin, F., & Le Pape, A. (1993). Radiofrequency map of an NMR coil by imaging. *Magnetic Resonance Imaging*, *11*(3), 437–441.  
[https://doi.org/10.1016/0730-725X\(93\)90078-R](https://doi.org/10.1016/0730-725X(93)90078-R)
- Aquino, D., Bizzi, A., Grisoli, M., Garavaglia, B., Bruzzone, M. G., Nardocci, N., Savoiardo, M., & Chiapparini, L. (2009). Age-related Iron Deposition in the Basal Ganglia: Quantitative Analysis in Healthy Subjects. *Radiology*, *252*(1), 165–172.  
<https://doi.org/10.1148/radiol.2522081399>
- Ashburner, J., & Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage*, *11*(6), 805–821. <https://doi.org/10.1006/nimg.2000.0582>
- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–851.  
<https://doi.org/10.1016/j.neuroimage.2005.02.018>
- Ashburner, J., & Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss–Newton optimisation. *NeuroImage*, *55*(3), 954–967.  
<https://doi.org/10.1016/j.neuroimage.2010.12.049>
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044.  
<https://doi.org/10.1016/j.neuroimage.2010.09.025>
- Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., & Gee, J. C. (2010). The optimal template effect in hippocampus studies of diseased populations. *NeuroImage*, *49*(3), 2457–2466.  
<https://doi.org/10.1016/j.neuroimage.2009.09.062>
- Basser, P. J., & Pierpaoli, C. (1996). Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI. *Journal of Magnetic Resonance, Series B*, *111*(3), 209–219. <https://doi.org/10.1006/jmrb.1996.0086>
- Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system – a technical review. *NMR in Biomedicine*, *15*(7–8), 435–455.  
<https://doi.org/10.1002/nbm.782>
- Benedetti, B., Charil, A., Rovaris, M., Judica, E., Valsasina, P., Sormani, M. P., & Filippi, M. (2006). Influence of aging on brain gray and white matter changes assessed by conventional, MT, and DT MRI. *Neurology*, *66*(4), 535–539.  
<https://doi.org/10.1212/01.wnl.0000198510.73363.c6>

- Bennett, I. J., & Madden, D. J. (2014). Disconnected aging: Cerebral white matter integrity and age-related differences in cognition. *Neuroscience*, *276*, 187–205. <https://doi.org/10.1016/j.neuroscience.2013.11.026>
- Blondell, S. J., Hammersley-Mather, R., & Veerman, J. L. (2014). Does physical activity prevent cognitive decline and dementia?: A systematic review and meta-analysis of longitudinal studies. *BMC Public Health*, *14*(1), 510. <https://doi.org/10.1186/1471-2458-14-510>
- Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., Estabrook, R., Bates, T. C., Mehta, P., von Oertzen, T., Gore, R. J., Hunter, M. D., Hackett, D. C., Karch, J., Brandmaier, A. M., Pritikin, J. N., Zahery, M., Kirkpatrick, R. M., Wang, Y., & Niesen, J. (2021). *OpenMx 2.19.6 User Guide*.
- Brandmaier, A. M., Wenger, E., Bodammer, N. C., Kühn, S., Raz, N., & Lindenberger, U. (2018). Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). *ELife*, *7*, e35718. <https://doi.org/10.7554/eLife.35718>
- Brickman, A. M., Zimmerman, M. E., Paul, R. H., Grieve, S. M., Tate, D. F., Cohen, R. A., Williams, L. M., Clark, C. R., & Gordon, E. (2006). Regional White Matter and Neuropsychological Functioning across the Adult Lifespan. *Biological Psychiatry*, *60*(5), 444–453. <https://doi.org/10.1016/j.biopsych.2006.01.011>
- Burdette, J. H., Laurienti, P. J., Espeland, M. A., Morgan, A., Telesford, Q., Vechlekar, C. D., Hayasaka, S., Jennings, J. M., Katula, J. A., Kraft, R. A., & Rejeski, W. J. (2010). Using network science to evaluate exercise-associated brain changes in older adults. *Frontiers in Aging Neuroscience*, *2*, 23. <https://doi.org/10.3389/fnagi.2010.00023>
- Burzynska, A. Z., Chaddock-Heyman, L., Voss, M. W., Wong, C. N., Gothe, N. P., Olson, E. A., Knecht, A., Lewis, A., Monti, J. M., Cooke, G. E., Wojcicki, T. R., Fanning, J., Chung, H. D., Awick, E., McAuley, E., & Kramer, A. F. (2014). Physical Activity and Cardiorespiratory Fitness Are Beneficial for White Matter in Low-Fit Older Adults. *PLoS ONE*, *9*(9), e107413. <https://doi.org/10.1371/journal.pone.0107413>
- Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., Lindenberger, U., Nyberg, L., Park, D. C., Reuter-Lorenz, P. A., Rugg, M. D., Steffener, J., & Rajah, M. N. (2018). Maintenance, reserve and compensation: The cognitive neuroscience of healthy ageing. *Nature Reviews Neuroscience*, *19*(11), 701–710. <https://doi.org/10.1038/s41583-018-0068-2>
- Canty, A., & Ripley, B. (2021). *boot: Bootstrap R (S-Plus) Functions* (1.3-28) [R].
- Chalmers, R. P., & Flora, D. B. (2015). faoutlier: An R Package for Detecting Influential Cases in Exploratory and Confirmatory Factor Analysis. *Applied Psychological Measurement*, *39*(7), 573–574. <https://doi.org/10.1177/0146621615597894>



- Chapman, S. B., Aslan, S., Spence, J. S., DeFina, L. F., Keebler, M. W., Didehbani, N., & Lu, H. (2013). Shorter term aerobic exercise improves brain, cognition, and cardiovascular fitness in aging. *Frontiers in Aging Neuroscience, 5*.  
<https://doi.org/10.3389/fnagi.2013.00075>
- Chapman, T. W., & Hill, R. A. (2020). Myelin plasticity in adulthood and aging. *Neuroscience Letters, 715*, 134645. <https://doi.org/10.1016/j.neulet.2019.134645>
- Choy, S. W., Bagarinao, E., Watanabe, H., Ho, E. T. W., Maesawa, S., Mori, D., Hara, K., Kawabata, K., Yoneyama, N., Ohdake, R., Imai, K., Masuda, M., Yokoi, T., Ogura, A., Taoka, T., Koyama, S., Tanabe, H. C., Katsuno, M., Wakabayashi, T., ... Sobue, G. (2020). Changes in white matter fiber density and morphology across the adult lifespan: A cross-sectional fixel-based analysis. *Human Brain Mapping, 41*(12), 3198–3211. <https://doi.org/10.1002/hbm.25008>
- Clark, C. M., Guadagni, V., Mazerolle, E. L., Hill, M., Hogan, D. B., Pike, G. B., & Poulin, M. J. (2019). Effect of aerobic exercise on white matter microstructure in the aging brain. *Behavioural Brain Research, 373*, 112042.  
<https://doi.org/10.1016/j.bbr.2019.112042>
- Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., Elavsky, S., Marquez, D. X., Hu, L., & Kramer, A. F. (2006). Aerobic Exercise Training Increases Brain Volume in Aging Humans. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 61*(11), 1166–1170.  
<https://doi.org/10.1093/gerona/61.11.1166>
- Colcombe, S. J., & Kramer, A. F. (2003). Fitness Effects on the Cognitive Function of Older Adults. *Psychological Science, 14*(2), 125–130.  
<https://doi.org/10.1111/1467-9280.t01-1-01430>
- Davies, R. R., Scahill, V. L., Graham, A., Williams, G. B., Graham, K. S., & Hodges, J. R. (2009). Development of an MRI rating scale for multiple brain regions: Comparison with volumetrics and with voxel-based morphometry. *Neuroradiology, 51*(8), 491–503. <https://doi.org/10.1007/s00234-009-0521-z>
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.
- Dennis, E. L., & Thompson, P. M. (2014). Functional brain connectivity using fMRI in aging and Alzheimer's disease. *Neuropsychology Review, 24*(1), 49–62.  
<https://doi.org/10.1007/s11065-014-9249-6>
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage, 31*(3), 968–980.  
<https://doi.org/10.1016/j.neuroimage.2006.01.021>

- Dhollander, T., Raffelt, D., & Connelly, A. (2016). A novel iterative approach to reap the benefits of multi-tissue CSD from just single-shell ( $b=0$ ) diffusion MRI data. *Proceedings of the 24th Annual Meeting of the International Society of Magnetic Resonance in Medicine*, 3010.
- Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R. S. J., Helms, G., & Weiskopf, N. (2011). Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *NeuroImage*, 55(4), 1423–1434. <https://doi.org/10.1016/j.neuroimage.2011.01.052>
- Erickson, K. I., Leckie, R. L., & Weinstein, A. M. (2014). Physical activity, fitness, and gray matter volume. *Neurobiology of Aging*, 35, S20–S28. <https://doi.org/10.1016/j.neurobiolaging.2014.03.034>
- Erickson, K. I., Raji, C. A., Lopez, O. L., Becker, J. T., Rosano, C., Newman, A. B., Gach, H. M., Thompson, P. M., Ho, A. J., & Kuller, L. H. (2010). Physical activity predicts gray matter volume in late adulthood: The Cardiovascular Health Study. *Neurology*, 75(16), 1415–1422. <https://doi.org/10.1212/WNL.0b013e3181f88359>
- Erickson, K. I., Voss, M. W., Prakash, R. S., Basak, C., Szabo, A., Chaddock, L., Kim, J. S., Heo, S., Alves, H., White, S. M., Wojcicki, T. R., Mailey, E., Vieira, V. J., Martin, S. A., Pence, B. D., Woods, J. A., McAuley, E., & Kramer, A. F. (2011). Exercise training increases size of hippocampus and improves memory. *Proceedings of the National Academy of Sciences*, 108(7), 3017–3022. <https://doi.org/10.1073/pnas.1015950108>
- Ferreira, L. K., & Busatto, G. F. (2013). Resting-state functional connectivity in normal brain aging. *Neuroscience and Biobehavioral Reviews*, 37(3), 384–400. <https://doi.org/10.1016/j.neubiorev.2013.01.017>
- Fling, B. W., Chapekis, M., Reuter-Lorenz, P. A., Anguera, J., Bo, J., Langan, J., Welsh, R. C., & Seidler, R. D. (2011). Age differences in callosal contributions to cognitive processes. *Neuropsychologia*, 49(9), 2564–2569. <https://doi.org/10.1016/j.neuropsychologia.2011.05.004>
- Ge, Y., Grossman, R. I., Babb, J. S., Rabin, M. L., Mannon, L. J., & Kolson, D. L. (2002). Age-related total gray matter and white matter changes in normal adult brain. Part II: Quantitative magnetization transfer ratio histogram analysis. *AJNR. American Journal of Neuroradiology*, 23(8), 1334–1341.
- Genc, S., Tax, C. M. W., Raven, E. P., Chamberland, M., Parker, G. D., & Jones, D. K. (2020). Impact of  $b$ -value on estimates of apparent fibre density. *Human Brain Mapping*, 41(10), 2583–2595. <https://doi.org/10.1002/hbm.24964>

- Giuliani, N. R., Calhoun, V. D., Pearlson, G. D., Francis, A., & Buchanan, R. W. (2005). Voxel-based morphometry versus region of interest: A comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophrenia Research*, *74*(2–3), 135–147. <https://doi.org/10.1016/j.schres.2004.08.019>
- Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Friston, K. J., & Frackowiak, R. S. J. (2001). A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *NeuroImage*, *14*(1), 21–36. <https://doi.org/10.1006/nimg.2001.0786>
- Good, C. D., Scahill, R. I., Fox, N. C., Ashburner, J., Friston, K. J., Chan, D., Crum, W. R., Rossor, M. N., & Frackowiak, R. S. J. (2002). Automatic Differentiation of Anatomical Patterns in the Human Brain: Validation with Studies of Degenerative Dementias. *NeuroImage*, *17*(1), 29–46. <https://doi.org/10.1006/nimg.2002.1202>
- Gorbach, T., Pudas, S., Lundquist, A., Orädd, G., Josefsson, M., Salami, A., de Luna, X., & Nyberg, L. (2017). Longitudinal association between hippocampus atrophy and episodic-memory decline. *Neurobiology of Aging*, *51*, 167–176. <https://doi.org/10.1016/j.neurobiolaging.2016.12.002>
- Guevara, M., Guevara, P., Román, C., & Mangin, J.-F. (2020). Superficial white matter: A review on the dMRI analysis methods and applications. *NeuroImage*, *212*, 116673. <https://doi.org/10.1016/j.neuroimage.2020.116673>
- Hedges, E., Zinser, J., Dimitrov, M., Antoniadis, M., Porffy, L., Pisani, S., Dickson, H., McGuire, P., & Kempton, M. J. (2020). M154. INTRA- AND INTER-SCANNER RELIABILITY OF GRAY MATTER VOLUME AND CORTICAL THICKNESS ESTIMATES: IMPLICATIONS FOR MULTICENTRE IMAGING STUDIES IN PSYCHOSIS. *Schizophrenia Bulletin*, *46*(Supplement\_1), S194–S194. <https://doi.org/10.1093/schbul/sbaa030.466>
- Helms, G., Dathe, H., & Dechent, P. (2008). Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation: Rational Approximation of the FLASH Signal. *Magnetic Resonance in Medicine*, *59*(3), 667–672. <https://doi.org/10.1002/mrm.21542>
- Helms, G., Dathe, H., Kallenberg, K., & Dechent, P. (2008). High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and  $T_1$  relaxation obtained from 3D FLASH MRI: Saturation and Relaxation in MT FLASH. *Magnetic Resonance in Medicine*, *60*(6), 1396–1407. <https://doi.org/10.1002/mrm.21732>
- Ho, A. J., Raji, C. A., Becker, J. T., Lopez, O. L., Kuller, L. H., Hua, X., Dinov, I. D., Stein, J. L., Rosano, C., Toga, A. W., & Thompson, P. M. (2011). The effects of physical activity, education, and body mass index on the aging brain. *Human Brain Mapping*, *32*(9), 1371–1382. <https://doi.org/10.1002/hbm.21113>

- Hoyer, W. J., Stawski, R. S., Wasylshyn, C., & Verhaeghen, P. (2004). Adult Age and Digit Symbol Substitution Performance: A Meta-Analysis. *Psychology and Aging, 19*(1), 211–214. <https://doi.org/10.1037/0882-7974.19.1.211>
- Hunter, M. D. (2018). State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(2), 307–324. <https://doi.org/10.1080/10705511.2017.1369354>
- Jaeger, J. (2018). Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in Neuropsychological Testing. *Journal of Clinical Psychopharmacology, 38*(5), 513–519. <https://doi.org/10.1097/JCP.0000000000000941>
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage, 62*(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- Jernigan, T. L., Archibald, S. L., Fennema-Notestine, C., Gamst, A. C., Stout, J. C., Bonner, J., & Hesselink, J. R. (2001). Effects of age on tissues and regions of the cerebrum and cerebellum. *Neurobiology of Aging, 22*(4), 581–594. [https://doi.org/10.1016/S0197-4580\(01\)00217-2](https://doi.org/10.1016/S0197-4580(01)00217-2)
- Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., & Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging: Prevalence of Multifiber Voxels in WM. *Human Brain Mapping, 34*(11), 2747–2766. <https://doi.org/10.1002/hbm.22099>
- Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., & Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage, 103*, 411–426. <https://doi.org/10.1016/j.neuroimage.2014.07.061>
- Johansson, J., Wählin, A., Lundquist, A., Brandmaier, A. M., Lindenberger, U., & Nyberg, L. (2022). Model of brain maintenance reveals specific change-change association between medial-temporal lobe integrity and episodic memory. *Aging Brain, 2*, 100027. <https://doi.org/10.1016/j.nbas.2021.100027>
- Kelley, S., Plass, J., Bender, A. R., & Polk, T. A. (2021). Age-Related Differences in White Matter: Understanding Tensor-Based Results Using Fixel-Based Analysis. *Cerebral Cortex, 31*(8), 3881–3898. <https://doi.org/10.1093/cercor/bhab056>
- Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., & Brennan, S. (2014). The impact of exercise on the cognitive functioning of healthy older adults: A systematic review and meta-analysis. *Ageing Research Reviews, 16*, 12–31. <https://doi.org/10.1016/j.arr.2014.05.002>

- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, *33*, 99–117. <https://doi.org/10.1016/j.dcn.2017.11.007>
- King, A. C. (1991). Group- vs Home-Based Exercise Training in Healthy Older Men and Women: A Community-Based Clinical Trial. *JAMA*, *266*(11), 1535. <https://doi.org/10.1001/jama.1991.03470110081037>
- Kirilina, E., Helbling, S., Morawski, M., Pine, K., Reimann, K., Jankuhn, S., Dinse, J., Deistung, A., Reichenbach, J. R., Trampel, R., Geyer, S., Müller, L., Jakubowski, N., Arendt, T., Bazin, P.-L., & Weiskopf, N. (2020). Superficial white matter imaging: Contrast mechanisms and whole-brain in vivo mapping. *Science Advances*, *6*(41), eaaz9281. <https://doi.org/10.1126/sciadv.aaz9281>
- Kleemeyer, M. M., Kühn, S., Prindle, J., Bodammer, N. C., Brechtel, L., Garthe, A., Kempermann, G., Schaefer, S., & Lindenberger, U. (2016). Changes in fitness are associated with changes in hippocampal microstructure and hippocampal volume among older adults. *NeuroImage*, *131*, 155–161. <https://doi.org/10.1016/j.neuroimage.2015.11.026>
- Kleinloog, J. P. D., Mensink, R. P., Ivanov, D., Adam, J. J., Uludağ, K., & Joris, P. J. (2019). Aerobic Exercise Training Improves Cerebral Blood Flow and Executive Function: A Randomized, Controlled Cross-Over Trial in Sedentary Older Men. *Frontiers in Aging Neuroscience*, *11*, 333. <https://doi.org/10.3389/fnagi.2019.00333>
- Köhncke, Y., Düzel, S., Sander, M. C., Lindenberger, U., Kühn, S., & Brandmaier, A. M. (2021). Hippocampal and Parahippocampal Gray Matter Structural Integrity Assessed by Multimodal Imaging Is Associated with Episodic Memory in Old Age. *Cerebral Cortex*, *31*(3), 1464–1477. <https://doi.org/10.1093/cercor/bhaa287>
- Kühn, S., Düzel, S., Eibich, P., Krekel, C., Wüstemann, H., Kolbe, J., Martensson, J., Goebel, J., Gallinat, J., Wagner, G. G., & Lindenberger, U. (2017). In search of features that constitute an “enriched environment” in humans: Associations between geographical properties and brain structure. *Scientific Reports*, *7*(1), 11920. <https://doi.org/10.1038/s41598-017-12046-7>
- Lehrl, S., Merz, J., Burkard, G., & Fischer, B. (1991). *Manual zum MWT-A*. Perimed.
- Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., & Mattay, V. S. (2012). Normal age-related brain morphometric changes: Nonuniformity across cortical thickness, surface area and gray matter volume? *Neurobiology of Aging*, *33*(3), 617.e1–617.e9. <https://doi.org/10.1016/j.neurobiolaging.2010.07.013>

- Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science*, *346*(6209), 572–578. <https://doi.org/10.1126/science.1254403>
- Lövdén, M., Schaefer, S., Noack, H., Bodammer, N. C., Kühn, S., Heinze, H.-J., Düzel, E., Bäckman, L., & Lindenberger, U. (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiology of Aging*, *33*(3), 620.e9–620.e22. <https://doi.org/10.1016/j.neurobiolaging.2011.02.013>
- Luque Laguna, P. A., Combes, A. J. E., Streffer, J., Einstein, S., Timmers, M., Williams, S. C. R., & Dell'Acqua, F. (2020). Reproducibility, reliability and variability of FA and MD in the older healthy population: A test-retest multiparametric analysis. *NeuroImage: Clinical*, *26*, 102168. <https://doi.org/10.1016/j.nicl.2020.102168>
- Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., & Weiskopf, N. (2010). Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T: Optimized RF Transmit Field Mapping at 3T. *Magnetic Resonance in Medicine*, *64*(1), 229–238. <https://doi.org/10.1002/mrm.22421>
- Maass, A., Düzel, S., Goerke, M., Becke, A., Sobieray, U., Neumann, K., Lövdén, M., Lindenberger, U., Bäckman, L., Braun-Dullaeus, R., Ahrens, D., Heinze, H.-J., Müller, N. G., & Düzel, E. (2015). Vascular hippocampal plasticity after aerobic exercise in older adults. *Molecular Psychiatry*, *20*(5), 585–593. <https://doi.org/10.1038/mp.2014.114>
- Marks, B., Katz, L., Styner, M., & Smith, J. (2011). Aerobic fitness and obesity: Relationship to cerebral white matter integrity in the brain of active and sedentary older adults. *British Journal of Sports Medicine*, *45*(15), 1208–1215. <https://doi.org/10.1136/bjism.2009.068114>
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). *OpenMx 2.0: Extended Structural Equation and Statistical Modeling*. <http://link.springer.com/10.1007/s11336-014-9435-8>
- Nyberg, L., & Lindenberger, U. (2020). Brain maintenance and cognition in old age. In D. Poeppel, G. R. Mangun, & M. S. Gazzaniga (Eds.), *The cognitive neurosciences* (Sixth edition, pp. 81–89). The MIT Press.
- Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, *16*(5), 292–305. <https://doi.org/10.1016/j.tics.2012.04.005>
- Nyberg, L., & Pudas, S. (2019). Successful Memory Aging. *Annual Review of Psychology*, *70*(1), 219–243. <https://doi.org/10.1146/annurev-psych-010418-103052>

- Ogoh, S., & Ainslie, P. N. (2009). Cerebral blood flow during exercise: Mechanisms of regulation. *Journal of Applied Physiology*, *107*(5), 1370–1380. <https://doi.org/10.1152/jappphysiol.00573.2009>
- O’Sullivan, M., Jones, D. K., Summers, P. E., Morris, R. G., Williams, S. C. R., & Markus, H. S. (2001). Evidence for cortical “disconnection” as a mechanism of age-related cognitive decline. *Neurology*, *57*(4), 632–638. <https://doi.org/10.1212/WNL.57.4.632>
- Pierpaoli, C., & Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magnetic Resonance in Medicine*, *36*(6), 893–906. <https://doi.org/10.1002/mrm.1910360612>
- Polk, S. E., Kleemeyer, M. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Düzel, S., & Wenger, E. (2022). *Aerobic exercise is associated with region-specific changes in volumetric, tensor-based, and fixel-based measures of white matter integrity in healthy older adults* [Manuscript under revision].
- Polk, S. E., Kleemeyer, M. M., Köhncke, Y., Brandmaier, A. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Wenger, E., & Düzel, S. (2022). Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2022.852737>
- Preibisch, C., & Deichmann, R. (2009). Influence of RF spoiling on the stability and accuracy of  $T_1$  mapping based on spoiled FLASH with varying flip angles: Influence of RF Spoiling on  $T_1$  Mapping. *Magnetic Resonance in Medicine*, *61*(1), 125–135. <https://doi.org/10.1002/mrm.21776>
- Pritikin, J. N., Hunter, M. D., & Boker, S. M. (2015). Modular Open-Source Software for Item Factor Analysis. *Educational and Psychological Measurement*, *75*(3), 458–474. <https://doi.org/10.1177/0013164414554615>
- Querido, J. S., & Sheel, A. W. (2007). Regulation of Cerebral Blood Flow During Exercise: *Sports Medicine*, *37*(9), 765–782. <https://doi.org/10.2165/00007256-200737090-00002>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raffelt, D. A., Smith, R. E., Ridgway, G. R., Tournier, J.-D., Vaughan, D. N., Rose, S., Henderson, R., & Connelly, A. (2015). Connectivity-based fixel enhancement: Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing fibres. *NeuroImage*, *117*, 40–55. <https://doi.org/10.1016/j.neuroimage.2015.05.039>

- Raffelt, D. A., Tournier, J.-D., Rose, S., Ridgway, G. R., Henderson, R., Crozier, S., Salvado, O., & Connelly, A. (2012). Apparent Fibre Density: A novel measure for the analysis of diffusion-weighted magnetic resonance images. *NeuroImage*, *59*(4), 3976–3994. <https://doi.org/10.1016/j.neuroimage.2011.10.045>
- Raffelt, D. A., Tournier, J.-D., Smith, R. E., Vaughan, D. N., Jackson, G., Ridgway, G. R., & Connelly, A. (2017). Investigating white matter fibre density and morphology using fixel-based analysis. *NeuroImage*, *144*, 58–73. <https://doi.org/10.1016/j.neuroimage.2016.09.029>
- Raghavan, S., Przybelski, S. A., Reid, R. I., Graff-Radford, J., Lesnick, T. G., Zuk, S. M., Knopman, D. S., Machulda, M. M., Mielke, M. M., Petersen, R. C., Jack, C. R., & Vemuri, P. (2020). Reduced fractional anisotropy of the genu of the corpus callosum as a cerebrovascular disease marker and predictor of longitudinal cognition in MCI. *Neurobiology of Aging*, *96*, 176–183. <https://doi.org/10.1016/j.neurobiolaging.2020.09.005>
- Ramanoël, S., Hoyau, E., Kauffmann, L., Renard, F., Pichat, C., Boudiaf, N., Krainik, A., Jaillard, A., & Baciú, M. (2018). Gray Matter Volume and Cognitive Performance During Normal Aging. A Voxel-Based Morphometry Study. *Frontiers in Aging Neuroscience*, *10*, 235. <https://doi.org/10.3389/fnagi.2018.00235>
- Raz, N., Gunning, F. M., Head, D., Dupuis, J. H., McQuain, J., Briggs, S. D., Loken, W. J., Thornton, A. E., & Acker, J. D. (1997). Selective aging of the human cerebral cortex observed in vivo: Differential vulnerability of the prefrontal gray matter. *Cerebral Cortex*, *7*(3), 268–282. <https://doi.org/10.1093/cercor/7.3.268>
- Raz, N., Gunning-Dixon, F., Head, D., Rodrigue, K. M., Williamson, A., & Acker, J. D. (2004). Aging, sexual dimorphism, and hemispheric asymmetry of the cerebral cortex: Replicability of regional differences in volume. *Neurobiology of Aging*, *25*(3), 377–396. [https://doi.org/10.1016/S0197-4580\(03\)00118-0](https://doi.org/10.1016/S0197-4580(03)00118-0)
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., Dahle, C., Gerstorff, D., & Acker, J. D. (2005). Regional Brain Changes in Aging Healthy Adults: General Trends, Individual Differences and Modifiers. *Cerebral Cortex*, *15*(11), 1676–1689. <https://doi.org/10.1093/cercor/bhi044>
- Rojas-Vite, G., Coronado-Leija, R., Narvaez-Delgado, O., Ramírez-Manzanares, A., Marroquín, J. L., Noguez-Imm, R., Aranda, M. L., Scherrer, B., Larriva-Sahd, J., & Concha, L. (2019). Histological validation of per-bundle water diffusion metrics within a region of fiber crossing following axonal degeneration. *NeuroImage*, *201*, 116013. <https://doi.org/10.1016/j.neuroimage.2019.116013>
- Ropele, S., Schmidt, R., Enzinger, C., Windisch, M., Martinez, N. P., & Fazekas, F. (2012). Longitudinal Magnetization Transfer Imaging in Mild to Severe Alzheimer Disease. *American Journal of Neuroradiology*, *33*(3), 570–575. <https://doi.org/10.3174/ajnr.A2812>



- Rosseel, Y. (2012). **lavaan**: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>
- Rowley, C. D., Campbell, J. S. W., Wu, Z., Leppert, I. R., Rudko, D. A., Pike, G. B., & Tardif, C. L. (2021). A model-based framework for correcting inhomogeneity effects in magnetization transfer saturation and inhomogeneous magnetization transfer saturation maps. *Magnetic Resonance in Medicine*, 86(4), 2192–2207. <https://doi.org/10.1002/mrm.28831>
- RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC. <http://www.rstudio.com/>
- Ruscheweyh, R., Willemer, C., Krüger, K., Duning, T., Warnecke, T., Sommer, J., Völker, K., Ho, H. V., Mooren, F., Knecht, S., & Flöel, A. (2011). Physical activity and memory functions: An interventional study. *Neurobiology of Aging*, 32(7), 1304–1319. <https://doi.org/10.1016/j.neurobiolaging.2009.08.001>
- Salvetti, X. M., Oliveira, J. A., Servantes, D. M., & Vincenzo de Paola, A. A. (2008). How much do the benefits cost? Effects of a home-based training programme on cardiovascular fitness, quality of life, programme cost and adherence for patients with coronary disease. *Clinical Rehabilitation*, 22(10–11), 987–996. <https://doi.org/10.1177/0269215508093331>
- Sexton, C. E., Betts, J. F., Demnitz, N., Dawes, H., Ebmeier, K. P., & Johansen-Berg, H. (2016). A systematic review of MRI studies examining the relationship between physical fitness and activity and the white matter of the ageing brain. *NeuroImage*, 131, 81–90. <https://doi.org/10.1016/j.neuroimage.2015.09.071>
- Sexton, C. E., Walhovd, K. B., Storsve, A. B., Tamnes, C. K., Westlye, L. T., Johansen-Berg, H., & Fjell, A. M. (2014). Accelerated Changes in White Matter Microstructure during Aging: A Longitudinal Diffusion Tensor Imaging Study. *Journal of Neuroscience*, 34(46), 15425–15436. <https://doi.org/10.1523/JNEUROSCI.0203-14.2014>
- Shvartz, E., & Reibold, R. C. (1990). Aerobic fitness norms for males and females aged 6 to 75 years: A review. *Aviation, Space, and Environmental Medicine*, 61(1), 3–11.
- Smith, J. C., Nielson, K. A., Woodard, J. L., Seidenberg, M., Durgerian, S., Antuono, P., Butts, A. M., Hantke, N. C., Lancaster, M. A., & Rao, S. M. (2011). Interactive effects of physical activity and APOE-ε4 on BOLD semantic memory activation in healthy elders. *NeuroImage*, 54(1), 635–644. <https://doi.org/10.1016/j.neuroimage.2010.07.070>

- Smith, P. J., Blumenthal, J. A., Hoffman, B. M., Cooper, H., Strauman, T. A., Welsh-Bohmer, K., Browndyke, J. N., & Sherwood, A. (2010). Aerobic Exercise and Neurocognitive Performance: A Meta-Analytic Review of Randomized Controlled Trials. *Psychosomatic Medicine*, *72*(3), 239–252. <https://doi.org/10.1097/PSY.0b013e3181d14633>
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. J. (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. *NeuroImage*, *31*(4), 1487–1505. <https://doi.org/10.1016/j.neuroimage.2006.02.024>
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., & Mohammadi, S. (2019). HMRI – A toolbox for quantitative MRI in neuroscience and clinical research. *NeuroImage*, *194*, 191–210. <https://doi.org/10.1016/j.neuroimage.2019.01.029>
- Tahedl, M. (2018). *B.A.T.M.A.N.: Basic and Advanced Tractography with MRtrix for All Neurophiles*. <https://doi.org/10.17605/OSF.IO/FKYHT>
- Taki, Y., Kinomura, S., Sato, K., Goto, R., Kawashima, R., & Fukuda, H. (2011). A longitudinal study of gray matter volume decline with age and modifying factors. *Neurobiology of Aging*, *32*(5), 907–915. <https://doi.org/10.1016/j.neurobiolaging.2009.05.003>
- Tian, Q., Erickson, K. I., Simonsick, E. M., Aizenstein, H. J., Glynn, N. W., Boudreau, R. M., Newman, A. B., Kritchevsky, S. B., Yaffe, K., Harris, T. B., & Rosano, C. (2014). Physical Activity Predicts Microstructural Integrity in Memory-Related Networks in Very Old Adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *69*(10), 1284–1290. <https://doi.org/10.1093/gerona/glt287>
- Tian, Q., Simonsick, E. M., Erickson, K. I., Aizenstein, H. J., Glynn, N. W., Boudreau, R. M., Newman, A. B., Kritchevsky, S. B., Yaffe, K., Harris, T., & Rosano, C. (2014). Cardiorespiratory fitness and brain diffusion tensor imaging in adults over 80 years of age. *Brain Research*, *1588*, 63–72. <https://doi.org/10.1016/j.brainres.2014.09.003>

- Tisserand, D. J., van Boxtel, M. P. J., Pruessner, J. C., Hofman, P., Evans, A. C., & Jolles, J. (2004). A Voxel-based Morphometric Study to Determine Individual Differences in Gray Matter Density Associated with Age and Cognitive Change Over Time. *Cerebral Cortex*, *14*(9), 966–973. <https://doi.org/10.1093/cercor/bhh057>
- Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, *202*, 116137. <https://doi.org/10.1016/j.neuroimage.2019.116137>
- United Nations, Department of Economic and Social Affairs, Population Division. (2019). *World Population Ageing 2019: Highlights (ST/ESA/SER.A/430)*.
- Van Petten, C., Plante, E., Davidson, P. S. R., Kuo, T. Y., Bajuscak, L., & Glisky, E. L. (2004). Memory and executive function in older adults: Relationships with temporal and prefrontal gray matter volumes and white matter hyperintensities. *Neuropsychologia*, *42*(10), 1313–1335. <https://doi.org/10.1016/j.neuropsychologia.2004.02.009>
- van Uffelen, J. G. Z., Chin A Paw, M. J. M., Hopman-Rock, M., & Mechelen, W. van. (2008). The Effects of Exercise on Cognition in Older Adults With and Without Cognitive Decline: A Systematic Review. *Clinical Journal of Sport Medicine*, *18*(6), 486–500. <https://doi.org/10.1097/JSM.0b013e3181845f0b>
- von Oertzen, T. (2010). Power equivalence in structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, *63*(2), 257–272. <https://doi.org/10.1348/000711009X441021>
- von Oertzen, T., Brandmaier, A. M., & Tsang, S. (2015). Structural Equation Modeling With *Onyx*. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 148–161. <https://doi.org/10.1080/10705511.2014.935842>
- Voss, M. W., Heo, S., Prakash, R. S., Erickson, K. I., Alves, H., Chaddock, L., Szabo, A. N., Mailey, E. L., Wójcicki, T. R., White, S. M., Gothe, N., McAuley, E., Sutton, B. P., & Kramer, A. F. (2013). The influence of aerobic fitness on cerebral white matter integrity and cognitive function in older adults: Results of a one-year exercise intervention. *Human Brain Mapping*, *34*(11), 2972–2985. <https://doi.org/10.1002/hbm.22119>
- Voss, M. W., Prakash, R. S., Erickson, K. I., Basak, C., Chaddock, L., Kim, J. S., Alves, H., Heo, S., Szabo, A. N., White, S. M., Wójcicki, T. R., Mailey, E. L., Gothe, N., Olson, E. A., McAuley, E., & Kramer, A. F. (2010). Plasticity of brain networks in a randomized intervention trial of exercise training in older adults. *Frontiers in Aging Neuroscience*, *2*(AUG), 1–17. <https://doi.org/10.3389/fnagi.2010.00032>

- Wansbeek, T., & Meijer, E. (2003). Measurement Error and Latent Variables. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* (pp. 162–179). Blackwell Publishing Ltd. <https://doi.org/10.1002/9780470996249.ch9>
- Wechsler, D. (1981). The psychometric tradition: Developing the Wechsler adult intelligence scale. *Contemporary Educational Psychology*, 6(2), 82–85. [https://doi.org/10.1016/0361-476X\(81\)90035-7](https://doi.org/10.1016/0361-476X(81)90035-7)
- Weiskopf, N., Callaghan, M. F., Josephs, O., Lutti, A., & Mohammadi, S. (2014). Estimating the apparent transverse relaxation time (R2\*) from images with different contrasts (ESTATICS) reduces motion artifacts. *Frontiers in Neuroscience*, 8. <https://doi.org/10.3389/fnins.2014.00278>
- Weiskopf, N., Edwards, L. J., Helms, G., Mohammadi, S., & Kirilina, E. (2021). Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nature Reviews Physics*, 3(8), 570–588. <https://doi.org/10.1038/s42254-021-00326-1>
- Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., & Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *NeuroImage*, 54(3), 2116–2124. <https://doi.org/10.1016/j.neuroimage.2010.10.023>
- Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., Ooi, C., Bullmore, E. T., & Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD\*, MT, and R2\* at 3T: A multi-center validation. *Frontiers in Neuroscience*, 7. <https://doi.org/10.3389/fnins.2013.00095>
- Wenger, E., Düzel, S., Polk, S. E., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., & Lindenberger, U. (2022). *Vamos en bici: Study protocol of an investigation of cognitive and neural changes following language training, physical exercise training, or a combination of both* [Preprint]. <https://doi.org/10.1101/2022.01.30.478181>
- Wenger, E., Polk, S. E., Kleemeyer, M. M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for MT and PD but attenuated for R1 and R2\* in healthy young adults. *Human Brain Mapping*. <https://doi.org/10.1002/hbm.25870>
- Whitwell, J. L., Josephs, K. A., Rossor, M. N., Stevens, J. M., Revesz, T., Holton, J. L., Al-Sarraj, S., Godbolt, A. K., Fox, N. C., & Warren, J. D. (2005). Magnetic Resonance Imaging Signatures of Tissue Pathology in Frontotemporal Dementia. *Archives of Neurology*, 62(9), 1402. <https://doi.org/10.1001/archneur.62.9.1402>
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45(1), S173–S186. <https://doi.org/10.1016/j.neuroimage.2008.10.055>

- Young, J., Angevaren, M., Rusted, J., & Tabet, N. (2015). Aerobic exercise to improve cognitive function in older people without known cognitive impairment. *Cochrane Database of Systematic Reviews*. <https://doi.org/10.1002/14651858.CD005381.pub4>
- Zimmerman, M. E., Brickman, A. M., Paul, R. H., Grieve, S. M., Tate, D. F., Gunstad, J., Cohen, R. A., Aloia, M. S., Williams, L. M., Clark, C. R., Whitford, T. J., & Gordon, E. (2006). The Relationship Between Frontal Gray Matter Volume and Cognition Varies Across the Healthy Adult Lifespan. *The American Journal of Geriatric Psychiatry*, *14*(10), 823–833. <https://doi.org/10.1097/01.JGP.0000238502.40963.ac>



*Appendix A.*  
**INDIVIDUAL WORKS**







Wenger, E., **Polk, S. E.**, Kleemeyer, M. M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for MT and PD but attenuated for R1 and R2\* in healthy young adults. *Human Brain Mapping*. <https://doi.org/10.1002/hbm.25870>





## RESEARCH ARTICLE

# Reliability of quantitative multiparameter maps is high for magnetization transfer and proton density but attenuated for $R_1$ and $R_2^*$ in healthy young adults

Elisabeth Wenger<sup>1</sup>  | Sarah E. Polk<sup>1</sup>  | Maike M. Kleemeyer<sup>1</sup>  |  
 Nikolaus Weiskopf<sup>2,3,4</sup>  | Nils C. Bodammer<sup>1</sup> | Ulman Lindenberger<sup>1,5</sup>  |  
 Andreas M. Brandmaier<sup>1,5,6</sup> 

<sup>1</sup>Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

<sup>2</sup>Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, University College London, London, UK

<sup>3</sup>Department of Neurophysics, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

<sup>4</sup>Felix Bloch Institute for Solid State Physics, Faculty of Physics and Earth Sciences, Leipzig University, Leipzig, Germany

<sup>5</sup>Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany

<sup>6</sup>Department of Psychology, MSB Medical School Berlin, Berlin, Germany

## Correspondence

Elisabeth Wenger, Center for Lifespan Psychology, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany.

Email: [wenger@mpib-berlin.mpg.de](mailto:wenger@mpib-berlin.mpg.de)

## Funding information

Wellcome Centre for Human Neuroimaging, Grant/Award Number: 203147/Z/16/Z; European Union's Horizon 2020, Grant/Award Number: 681094; European Research Council, Grant/Award Number: FP7/2007-2013; Bundesministerium für Bildung und Forschung, Grant/Award Numbers: 01EW1711A & B, 01GQ1421B; Max Planck Institute for Human Development; Max Planck Society

## Abstract

We investigate the reliability of individual differences of four quantities measured by magnetic resonance imaging-based multiparameter mapping (MPM): magnetization transfer saturation (MT), proton density (PD), longitudinal relaxation rate ( $R_1$ ), and effective transverse relaxation rate ( $R_2^*$ ). Four MPM datasets, two on each of two consecutive days, were acquired in healthy young adults. On Day 1, no repositioning occurred and on Day 2, participants were repositioned between MPM datasets. Using intraclass correlation effect decomposition (ICED), we assessed the contributions of session-specific, day-specific, and residual sources of measurement error. For whole-brain gray and white matter, all four MPM parameters showed high reproducibility and high reliability, as indexed by the coefficient of variation (CoV) and the intraclass correlation (ICC). However, MT, PD,  $R_1$ , and  $R_2^*$  differed markedly in the extent to which reliability varied across brain regions. MT and PD showed high reliability in almost all regions. In contrast,  $R_1$  and  $R_2^*$  showed low reliability in some regions outside the basal ganglia, such that the sum of the measurement error estimates in our structural equation model was higher than estimates of between-person differences. In addition, in this sample of healthy young adults, the four MPM parameters showed very little variability over four measurements but differed in how well they could assess between-person differences. We conclude that  $R_1$  and  $R_2^*$  might carry only limited person-specific information in some regions of the brain in healthy young adults, and, by implication, might be of restricted utility for studying associations to between-person differences in behavior in those regions.

## KEYWORDS

intraclass correlation effect decomposition, MRI, multiparameter mapping, reliability

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Human Brain Mapping* published by Wiley Periodicals LLC.

## 1 | INTRODUCTION

Research on human development seeks to delineate the variable and invariant properties of age-graded changes in the organization of brain-behavior-environment systems (Lindenberger et al., 2006). Magnetic resonance imaging (MRI) has become an indispensable tool for the noninvasive assessment of brain anatomy and microstructure and will continue to contribute knowledge on how brain structure changes in response to new environmental challenges or aging.

Quantitative MRI can help us to characterize the brain's micro-anatomy by using the magnetophysical properties of water molecules in brain tissue that govern MRI contrasts, which are then in turn used as surrogate parameters to describe histological properties (Tofts, 2003; Weiskopf et al., 2021). Recently, a comprehensive quantitative multiparameter mapping (MPM) approach was developed, which provides high-resolution maps of the longitudinal relaxation rate ( $R_1 = 1/T_1$ ), proton density (PD), magnetization transfer saturation (MT), and effective transverse relaxation rate ( $R_2^* = 1/T_2^*$ ) (Helms et al., 2009; Helms, Dathe, & Dechent, 2008; Weiskopf et al., 2011). These multiparameter maps are related to microstructural properties of myelin, iron deposits, and water, among other things (Draganski et al., 2011), even though it is not a simple one-to-one mapping and the exact relation to underlying physiological processes at the cellular and molecular level is still to be resolved (Weiskopf et al., 2021).

Central questions in lifespan psychology often pertain to the range and direction of within-person change and variability – be it longitudinal change observed over years and decades (Raz & Rodrigue, 2006), intervention-induced change over weeks and months (May, 2011), or fluctuations that occur from day to day and from moment to moment (Schmiedek et al., 2010). Random measurement error and systematic drifts can compromise the reliable measurement of change (Karch et al., 2019). Given that the expected effect sizes we typically seek to detect with structural MR are often no larger than 2%–3% of the quantity under investigation, be it gray matter volume, mean diffusivity, or other structural brain measures, measurement error and drift can easily jeopardize the reliable assessment of within-person changes and between-person differences. If measurement artifacts are of similar magnitude as effects of interest, then reliability is low, and effects of interest cannot be detected. Reliability is a pivotal issue in longitudinal studies, but it also matters for cross-sectional studies, when researchers either are interested in stable between-person differences or when they use time- or age-related differences between people as a proxy for change. Thus, in both cross-sectional and longitudinal designs, the stability of MR measures cannot simply be assumed but must instead be tested explicitly (Noble et al., 2020).

Different scientific communities such as physics and psychometrics can rely on two fundamentally different conceptions of reliability and error: physics widely uses the coefficient of variation (CoV) and less frequently the intraclass correlation coefficient (ICC), whereas the opposite is the case for psychometrics. Each of them is equally important but notably provide answers to very different questions. Physicists typically inquire how reliably a given measurement can detect a

given quantity. Therefore, it is common and well-justified to use CoV as the main measure to assess repeatability, as was done, for example, in a previous multicenter study of MPM (Weiskopf et al., 2013) or another study testing within-site and between-site reproducibility (Leutritz et al., 2020) of MPM. The CoV is a standardized measure of dispersion and is often expressed as a percentage. It is widely used in analytical chemistry, engineering, and physics to express precision of a measurement and repeatability on well-defined objects of measurement. However, Brandmaier, Wenger, et al. (2018) showed that CoV does not distinguish between error variance and true construct-related variance, that is, between-person differences in the construct of interest, and may therefore not be particularly informative for correlational studies interested in assessing and explaining between-person differences. Instead, cognitive neuroscience commonly relies on a different conception of reliability, which refers to the precision of assessing between-person differences. This is typically expressed as a ratio index, the ICC, which relates the variance within persons (or groups of persons) to the total variance, and therefore represents the strength of association between any pair of measurements made on the same object (Bartko, 1966). It is important to keep in mind that ICC will increase when within-subject measurements become more similar or when the true scores of participants become more distinct from one another. By definition, ICC values must thus be interpreted contingent upon the characteristics of a given population.

To assess the adequacy of MPM parameters for correlational studies of human neuroscience, we investigated the reliability of MPM parameters within participants across four different measurement occasions in relation to between-person differences, using ICC. Specifically, we made use of intraclass correlation effect decomposition (ICED), which has been recently introduced by Brandmaier, Wenger, et al. (2018). ICED estimates overall reliability while attributing the overall error variance to different sources by making use of the design features of a given study. We acquired data from 15 volunteers, who each were assessed four times. On Day 1, participants were measured twice back-to-back, without repositioning between MPM datasets. On Day 2, participants were also measured twice, but this time with a break in between measurements, which afforded repositioning of the participants' head. With this study design, we are able to tease apart three sources of error variance: variance originating from (1) repositioning the subject between two measurements on the same day (session-specific error variance; mostly due to different head position inside the coil); (2) repositioning the subject on another day (day-specific error variance; e.g., due to different environmental properties, intrasubject changes, or scanner-related properties); and (3) other sources of error (residual error variance). Given that the quantities derived from multiparameter maps, if measured reliably, can contribute to a better understanding of individual differences in brain physiology and age-related changes therein, estimating the size of these sources of error variance is of great methodological interest. In particular, if the sum of these three sources of error is small relative to the magnitude of between-person differences, then reliability as indexed by ICC is high, which bodes well for the investigation of individual differences in brain physiology and potential relations to

individual differences in behavior. Conversely, if the sum of these three sources of error is relatively large, then these parameters are not well suited for investigating individual differences of any sort, including associations to behavior.

## 2 | METHODS

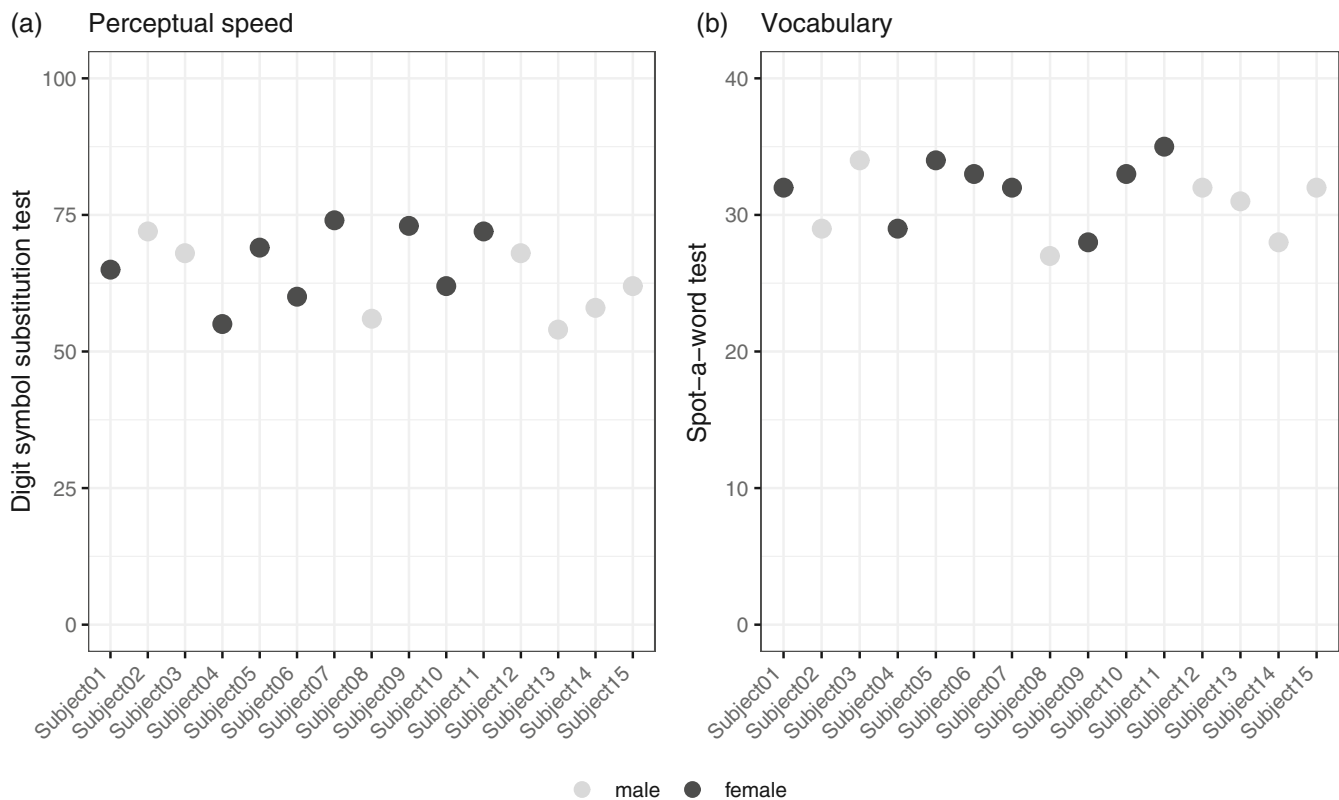
### 2.1 | Participants and procedure

Fifteen healthy volunteers (8 females, mean age = 27.30, SD =  $\pm 3.34$ , range = 22–31 years) participated in the study. All participants had normal hearing, normal or corrected-to-normal vision, no history of psychological or neurological diseases, and no contraindication to participate in an MR study, like metallic implants, tinnitus, or claustrophobia. The sample was quite representative regarding general cognitive functioning, as indicated by perceptual speed performance measured via the Digit Symbol substitution test (Wechsler, 1981) and vocabulary via the Spot-a word test (Lehrl et al., 1991;  $M = 31.3$ , SD = 2.5; see Figure 1). For comparison, data from a meta-analysis by Hoyer et al. (2004) showed a mean performance in the Digit Symbol substitution test of 69.8 in younger adults, and data from another Berlin-based training study with 100 younger adults had a mean performance of 60.3 (SD = 9.5) in this test (Schmiedek et al., 2010), and data from yet another Berlin-based training study with 44 younger

adults reported a very comparable mean performance in the vocabulary test of 30.3 (SD = 2.7; Lövdén et al., 2012). These behavioral test results give reason to believe that the chosen sample is representative of a young adult population on the cognitive level and may therefore most likely also show to-be-expected variance in brain structure, even though this link from cognition to brain structure is of course speculative.

Each participant was scanned four times, distributed over two consecutive days. On Day 1, participants were scanned for the first time with the full MPM protocol (Measurement 1). No repositioning was done, that is, participants remained inside the scanner, but all scanner adjustments and settings were reset. Then, participants were measured a second time with the MPM protocol (Measurement 2). For the Day 2 measurement, all participants were re-invited to be scanned on the following day around the same time of the day and the full MPM protocol was acquired again (Measurement 3). After that, participants were moved out of the scanner, the head coil was removed and participants briefly got up and walked around before lying back down on the scanner bed and being moved back in. Participants were then scanned for the fourth time (Measurement 4).

The study received ethical approval by the ethics committee of the German Association of Psychology (Deutsche Gesellschaft für Psychologie, DGPs) and was carried out as a pilot study in preparation for a training intervention study. All participants provided written informed consent prior to participation.



**FIGURE 1** General cognitive functioning of all 15 participants, as measured by the Digit symbol substitution test for perceptual speed (Wechsler, 1981) and the Spot-a-word test (MWT-A) for vocabulary knowledge (Lehrl et al., 1991;  $M = 31.3$ , SD = 2.5)

## 2.2 | MR image acquisition

All MPM datasets were acquired on a Siemens Tim Trio 3 T MR scanner (Erlangen, Germany; VB17a software version) with a standard radio-frequency (RF) 32-channel receive head coil and RF transmit body coil. The MPM protocol comprised one static magnetic ( $B_0$ ) GRE-field map, one RF transmit field map ( $B_1^+$ ), and three multiecho 3D FLASH (fast low angle shot) sequences. The MPM acquisition and postprocessing have been developed and described in previous studies (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008; Weiskopf et al., 2011, 2013) and acquisition parameters were chosen in accordance with previously published work (Weiskopf et al., 2013).

The  $B_0$  gradient echo field mapping sequence was acquired with the following parameters: 64 transverse slices, slice thickness = 2 mm with 50% distance factor, repetition time (TR) = 1020 ms, echo times (TE) TE<sub>1</sub>/TE<sub>2</sub> = 10/12.46 ms, flip angle  $\alpha = 90^\circ$ , matrix =  $64 \times 64$ , field of view (FOV) =  $192 \times 192$  mm, right-left phase encoding (PE) direction, bandwidth (BW) = 260 Hz/Px, flow compensation, acquisition time = 2:14 min.

Maps of the local RF transmit/ $B_1^+$  field were acquired following recommendations by Lutti and colleagues (Lutti et al., 2010) and were measured and estimated from a 3D EPI acquisition of spin and stimulated echoes (SE and STE) with different flip angles. The following parameters were used: 4 mm isotropic resolution, matrix =  $64 \times 48 \times 48$ , FOV =  $256 \times 192 \times 192$  mm, parallel imaging using GRAPPA factor  $2 \times 2$  in PE and partition directions, TR = 500 ms, TE<sub>SE/STE</sub>/mixing time = 39.06/33.80 ms. Eleven pairs of SE/STE image volumes were measured successively employing decreasing flip angles  $\alpha$  from  $115^\circ$  to  $65^\circ$  in steps of  $-5^\circ$  (applied in a  $\alpha$ - $2\alpha$ - $\alpha$  series of RF pulses to produce SEs and STEs; see Akoka et al., 1993). Acquisition time was 3 min.

The three different multiecho FLASH sequences were acquired with predominantly  $T_1$  weighting ( $T_1w$ ), proton density weighting (PDw), or magnetization transfer weighting (MTw) by appropriate choice of repetition time (TR) and flip angle  $\alpha$  ( $T_1w$ : TR/ $\alpha = 24.5$  ms/ $21^\circ$ ; PDw and MTw: TR/ $\alpha = 24.5$  ms/ $6^\circ$ ) and by applying an off-resonance Gaussian-shaped RF pulse (4 ms duration,  $220^\circ$  nominal flip angle, 2 kHz frequency offset from water resonance) prior to excitation in case of the MTw sequence version. Multiple gradient echoes with alternating readout polarity were acquired at six equidistant echo times (TE) between 2.34 and 14.04 ms for the  $T_1w$  and MTw acquisitions with two additional echoes at TE = 16.38 ms and 18.72 ms for the PDw acquisition. A high readout bandwidth (BW) = 465 Hz/pixel was used to minimize off-resonance artifacts. For an effective spoiling of transverse magnetization after each TR, gradient spoilers combined with RF spoiling were used with a phase increment of  $137^\circ$ .

To speed up data acquisition, GRAPPA parallel imaging with an acceleration factor of two was applied in the phase-encoding (anterior-posterior) direction (outer/slow phase encoding loop) and 6/8 partial Fourier acquisitions in the partition (left-right) direction (inner/fast phase encoding loop). Additional acquisition parameters

were as follows: 1 mm isotropic resolution, 176 slices per slab, FOV =  $256 \times 240$  mm, and acquisition time of each of the three FLASH sequences = 7:03 min.

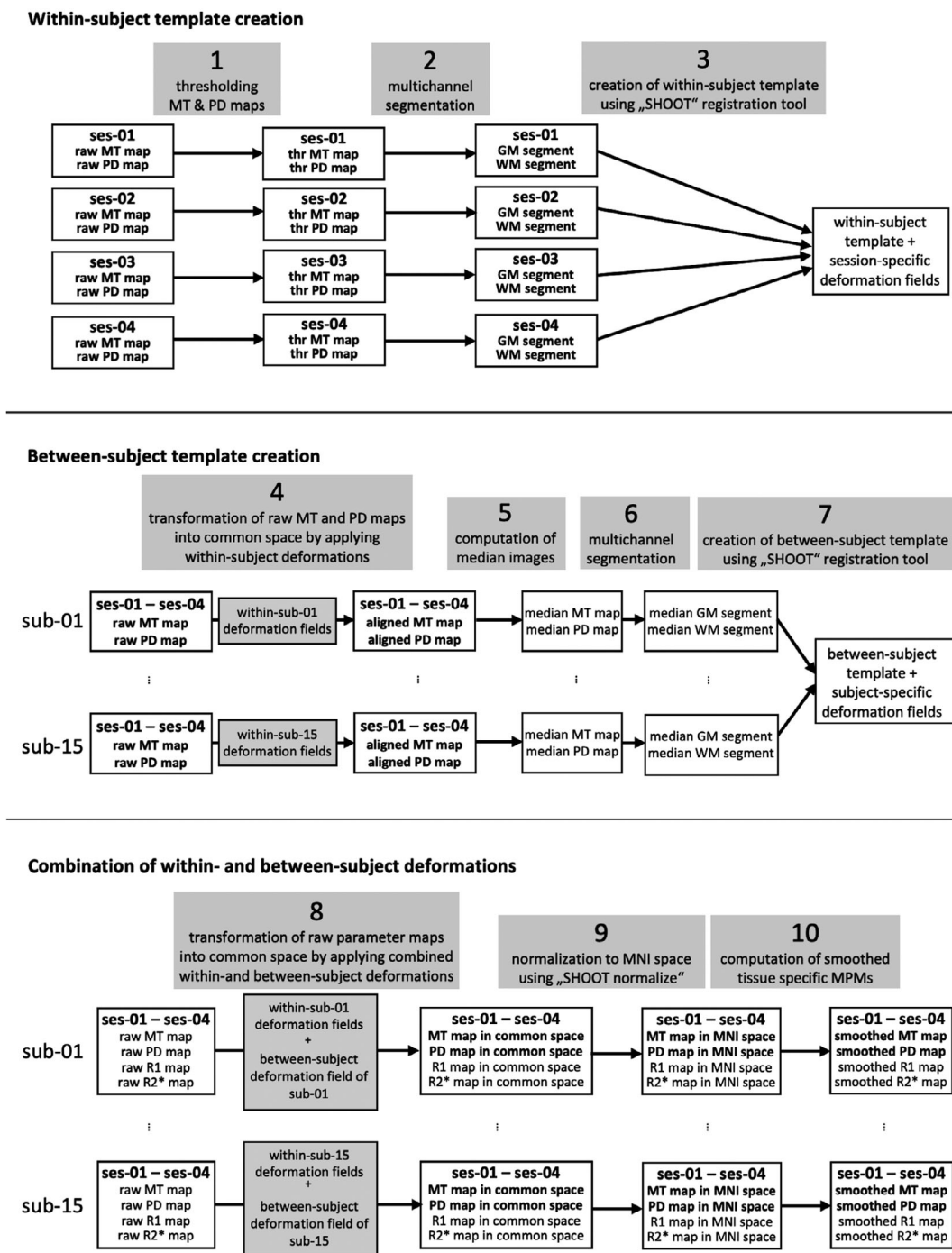
## 2.3 | Estimation of parameter maps

All data analyses and processing were performed in SPM12 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) running on Matlab 2017b (The MathWorks Inc., Natick, MA, USA) using the hMRI toolbox (Tabelow et al., 2019; <https://hmri-group.github.io/hMRI-toolbox/>). The *Create hMRI maps* module was used to compute quantitative and semi-quantitative estimates of  $R_2^*$ ,  $R_1$ , PD, and MT from unprocessed multiecho  $T_1$ -, PD-, and MT-weighted RF-spoiled gradient echo acquisitions.

As has been described in more detail elsewhere (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008; Weiskopf et al., 2011, 2013), the signal from the PD-, and  $T_1$ -weighted echoes can be described by the Ernst equation, whereas the signal strength throughout a train of gradient-recalled echoes follows a largely exponential decay with time constant  $T_2^*$ —for all three contrasts. For the MT-weighted contrast, Helms et al. have suggested to treat the MT-weighting preparation of the sequence like a first pulse in a dual-excitation FLASH sequence and introduced—based on the associated extended Ernst equation—a novel semi-quantitative parameter for describing the MT saturation effect (Helms, Dathe, & Dechent, 2008). This novel MT parameter is “semi-quantitative” since it still depends on the efficiency of the applied MT saturation, but—different from the frequently used magnetization transfer ratio (MTR)—an influence by the local  $T_1$  and also by transmit field inhomogeneities is largely canceled.

The  $R_2^*$ , that is, the effective transverse relaxation rate ( $R_2^* = 1/T_2^*$ ) was estimated by applying the ESTATICS approach (Weiskopf et al., 2014) assuming mono-exponential signal decay with increasing TE with the same  $R_2^*$  for all three contrast weightings. That is, for all three contrasts (PDw,  $T_1w$ , and MTw)—and within each contrast for the number of available TEs as datapoints—a joint log-linear fit using ordinary least squares (OLS) is applied. Thereby the slope corresponds to  $R_2^*$ , identically for all three contrasts, whereas the intercept, that is, the signal values extrapolated to TE = 0, are representing the three different contrasts without any influence of this common transversal relaxation. With this approach, relatively stable values for  $R_2^*$  are estimated for each voxel; additionally, and values largely unaffected by  $R_2^*$  are estimated for all three contrasts (i.e., extrapolated to TE = 0). Due to their minimized dependency on  $R_2^*$ , these images are an optimal basis for further calculations.

As a next step, uncorrected  $R_1$ , PD, and MT maps are calculated from the extrapolated  $T_1w$ -, PDw, and MTw measurements (for TE = 0) by applying the Ernst equation according to Helms and colleagues (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008). Quantitative maps of  $R_1$ , that is, the longitudinal relaxation rate ( $R_1 = 1/T_1$ ), were corrected for local RF transmit field inhomogeneities. To do so, the acquired 3D-EPI-based  $B_1^+$  maps (Lutti et al., 2012) were used after correcting them for EPI-specific



**FIGURE 2** Overview of the registration procedure of the multisubject scan-rescan (longitudinal) data

distortions by means of the gradient echo-based  $B_0$  fieldmaps. Also the MT maps were corrected for residual local RF transmit field inhomogeneities—using a semi-empirical approach (Rowley et al., 2021; Weiskopf et al., 2013). Imperfect RF spoiling was also corrected for in the  $T_1$  maps using the approach described by Preibisch and Deichmann (2009), which was adapted to the FLASH acquisition parameters used here. PD maps were estimated from the signal amplitude maps by adjusting for global and local receive sensitivity inhomogeneities using the “unified segmentation” approach

(Ashburner & Friston, 2005). The mean white matter PD value was calibrated to 69% units, since the global mean PD cannot be estimated accurately without an external standard.

To achieve an improved within-subject coregistration of the created maps, we adapted a longitudinal processing pipeline of the data (see Figure 2 for an overview of the longitudinal registration procedure). To do this, we first thresholded all MT maps (between 0 and 5) and all PD maps (between 0 and 200) in order to improve the segmentation performance. We then conducted multichannel segmentations



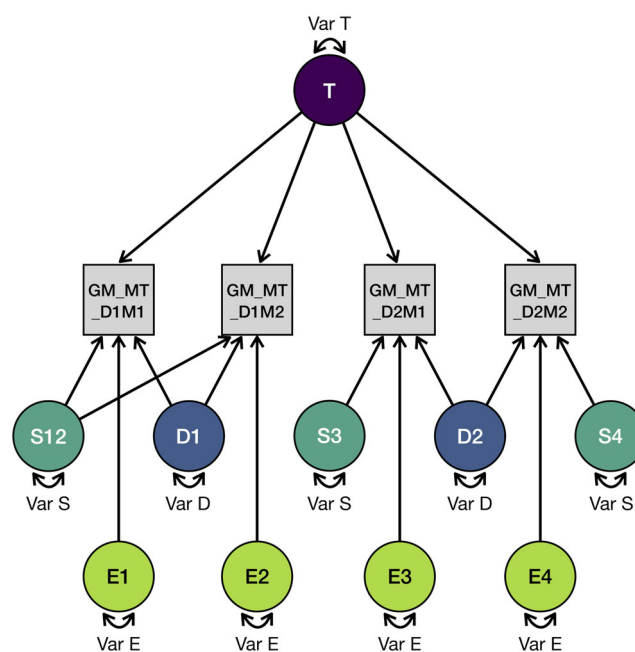
using the thresholded MT and PD map pairs for each session and each subject. The resulting gray and white matter segments from the four sessions per subject were used to create an unbiased within-subject template using the diffeomorphic registration tool with geodesic shooting “SHOOT” (Ashburner & Friston, 2011), available in SPM12 onwards. Currently, the shoot toolbox only works with images that have isotropic voxels, identical dimensions and are in approximate alignment with each other. Therefore, the registration was based on images that were first “imported” via the New Segment toolbox. The resulting deformation fields of this within-subject registration were applied to the raw MT and PD maps to warp them into each subject’s template space, and the median MT and PD map across all sessions of one individual was computed. These median maps were then also subjected to a multichannel segmentation, and the resulting median gray and white matter segments of all subjects were used to create a between-subject template, again using SHOOT, thereby also creating deformation fields from each subject’s template space to the group template space. Subsequently, we combined these two deformation fields (see Figure 2), namely the one from native space to the subject’s template space and the second from the subject’s template space to group template space. All four parameter maps (MT, PD,  $R_1$ ,  $R_2^*$ ) of all four sessions were then normalized to MNI space based on this combined deformation field, again using SHOOT. Also, the median gray and white matter segments were spatially normalized to MNI space (with application of Jacobian modulation). Following the voxel-based quantification (VBQ) approach by Draganski et al. (2011), we used these normalized tissue class segments together with the four parameter maps to finally compute “smoothed tissue specific MPMs” applying a 6 mm FWHM smoothing kernel and weighted averaging (the full set of scripts can be found on [https://git.mpib-berlin.mpg.de/plasticity/aktiv/hmri\\_scripts.git](https://git.mpib-berlin.mpg.de/plasticity/aktiv/hmri_scripts.git)).

These segmented and smoothed maps were used to extract the mean and standard deviation values across voxels in each one of the predefined regions of interest (ROIs) for every individual, at every measurement time point from the Harvard-Oxford cortical and subcortical structural atlases (<https://identifiers.org/neurovault.collection:262>; Desikan et al., 2006). In the following, we focused on a selection of ROIs that are typically of interest to neuroscientists when for example investigating language learning and effects of physical exercise. However, results for all regions of interest of the Harvard-Oxford atlas can be found here: <https://osf.io/6p9bf/>. We therefore extracted means and SD for whole gray matter (cortex; GM), whole white matter (WM), inferior frontal gyrus (IFG) pars triangularis (pars

tri), IFG pars opercularis (pars oper), orbitofrontal cortex (OFC), anterior cingulate cortex (ACC), precuneus, middle temporal gyrus (MTG), caudate nucleus, putamen, and pallidum (see Figure 3). Additionally, we also performed a whole-brain voxel-wise analysis, such that reliability was also estimated for every individual voxel.

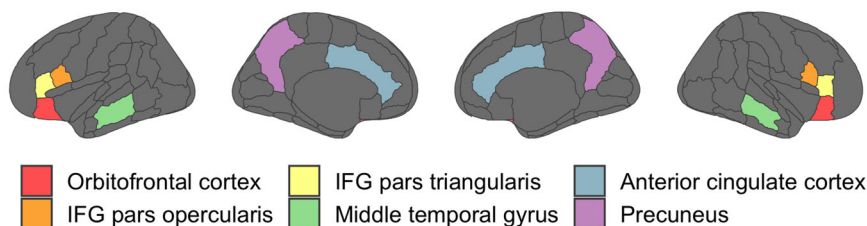
## 2.4 | Statistical analysis

We used intra-class effect decomposition (ICED) to estimate reliability (Brandmaier, Wenger, et al., 2018). This recently introduced approach uses structural equation modeling (SEM) of data to decompose reliability in orthogonal sources of measurement error that can be attributed to different measurement characteristics. Using ICED, we are



**FIGURE 4** Path diagram of a structural equation model derived from  $\Omega$ nyx. This diagram exemplifies the parameter magnetization transfer saturation (MTsat) in whole gray matter (GM). In our repeated measures design, each participant was scanned four times: Twice on Day 1 without repositioning (D1M1 and D1M2), and twice on Day 2 with repositioning (D2M1 and D2M2). Data were standardized across days and measurements, and a saturated mean structure was used. Var T = true score variance, Var S = session-specific variance, Var D = day-specific variance, Var E = residual error variance

## Regions of interest (Harvard-Oxford cortical atlas)



**FIGURE 3** Cortical regions of interest (ROIs) from the Harvard-Oxford atlas. In addition to the six selected regions displayed here, we also investigated the subcortical regions caudate, putamen, and pallidum, as well as whole gray matter and white matter additionally, we investigated reliability in a voxel-wise manner

able to estimate the main effects of session, day, and residual variance on measurement error. It therefore allows to distinguish between the following variance compartments: true-score variance (Var T; representing true between-person differences in the construct of interest), day-specific error variance (Var D), session-specific error variance (Var S; here capturing the effect of repositioning a person between the MPM datasets, accompanied by a new prescan, i.e. new adjustments of RF amplitude, center frequency and  $B_0$  shim), and residual error variance (Var E). Model specification and estimation were performed in  $\Omega$ nyx (von Oertzen et al., 2015) and lavaan, an SEM package for the statistical programming language R (Rosseel, 2012).

The path diagram in Figure 4 illustrates the ICED model for estimating the individual variance components of the total observed variance of the MPM parameter magnetization transfer saturation (MT) in gray matter. The four measurements are labeled in the path diagram as “GM\_MT\_D1M1,” “GM\_MT\_D1M2,” “GM\_MT\_D2M1,” and

“GM\_MT\_D2M2.” The same labeling convention was applied to all parameters (MT, PD,  $R_1$ ,  $R_2^*$ ) in all ROIs and in every voxel.

First, the baseline SEM model was generated, which estimated all four variance parameters related to the four sources of variance (Var T, Var D, Var S, and Var E), with a lower bound of 0.0001 applied to the estimates. Three additional null models were generated in which the true, day-related, and session-related variance were, respectively, set to zero, one at a time. To assess the significance of the magnitudes of these separate sources of error, likelihood ratio tests were used to compare the unconstrained models against the respective null models. A Wald test was used to test the residual error variance component, as the null model without an orthogonal error structure cannot be estimated. We report all variances rescaled such that they add up to one; this way, the variances can directly be interpreted as relative contributions to overall variance or variance explained.

**TABLE 1** Means and standard deviation for each ROI and for each MPM parameter: Magnetization transfer saturation (MTsat), proton density (PD), longitudinal relaxation rate ( $R_1$ ), and effective transverse relaxation rate ( $R_2^*$ )

ROIs	MTsat (p.u.)	PD (p.u.)	$R_1$ ( $s^{-1}$ )	$R_2^*$ ( $s^{-1}$ )
Gray matter	0.874 ± 0.011	80.61 ± 0.26	0.622 ± 0.007	16.9 ± 0.2
White matter	1.706 ± 0.017	69.69 ± 0.10	0.967 ± 0.012	21.5 ± 0.2
IFG (pars tri.)	0.912 ± 0.016	80.11 ± 0.38	0.644 ± 0.011	16.2 ± 0.4
IFG (pars oper.)	0.893 ± 0.013	80.46 ± 0.33	0.629 ± 0.010	15.9 ± 0.4
OFC	0.898 ± 0.027	79.96 ± 0.58	0.635 ± 0.016	17.5 ± 0.4
ACC	0.846 ± 0.014	81.98 ± 0.29	0.598 ± 0.012	14.9 ± 0.4
Precuneus	0.899 ± 0.014	80.57 ± 0.27	0.615 ± 0.012	17.5 ± 0.5
MTG (posterior)	0.879 ± 0.013	80.53 ± 0.37	0.617 ± 0.008	16.7 ± 0.3
Caudate	0.913 ± 0.024	81.31 ± 0.40	0.684 ± 0.019	19.2 ± 0.6
Putamen	1.039 ± 0.018	79.26 ± 0.44	0.744 ± 0.013	21.3 ± 0.5
Pallidum	1.211 ± 0.023	77.39 ± 0.59	0.842 ± 0.017	29.7 ± 0.7

Note: Values extracted for ROIs from the Harvard-Oxford atlas, from the longitudinally processed normalized and segmented parameter maps. Means were calculated across all voxels and all participants. Standard deviations were calculated across the means of the four measurement points (normalized by  $n - 1 = 3$ ). p.u. = percentage units.

**TABLE 2** Parameter-specific means and SDs of CoVs across participants, where each participant's CoV is calculated by dividing the SD across the four measurement points (normalized by  $n - 1 = 3$ ) by the mean of the four extracted values

Coefficient of variation				
ROIs	MT	PD	$R_1$	$R_2^*$
Gray matter	1.0% ± 0.6%	0.2% ± 0.2%	0.9% ± 0.7%	1.0% ± 0.5%
White matter	0.9% ± 0.5%	0.1% ± 0.1%	1.1% ± 0.6%	0.8% ± 0.4%
IFG (pars tri.)	1.6% ± 0.6%	0.4% ± 0.2%	1.6% ± 0.7%	2.0% ± 1.2%
IFG (pars oper.)	1.3% ± 0.5%	0.3% ± 0.2%	1.4% ± 0.8%	2.0% ± 1.1%
OFC	2.3% ± 1.9%	0.6% ± 0.4%	2.0% ± 1.3%	2.2% ± 1.1%
ACC	1.4% ± 0.8%	0.3% ± 0.1%	1.7% ± 1.0%	1.9% ± 1.7%
Precuneus	1.4% ± 0.8%	0.3% ± 0.2%	1.7% ± 0.9%	2.0% ± 2.0%
MTG (posterior)	1.2% ± 0.6%	0.4% ± 0.2%	1.0% ± 0.8%	1.7% ± 1.1%
Caudate	2.4% ± 1.0%	0.4% ± 0.3%	2.4% ± 1.4%	2.9% ± 1.2%
Putamen	1.5% ± 0.7%	0.4% ± 0.4%	1.5% ± 0.8%	1.8% ± 1.0%
Pallidum	1.7% ± 0.9%	0.6% ± 0.4%	1.8% ± 0.8%	1.9% ± 1.2%

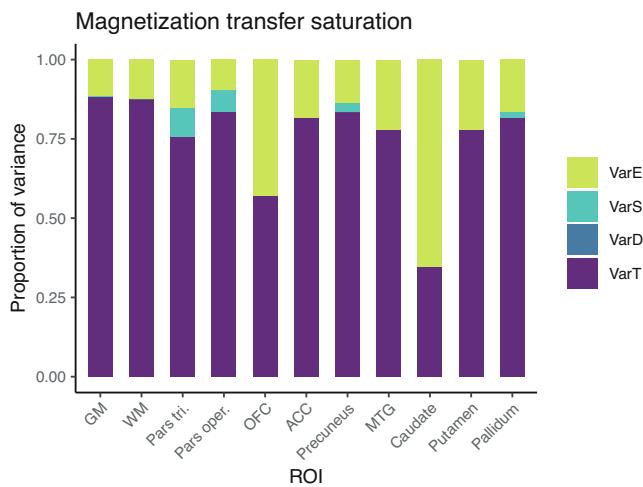
TABLE 3 Variance estimates and reliability measures

ROI	Modality	RMSEA	Var T	Var			ICC	95% CI	ICED ICC2	95% CI
				Var D	Var S	Var E				
Gray matter 1,016,729 voxels	MT	0.215	0.883*	0.000	0.003	0.114*	0.883	0.759–0.951	0.967	0.930–0.986
	PD	0.449	0.873*	0.000	0.021	0.106*	0.873	0.636–0.963	0.958	0.855–0.991
	R1	0.357	0.728*	0.000	0.000	0.272*	0.728	0.348–0.900	0.914	0.738–0.975
	R2*	0.000	0.781*	0.090	0.072	0.057*	0.781	0.547–0.866	0.885	0.740–0.939
White matter 498,645 voxels	MT	0.194	0.875*	0.000	0.000	0.125*	0.875	0.764–0.943	0.965	0.912–0.982
	PD	0.291	0.808*	0.000	0.020	0.172*	0.808	0.458–0.914	0.937	0.682–0.979
	R1	0.386	0.733*	0.000	0.029	0.238*	0.733	0.385–0.854	0.906	0.737–0.956
	R2*	0.125	0.883*	0.062	0.000	0.055*	0.883	0.624–0.950	0.952	0.825–0.986
IFG (pars tri.) 9,503 voxels	MT	0.000	0.757*	0.000	0.092	0.151*	0.757	0.598–0.864	0.891	0.769–0.948
	PD	0.522	0.769*	0.000	0.000	0.231*	0.769	0.465–0.903	0.930	0.807–0.973
	R1	0.274	0.562*	0.000	0.000	0.438*	0.562	0.242–0.769	0.837	0.512–0.929
	R2*	0.158	0.389*	0.000	0.229	0.381*	0.389	0.030–0.671	0.627	0.082–0.902
IFG (pars oper.) 11,674 voxels	MT	0.000	0.835*	0.000	0.069	0.096*	0.835	0.654–0.914	0.927	0.816–0.971
	PD	0.420	0.816*	0.000	0.000	0.184*	0.816	0.536–0.926	0.946	0.783–0.981
	R1	0.122	0.547*	0.000	0.000	0.453*	0.547	0.173–0.797	0.828	0.510–0.935
	R2*	0.113	0.427*	0.199	0.000	0.374*	0.427	0.000–0.742	0.688	0.000–0.902
OFC 25,157 voxels	MT	0.258	0.570*	0.000	0.000	0.430*	0.570	0.246–0.819	0.841	0.557–0.953
	PD	0.307	0.584*	0.000	0.000	0.416*	0.584	0.354–0.685	0.849	0.745–0.900
	R1	0.132	0.424*	0.000	0.000	0.575*	0.424	0.000–0.778	0.747	0.042–0.931
	R2*	0.290	0.523*	0.000	0.134	0.343*	0.523	0.202–0.770	0.761	0.433–0.917
ROI	Modality	RMSEA	Var T	Var D	Var S	Var E	ICC	95% CI	ICED ICC2	95% CI
ACC 20,844 voxels	MT	0.192	0.817*	0.000	0.000	0.182*	0.817	0.630–0.928	0.947	0.869–0.980
	PD	0.541	0.828*	0.000	0.000	0.172*	0.828	0.693–0.911	0.950	0.877–0.974
	R1	0.153	0.453*	0.000	0.000	0.547*	0.453	0.124–0.727	0.768	0.380–0.917
	R2*	0.338	0.754*	0.000	0.000	0.246*	0.754	0.410–0.919	0.924	0.737–0.976
Precuneus 44,699 voxels	MT	0.000	0.834*	0.000	0.031	0.135*	0.834	0.629–0.942	0.942	0.809–0.978
	PD	0.241	0.840*	0.016	0.000	0.145*	0.840	0.660–0.945	0.950	0.894–0.983
	R1	0.208	0.612*	0.000	0.000	0.388*	0.612	0.307–0.803	0.863	0.623–0.941
	R2*	0.304	0.209	0.626*	0.000	0.165*	0.209	0.000–0.671	0.371	0.000–0.856
MTG (post.) 21,879 voxels	MT	0.300	0.780*	0.000	0.000	0.220*	0.780	0.542–0.898	0.934	0.831–0.973
	PD	0.177	0.836*	0.017	0.031	0.117*	0.836	0.701–0.919	0.938	0.834–0.976
	R1	0.000	0.660*	0.105	0.129	0.106*	0.660	0.151–0.875	0.810	0.378–0.960
	R2*	0.173	0.532*	0.297*	0.000	0.171*	0.532	0.234–0.759	0.735	0.428–0.920
Caudate 5,750 voxels	MT	0.000	0.346*	0.000	0.000	0.654*	0.346	0.083–0.539	0.679	0.197–0.822
	PD	0.000	0.662*	0.000	0.000	0.337*	0.662	0.214–0.858	0.887	0.635–0.967
	R1	0.294	0.433*	0.000	0.000	0.567*	0.433	0.222–0.656	0.753	0.594–0.897
	R2*	0.000	0.571*	0.000	0.000	0.429*	0.571	0.183–0.845	0.842	0.458–0.951
Putamen 9,626 voxels	MT	0.236	0.780*	0.000	0.000	0.220*	0.780	0.550–0.893	0.934	0.821–0.969
	PD	0.418	0.596*	0.000	0.211	0.193*	0.596	0.165–0.861	0.771	0.264–0.957
	R1	0.220	0.666*	0.000	0.034	0.301*	0.666	0.328–0.851	0.876	0.618–0.963
	R2*	0.000	0.766*	0.077	0.014	0.143*	0.766	0.227–0.918	0.903	0.429–0.978
Pallidum 4,244 voxels	MT	0.244	0.817*	0.000	0.017	0.166*	0.817	0.641–0.917	0.941	0.857–0.976
	PD	0.034	0.483*	0.000	0.197	0.320*	0.483	0.192–0.741	0.710	0.339–0.900
	R1	0.213	0.715*	0.000	0.000	0.285*	0.715	0.338–0.850	0.909	0.726–0.959
	R2*	0.000	0.861*	0.033	0.023	0.082*	0.861	0.669–0.945	0.945	0.833–0.985

Abbreviation: RMSEA, root mean squared error of approximation.

\* $p$  of chisq difference from null model < .05 (VarT, VarD, VarS), or  $p$  of Wald test < .05 (Var E).





**FIGURE 5** Distribution of magnitudes of sources of variance for magnetization transfer across the different ROIs. Note that absolute magnitudes are displayed, irrespective of significance. Session- and day-specific variances were not significantly different from zero for MT

These variance components were then used to calculate ICC, an index of reliability between MPM datasets within the study. We defined ICC as the ratio of between-person variance to total variance at the level of observed variables. In addition, we calculated ICC2 as a measure of reliability on the construct level. As such, it is defined as the ratio of true score variance to total (effective) variance, where the effective error is the single residual error term that arises from all variance components other than the construct that is to be measured. When assuming no day-specific and session-specific effects, we would obtain exactly the classical definition of ICC2, which scales the residual error variance with the number of measurement occasions (Brandmaier, Wenger, et al., 2018, for exact definitions and formulae). In addition, ICC is a coefficient describing test-retest reliability of a single measurement (i.e., how well can a single measurement measure the underlying quantitative value), whereas ICC2 is a coefficient describing test-retest reliability of the entire design (i.e., how well can we measure the underlying quantitative value with multiple measurements; here a total of four measurements in the given study design). As ICC2 is contingent upon a certain study design (with a specific number of measurement occasions), we prefer to rely on ICC in our interpretation of the data, as it relies on a single occasion of measurement per person and thus provides a lower bound of reliability. ICC2 values are additionally reported for reference. Note that as the variance estimates were rescaled such that they represent proportions of the total variance, true score variance is equal to the ICC here.

Bootstrapped 95% confidence intervals, using 1000 samples, were generated for ICC and ICC2 values of each parameter (MT, PD,  $R_1$ , and  $R_2^*$ ) in all ROIs (using *boot.ci* in lavaan).

We report mean values of all MPM parameters and CoVs. We calculated CoV for each parameter in each ROI by dividing the standard deviation of the four extracted means (SD, normalized by  $N - 1$  sample size to avoid bias; in our case, 4 scans  $- 1$ ) by the overall mean across all four measurement points ( $\text{CoV} = \text{SD}/\text{Mean}$ ). In addition, we

visualize true score variance, that is ICC, in every voxel in whole-brain maps where the whiter a voxel is displayed, the closer its ICC is to 1.

## 3 | RESULTS

### 3.1 | Mean MPM parameter values and CoVs

The means and SDs, as well as CoVs of all four MPM parameters in all ROIs are presented in Tables 1 and 2.

### 3.2 | Reliability of MPM parameters and variances explained

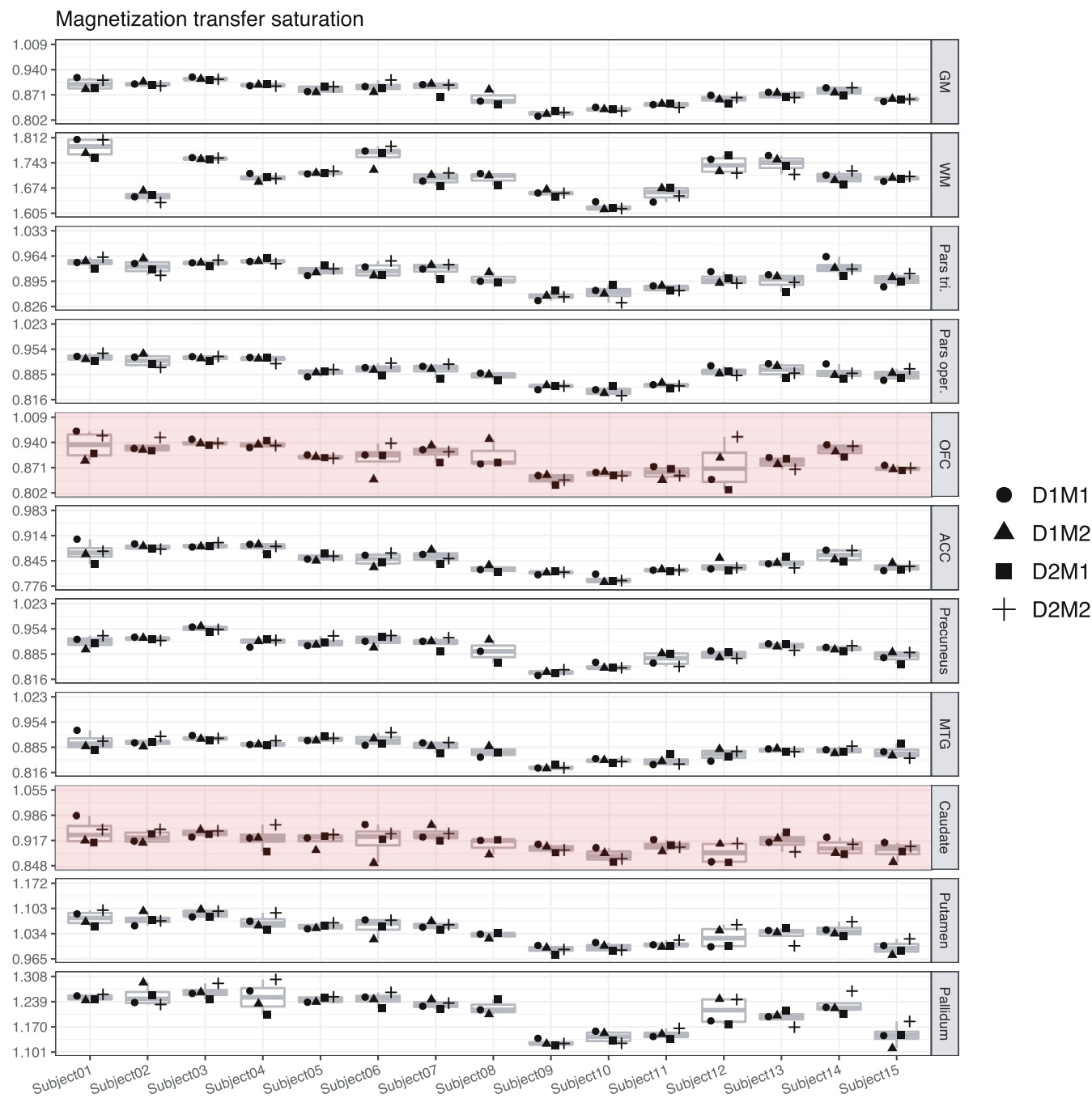
In Table 3, we summarize all estimates for the four sources of variance: true-score variance (i.e., variance attributable to between-person differences), day-specific error variance, session-specific error variance, and residual error variance, as well as ICC and ICC2, calculated using ICED. As the variance components are rescaled such that they add up to one, they can directly be interpreted as relative contributions to the total observed variance.

#### 3.2.1 | Variance in MT

In both whole gray and white matter, most variance in MT could be attributed to true score variance (ca. 88%). There were no day- or session-specific effects, and relatively small amounts of variance appeared as residual error (11% and 13%). This was also reflected in high ICC values (GM: 0.88, WM: 0.88). In the localized ROIs, true-score variance in MT varied quite a bit, from high (84% in pars opercularis, 83% in precuneus, 82% in ACC, 82% in pallidum, 78% in MTG, 78% in putamen, 76% in pars triangularis), to lower (57% in OFC), and very low (35% in caudate). There were no significant day- or session-specific effects in any of the regions, indicating a robustness against repositioning. Therefore, ICCs of MT were excellent for most of the regions and were only lower for OFC and caudate. In particular, caudate showed a poor reliability estimation of MT compared to the other regions. This is also reflected in the wide confidence interval of the ICC for MT in caudate, ranging from 0.07 to 0.54 (Figures 5 and 6).

#### 3.2.2 | Variance in PD

In both whole GM and WM, the relative proportion of true-score variance for PD was very high (87% and 81%), relatively small amounts appeared as residual error variance (11% and 17%) and again no significant effects of day or session appeared indicating no effects of repositioning. This was also reflected in high ICC values. Note, though, that per construction the absolute variance of PD in WM is negligible as PD is set to 69 p.u., see also second row from Figures 7 and 8.

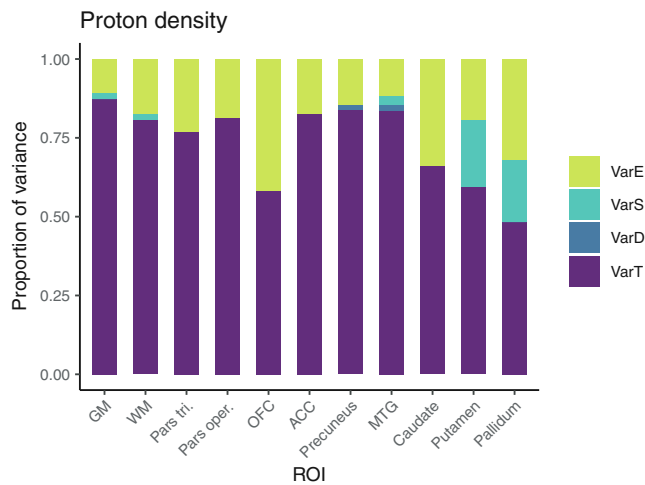


**FIGURE 6** Boxplots of MT values for each of the 15 participants and their four measurements. Parts of the plots marked in red indicate an attenuated ICC value ( $<0.75$ ) for this region. Note that the range displayed on the y-axis differs across ROIs; importantly, though, the width of the displayed range is constant across ROIs to ensure comparability and is always 0.207

True-score variance for PD was also high in precuneus (84%), MTG (84%), ACC (83%), pars opercularis (82%), and pars triangularis (77%), and slightly lower in caudate (66%), putamen (60%), OFC (58%), and pallidum (48%), again with no significant day- or session-specific variances. ICC values for PD were all excellent, except for lower values in caudate (0.66, CI ranging from 0.15 to 0.85), in putamen (0.60, CI ranging from 0.17 to 0.86), in OFC (0.58, CI ranging from 0.41 to 0.70), and pallidum (0.48, CI ranging from 0.19 to 0.74).

### 3.2.3 | Variance in $R_1$

In general,  $R_1$  and  $R_2^*$  values exhibited overall much smaller proportions of true-score variances than MT and PD parameters. In both whole GM and WM, true-score variance for  $R_1$  accounted for 73% of the estimated variance, and larger amounts appeared as residual error variance (27% and 24%), while effects of day and session were still not significantly different from zero, that is, showed no effects of



**FIGURE 7** Distribution of magnitudes of sources of variance for PD across the different ROIs. Note that absolute magnitudes are displayed, irrespective of significance. Session- and day-specific variances were not significantly different from zero for PD

repositioning. Accordingly, ICC values for  $R_1$  in whole GM and WM were only fair, both at 0.73. For the smaller, localized ROIs, true-score variances for  $R_1$  were considerably smaller, the highest being in pallidum (72%), putamen (67%), MTG (66%), precuneus (61%), pars triangularis (56%), and pars opercularis (55%), and the lower ones in ACC (45%), caudate (43%), and OFC (42%), the latter three with a higher proportion of residual error variance than true-score variance at 55%, 57%, and 58%. Accordingly, ICC values were rather low for ACC, OFC, and caudate with wide confidence intervals and only slightly better for pars triangularis, pars opercularis, precuneus, and MTG. Indeed, for  $R_1$ , none of the ICC values was above our desired threshold for reliability of 0.75 (Figures 9 and 10).

### 3.2.4 | Variance in $R_2^*$

In GM and WM, true-score variance was found to be 78% and 88% of the total variance, with a residual error variance of 6% in each. ICC values were excellent at 0.78 and 0.88. The proportion of true-score variance for  $R_2^*$  was excellent in ROIs of the basal ganglia, namely 77% in putamen, 86% in pallidum, and was also excellent in ACC (75%). However, it was considerably lower in the remaining ROIs, with 57% in caudate, 53% in MTG, 52% in OFC, and 43% in pars opercularis, and was poor in pars triangularis with 39% and in precuneus with 21%, a value not significantly different from the null model. For  $R_2^*$ , there were significant day-specific effects in MTG (Var D = 30%) and precuneus (Var D = 63%). However, as total variance in these regions was numerically so small, the allocation of this small variance to day-specific effects should not be over-interpreted as true effects of repositioning here. For  $R_2^*$ , ICC values were overall relatively low, compared to MT and PD, except for putamen and pallidum and ACC, and reached an unacceptable value of 0.21 in precuneus (Figures 11 and 12).

### 3.2.5 | Voxel-specific analysis

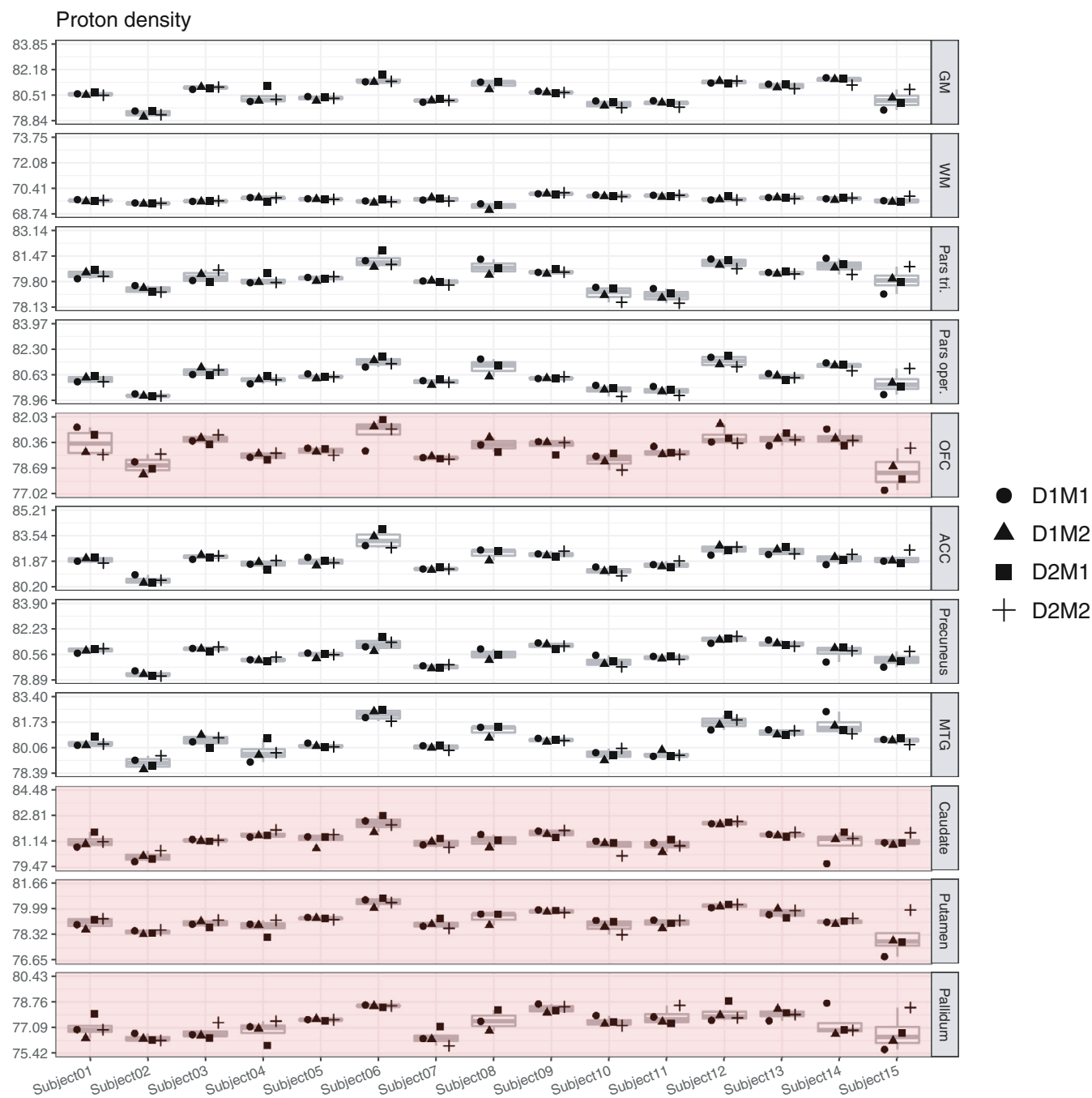
In addition to the ROI analysis, we performed the above-described reliability estimation also in every single voxel. Figure 13 displays true score variance, that is identical to ICC in our case, in all gray matter voxels, such that the whiter a voxel appears, the higher its ICC value. These whole-brain maps of variability for all four parameter maps for both gray and white matter can be found in OSF at <https://osf.io/6p9bf/>. A visual inspection of these images mirrors the above described pattern of overall good ICC for MT and PD but attenuated values for  $R_1$  and  $R_2^*$ .

## 4 | DISCUSSION

In this study, we investigated the test-retest reliability of four MPM parameters, namely MT, PD,  $R_1$ , and  $R_2^*$ , in whole gray and white matter as well as in selected gray matter ROIs that are commonly of interest in cognitive neuroscience studies. To evaluate reliability, we used ICED (Brandmaier, Wenger, et al., 2018), which partitions multiple sources of unreliability into its constituent components and therefore provides a more detailed picture of the parameter properties than CoV or ICC alone.

A basic check of value plausibility shows that the measured parameter values fall well within the range of those from previously published studies. For example, the  $R_1 = 0.62/0.98 \text{ s}^{-1}$  in GM/WM was similar to the  $R_1$  reported previously  $0.61/1.04 \text{ s}^{-1}$  (Weiskopf et al., 2013) and  $0.63/1.19 \text{ s}^{-1}$  (Wright et al., 2008). The same holds for our PD estimates of 80.6/81.3 p.u. in GM/caudate, which were comparable to previous studies reporting 84.4/82.7 p.u. (Weiskopf et al., 2013), 81.1/81.5 p.u. (Volz et al., 2012), and 82.2/84.8 p.u. (Neeb et al., 2008) for the same structures. The small numerical difference here might originate from the usage of effective PD, that is, not extrapolating to  $TE = 0$  in some previous publications as, for example, in the study by Weiskopf et al. (2013). Our estimates of  $R_2^*$  were also similar to previously published values, namely 16.9/21.5  $\text{s}^{-1}$  compared to 15.2/21.0  $\text{s}^{-1}$  in GM/WM (Weiskopf et al., 2013) and 19.5  $\text{s}^{-1}$  and 21.7  $\text{s}^{-1}$  in WM (Baudrexel et al., 2009; Martin et al., 2008). For all parameters, the choice of resolution can lead to a certain change of the measured parameter values by partial volume effects. In case of  $R_2^*$ , the chosen resolution can produce even more substantial deviations due to the fact that for larger voxels the intravoxel  $B_0$  field is more inhomogeneous, which causes larger  $R_2^*$  values.

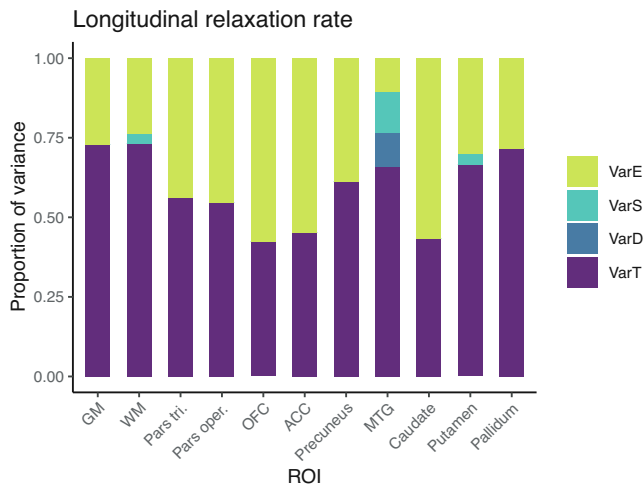
All four MPM parameters showed excellent reliabilities for whole gray and white matter across the four measurements. However, we noted marked differences in reliability among the four MPM parameters for different regions of the brain: MT and PD exhibited excellent reliability and were robust against participant repositioning within a scanning session and on different days in nearly all regions, except for OFC (ICCs fair at 0.57 for MT and 0.58 for PD), caudate (ICC for MT poor at 0.35, ICC for PD good at 0.66), pallidum (ICC for PD fair at 0.48), and putamen (ICC for PD good at 0.60). In contrast,  $R_1$  and  $R_2^*$



**FIGURE 8** Boxplots of PD values for each of the 15 participants and their four measurements. Parts of the plots marked in red indicate an attenuated ICC value ( $<0.75$ ) for this region. Note that the range displayed on the y-axis differs across ROIs; importantly, though, the width of the displayed range is constant across ROIs to ensure comparability and is always 5.01

showed only fair reliability in most regions ( $<0.60$ ) with the exception of regions of the basal ganglia, namely putamen and pallidum, and even poor reliability ( $<0.40$ ) for  $R_2^*$  in pars triangularis and precuneus. In some regions, the sum of the error estimates exceeded between-person differences (e.g., for  $R_1$  in ACC, OFC, and caudate), effectively rendering it hard if not impossible to interpret between-person differences and continue with correlational approaches. For  $R_2^*$  in precuneus, the proportion of true-score

variance was not significantly different from zero, that is, all participants yielded very similar values on this parameter in precuneus. For that region,  $R_2^*$  did not convey any person-specific information that would be potentially associated with any sort of between-person differences in behavior. Given that cognitive neuroscience often aims at delineating brain-behavior relations (Blakemore & Lindenberger, 2020; Krakauer et al., 2017), knowing about these differences in reliability is of critical importance.



**FIGURE 9** Distribution of magnitudes of sources of variance for  $R_1$  across the different ROIs. Note that absolute magnitudes are displayed, irrespective of significance. Also for  $R_1$ , session- and day-specific variances were not significantly different from zero

These observations on differential reliability of the four MPM parameters reported above exist in parallel to observations of high precision of these parameters, when calculated as the variation over four different measurements on the same scanner, with CoVs ranging between 0.2% and 2.4%. Our estimates of CoVs are even better than those reported before, for example, in an intersite and intrasite validation study (Weiskopf et al., 2013), or in another study testing within-site and between-site reproducibility of MPM (Leutritz et al., 2020). In general, other studies have also found quantitative MRI to be highly reproducible across software updates, different sites, and even across different vendors (Gracien et al., 2020; Lee et al., 2019; Leutritz et al., 2020). It is important to note that the findings of high reproducibility as assessed with CoVs, and those of varying reliability between the four MPM parameters across different ROIs as assessed with ICCs do not contradict each other, but are simply different pieces of the same puzzle. To repeat, these two notions of reliability both capture the precision of measurement; however, they standardize them with respect to different sample values; CoV standardizes against the sample mean, whereas ICC standardizes against the between-person variance. In experimental approaches, the former logic may be more apt to quantify a standardized precision of measurement, whereas the precision necessary for correlational (individual differences) studies is better captured by the latter.

By definition, ICC values must be interpreted contingent upon the characteristics of a given population. This is important to keep in mind when interpreting a given ICC. To the extent a given study population has lower or higher true score variability, reliability will increase or decrease proportionally. Here, we have measured 15 healthy younger adults aged 20–30 years. In this sample, ICCs were high for MT and PD but attenuated for  $R_1$  and  $R_2^*$ . It is an empirical question whether similar results would be obtained in samples representing an older or diseased population. For example, it may very well be that the underlying biological characteristics

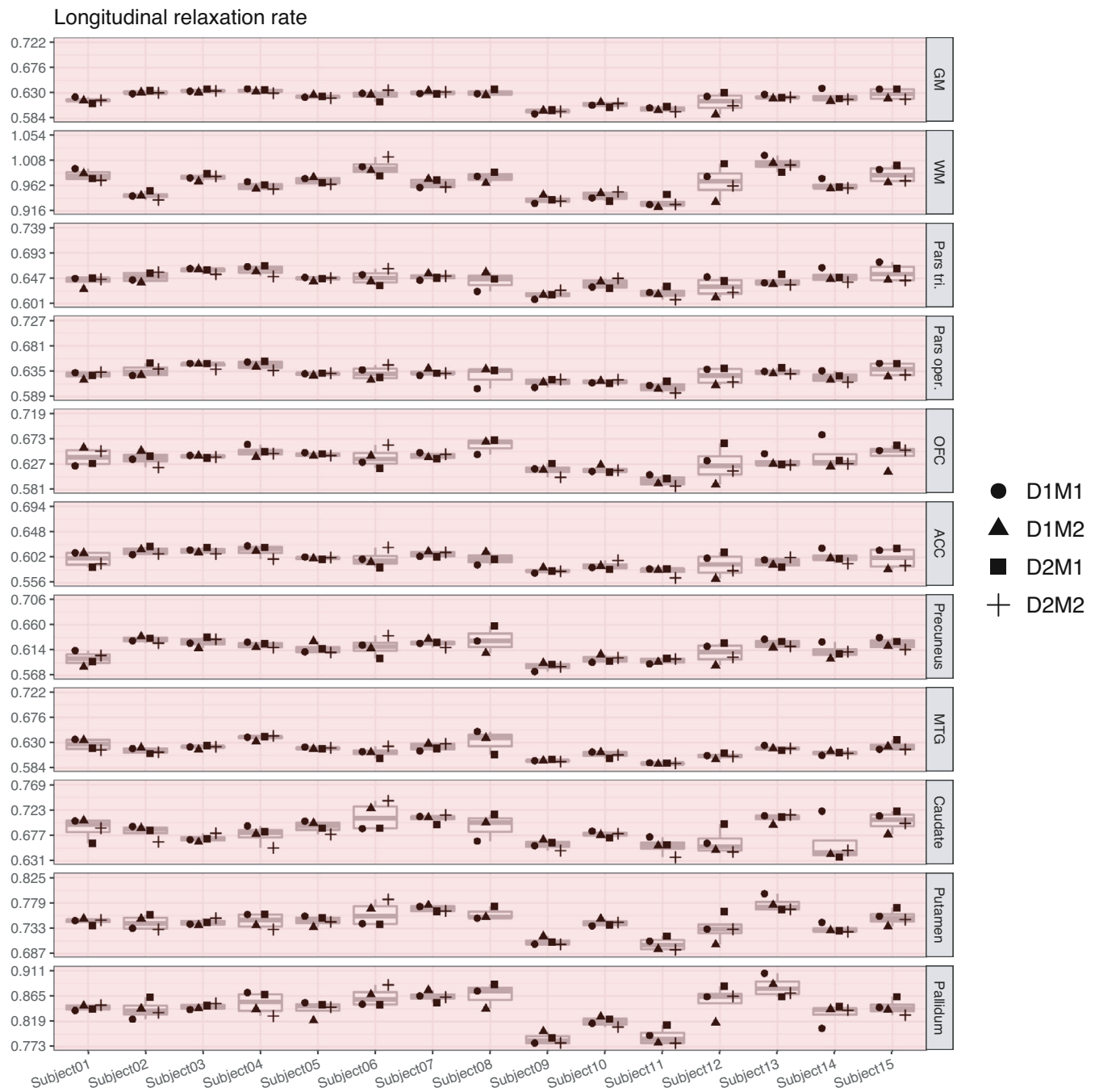
associated with these parameters simply do not yet show or do not anymore show any substantial variance in the 15 younger adults included in this study as childhood and adolescence is over and older age has not yet begun. Also, the chosen ROI approach here might limit the ability of MPM to pick up small changes. Classical voxel-based mass univariate mapping approaches might be better suited to deliver more fine-grained results that can be picked up at the individual voxel level, but go unnoticed in the averaging process done to form an ROI.

According to a commonly accepted biophysical interpretation,  $R_1$  depends on the mobility of water in its microenvironment, which is affected by certain aspects of cell membranes, and is thus related to myelination due to the presence of dense multiple myelin sheaths (Weiskopf et al., 2021).  $R_2^*$  is dependent on the magnetic field distribution, which in turn is particularly affected by iron (Cherubini et al., 2009; Duyn et al., 2007; Péran et al., 2007, 2009; Weiskopf et al., 2021). Healthy young adults may simply not differ substantially in iron deposition and molecular mobility near membranes. In other words, there may not have been between-person differences to be detected by  $R_1$  and  $R_2^*$  in this specific sample. The fact that reliability of  $R_2^*$  was considerably higher in regions of the basal ganglia speaks to the fact that this part of the brain is indeed specifically amenable to iron deposition and may therefore also be more likely to show individual differences in  $R_2^*$  than other cortical regions. Future studies using ICC as a reliability measure of MPM parameters in more diverse samples will show whether reliability attenuation of  $R_1$  and  $R_2^*$  in most cortical regions generalizes beyond the present sample of healthy young adults.

The importance of considerations on reliability of different parameters cannot be overstated. As statistical power is directly influenced by precision of measurement (besides sample size, test size, and population effect size), it is essential to be aware of the reliability of each parameter that is used to characterize, for example, age-related changes in brain structure, as is often intended when using the MPM protocol. With knowledge of reliability estimates, it is then possible to perform power analyses in the context of individual differences in longitudinal designs (Brandmaier, von Oertzen, et al., 2018), thereby enabling informed study design decisions to optimize conditions to detect a hypothesized effect. The “replicability revolution” in psychological science is an example of how changing norms can shape research practices and standards. In only a few years, practices to foster replicability, be it preregistration of hypotheses and planned analyses or publishing of analysis scripts or even data, have rapidly gained in popularity (Nosek et al., 2018). Similar norms would be highly desirable in the context of reliability: Researchers should report the reliabilities of all MRI-derived estimates whenever these are used to study individual differences, similar to what has been proposed for task-fMRI (Elliott et al., 2020). In doing so, researchers can make more informed decisions about sample size and select reliable measures for a given research question.

At the risk of redundancy, we would like to reiterate that all four MPM parameters showed excellent reproducibility, that is, they show very little variability when measurements are repeated in the same





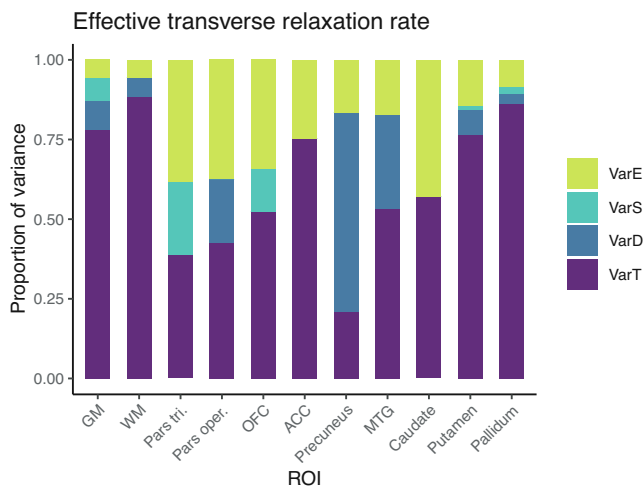
**FIGURE 10** Boxplots of  $R_1$  values for each of the 15 participants and their four measurements. ICC values were below 0.75 in all investigated regions (marked in red). Note that the range displayed on the y-axis differs across ROIs; importantly, though, the width of the displayed range is constant across ROIs to ensure comparability and is always 0.138

participant four times. However, the four parameter estimates do differ in how informative they are for assessing between-person differences in young adults. In particular,  $R_1$  and  $R_2^*$  seem to carry only limited person-specific information as can be gathered from the very low variability among participants displayed in Figures 10 and 12.

In the following, we discuss this differential reliability in light of a number of potential influences on parameter estimation.

#### 4.1 | Potential causes of variations in parameter maps

The dual flip angle mapping approach used in MPM (Helms, Dathe, & Dechent, 2008) provides signal amplitude (proportional to PD) and  $R_1$  maps that need to be corrected for RF transmit and receive field inhomogeneities. As recommended for the preprocessing stream, we used



**FIGURE 11** Distribution of magnitudes of sources of variance for  $R_2^*$  across the different ROIs. Note that absolute magnitudes are displayed, irrespective of significance. Notably, there were significant day-specific variances for  $R_2^*$  in MTG and precuneus

highly accurate and precise RF transmit field maps with a total error of less than 3% (Lutti et al., 2010, 2012) and corrected for imperfect RF spoiling. The previous effects lead to deviations from the Ernst signal equation underlying the  $R_1$  estimation and imprecisions in their correction may thus lead to inaccuracies (Corbin & Callaghan, 2021; Preibisch & Deichmann, 2009; Yarnykh, 2010). For example, even small deviations of 3% in the RF transmit field mapping can cause errors of up to 6% in the  $R_1$  maps due to the quadratic dependence of the estimated  $R_1$  on the local flip angle (Weiskopf et al., 2011). MT saturation maps are largely self-correcting and independent of the RF transmit and receive fields (Helms, Dathe, & Dechent, 2008). In addition, residual effects of RF transmit field inhomogeneities were further reduced in postprocessing based on the measured RF transmit maps (Weiskopf et al., 2013).

There is some bias to be expected in  $R_2^*$  (and PD) maps as the mono-exponential decay model applied by the ESTATICS approach and extrapolation to  $TE = 0$  is just a simplification that does not perfectly fit due to partial voluming of different cell compartments with different transversal relaxation rates as intra- and extracellular and myelin-associated spaces. Thus  $R_2^*$  is poorly defined and also the extrapolation of the signal to  $TE = 0$  for the PD mapping is inaccurate (Neeb et al., 2008; Tabelow et al., 2019; Weiskopf et al., 2013). Thus, differences in shim, which can also be caused by differences in relative orientation of the head, may have influenced  $R_2^*$  maps (Draganski et al., 2011). At the same time, the high spatial resolution of 1 mm should reduce the effects of susceptibility artifacts on the signal decay due to a smaller within voxel spin phase coherence loss (Weiskopf et al., 2007).  $R_2^*$  of highly structured brain structures can also exhibit a field orientation dependence due to their microstructural geometry even independent of shim quality (Papazoglou et al., 2019; Wharton & Bowtell, 2012). Similar effects have actually been demonstrated for

both  $R_1$  and MT within WM (Schyboll et al., 2020). Also, nonlinear  $B_0$  inhomogeneities may have an influence on the  $R_2^*$  maps, as 3D  $R_2^*$ -mapping by short TE trains have been shown to be affected by local  $B_0$  gradients (Helms & Dechent, 2010). Since the longest echo time acquired in the PD-weighted sequence was 18.79 ms, the estimation of the long  $T_2^*$  ( $=1/R_2^*$ ) found in GM, WM, or CSF is relatively poorly conditioned. The precision of the  $R_2^*$  maps may be improved by increasing the maximal echo time, but this would also prolong the total acquisition time (Weiskopf et al., 2013). Additionally, the generally relatively poor reproducibility and performance of shimming routines might influence the reliable estimation of  $R_2^*$  (Leutritz et al., 2020).

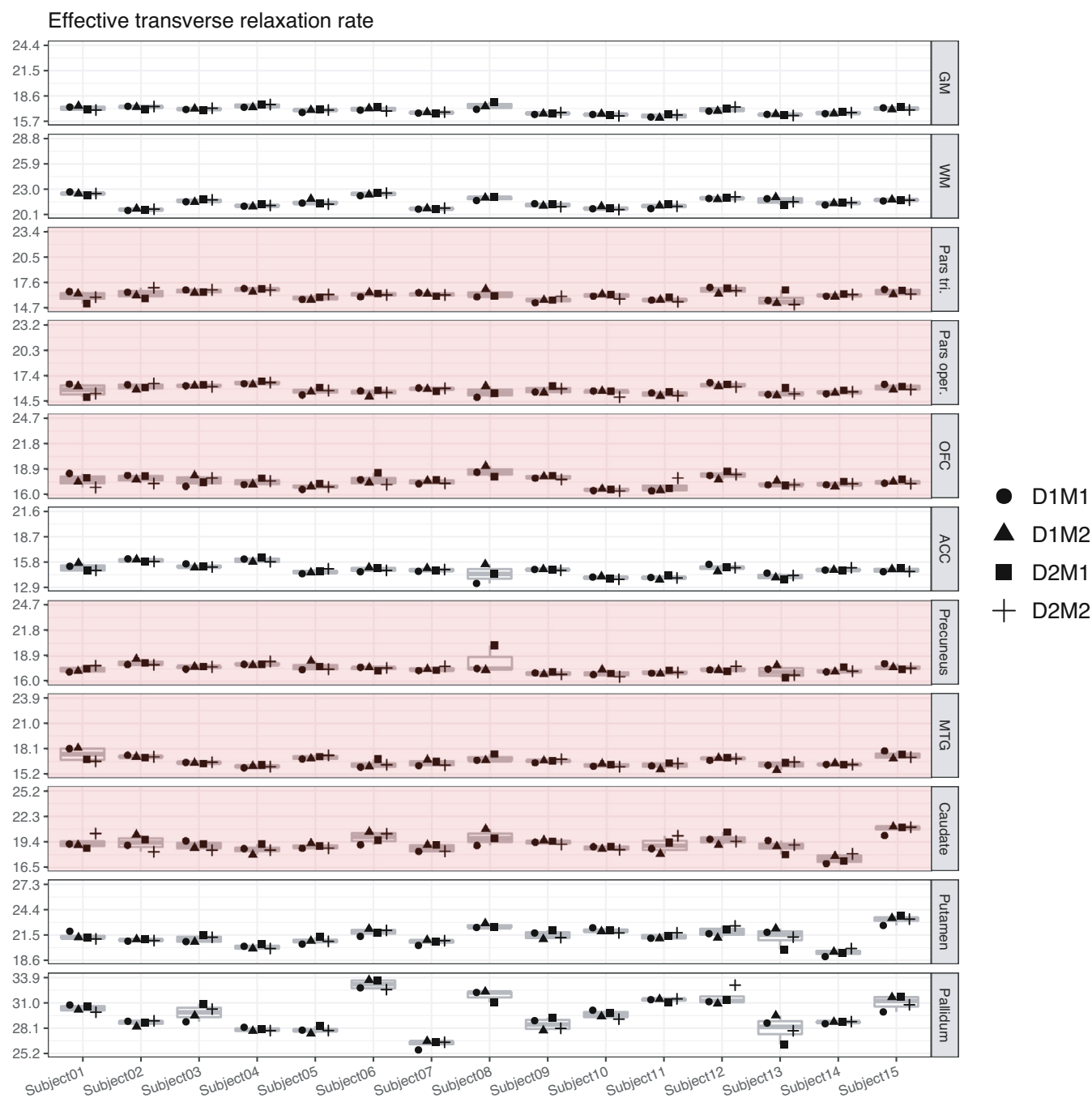
The RF receive field effect on the PD map was minimized by image postprocessing. Unified segmentation (Ashburner & Friston, 2005) was adapted to robustly determine and correct for the multiplicative receive coil sensitivity profile in the PD maps, similar to the previously developed UNICORT approach for correcting  $R_1$  maps (Weiskopf et al., 2011). Indeed, this seems to be working quite well, as the PD estimates exhibited consistently high ICC values across all the regions in our study (with only a slight attenuation in OFC and caudate).

As with any other MR sequence, MPM performance may be impaired in noncompliant volunteers. For example, some participants may have difficulties to minimize head or body motion, which can change the magnetic field in the head and affect data quality (Versluis et al., 2010; Weiskopf et al., 2013). The parameter maps are estimated from three acquired FLASH sequences and are sensitive to existing artifacts in any of these. When inspecting our plots above depicting the individual means of all four measurements for every participant, it is obvious that some participants' data were in general more variable than others (e.g., Participants 8 and 12). Some of these problems may be alleviated by using prospective motion correction (Callaghan et al., 2015; Maclaren et al., 2012) and phase navigator techniques (Versluis et al., 2010).

We also note that the physical and biophysical models underlying the modeling, analysis and interpretation of quantitative MRI and MPM data pose additional constraints. The brain tissue is a highly complex structure with a plethora of cells, cellular process and extensive vascularization. Thus, for example, the reduction and description by single compartments underlying standard relaxation parameters such as  $R_1$  and  $R_2^*$  can only partially capture the tissue's complexity, effectively causing instabilities in the aggregate parameter measures (Weiskopf et al., 2021).

## 5 | CONCLUSION

This study used ICED (Brandmaier, Wenger, et al., 2018) to investigate the reliability of MPM parameters assessed with 3 T MRI. ICED allowed us to separate sources of unreliability due to session- and day-specific effects from residual error variance. In line with earlier validation studies, we found high reproducibility of all four MPM parameters using CoV throughout all assessed regions of the brain.

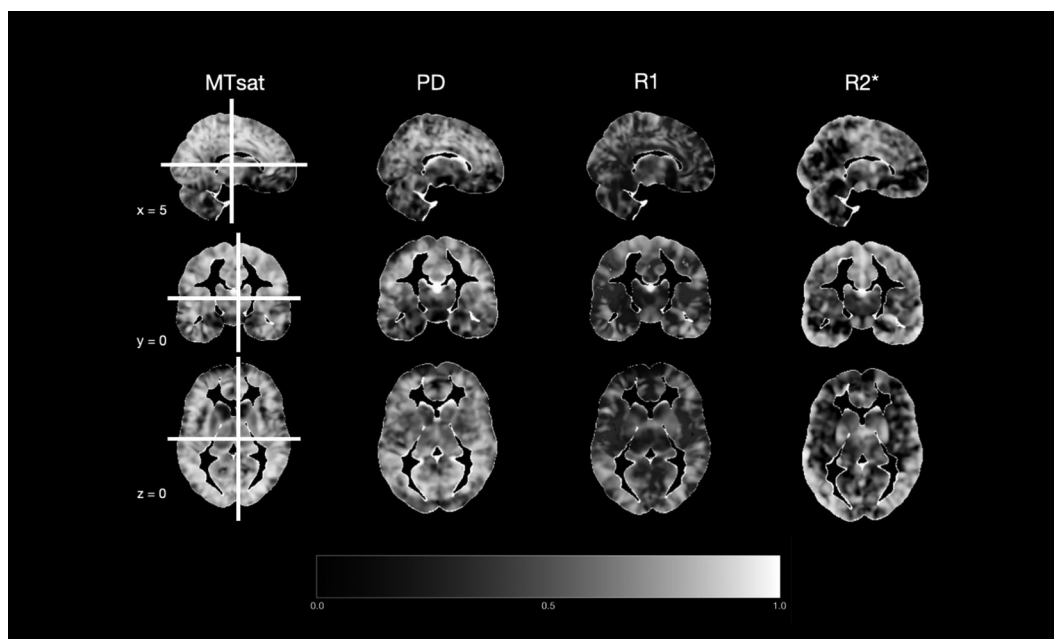


**FIGURE 12** Boxplots of  $R_2^*$  values for each of the 15 participants and their four measurements. Parts of the plots marked in red indicate an attenuated ICC value ( $<0.75$ ) for this region. Note that the range displayed on the y-axis differs across ROIs; importantly, though, the width of the displayed range is constant across ROIs to ensure comparability and is always 8.7

Going beyond common practice, we placed special emphasis on representing the precision with which MPM parameters capture between-person differences. To this end, we calculated the ICC based on ICED, which quantifies variance within persons in relation to total variance. We found that reliabilities of between-person differences were high for all four parameters in relation to whole gray and white matter. However, across different regions of the brain, the reliabilities of the four parameters varied greatly. Specifically, MT and PD emerged as highly reliable parameters that are robust against

participant repositioning in nearly all regions, whereas true-score variances were lower for  $R_1$  and  $R_2^*$ . In some regions, residual-error variances of  $R_1$  and  $R_2^*$  exceeded true-score variances, rendering the interpretation of between-person differences for these parameters unviable. We conclude that  $R_1$  and  $R_2^*$  carried little person-specific information in regions outside the basal ganglia in the present sample of healthy young adults, and recommend researchers to routinely check the reliability of MRI parameters before examining their associations to individual differences in behavior.





**FIGURE 13** Voxel-specific estimation of true score variance, that is, ICC, for all four MPM parameters in gray matter. These whole-brain maps of variability for gray matter as displayed here, as well as the ones for white matter can be found on OSF at <https://osf.io/6p9bf/>. In general, the whiter a voxel appears here, the higher its ICC value

## ACKNOWLEDGMENTS

This work was supported by the Max Planck Society and the Max Planck Institute for Human Development and is part of the BMBF funded Energi Consortium (01GQ1421B). Nikolaus Weiskopf was supported by: the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement n° 616905; the European Union's Horizon 2020 research and innovation program under the grant agreement No 681094; the BMBF (01EW1711A & B) in the framework of ERA-NET NEURON. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome (203147/Z/16/Z). The authors are very grateful to Siawoosh Mohammadi for advice on data preprocessing, to Michael Krause for continuous assistance in implementation of data preprocessing on the computing cluster and to Martina Callaghan for continued support during implementation of the MPM protocol and highly valuable comments on the final manuscript draft. The authors thank the MRI team at the Max Planck Institute for Human Development (Sonali Beckmann, Nadine Taube; Thomas Feg, and Davide Santoro) as well as all participants for their time and support. Open access funding enabled and organized by Projekt DEAL.

## CONFLICTS OF INTEREST

The Max Planck Institute for Human Cognitive and Brain Sciences has an institutional research agreement with Siemens Healthcare. Nikolaus Weiskopf holds a patent on acquisition of MRI data during spoiler gradients (US 10,401,453 B2). Nikolaus Weiskopf was a speaker at an event organized by Siemens Healthcare and was reimbursed for the travel expenses.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study as well as the R analysis script are openly available in OSF at <https://osf.io/6p9bf/>.

## ORCID

Elisabeth Wenger  <https://orcid.org/0000-0002-1830-1691>  
 Sarah E. Polk  <https://orcid.org/0000-0002-3862-8333>  
 Maïke M. Kleemeyer  <https://orcid.org/0000-0002-9388-5535>  
 Nikolaus Weiskopf  <https://orcid.org/0000-0001-5239-1881>  
 Ulman Lindenberger  <https://orcid.org/0000-0001-8428-6453>  
 Andreas M. Brandmaier  <https://orcid.org/0000-0001-8765-6982>

## REFERENCES

- Akoka, S., Franconi, F., Seguin, F., & Le Pape, A. (1993). Radiofrequency map of an NMR coil by imaging. *Magnetic Resonance Imaging*, 11(3), 437–441. [https://doi.org/10.1016/0730-725x\(93\)90078-r](https://doi.org/10.1016/0730-725x(93)90078-r)
- Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3), 839–851. <https://doi.org/10.1016/j.neuroimage.2005.02.018>
- Ashburner, J., & Friston, K. J. (2011). Diffeomorphic registration using geodesic shooting and Gauss-Newton optimisation. *NeuroImage*, 55(3–3), 954–967. <https://doi.org/10.1016/j.neuroimage.2010.12.049>
- Bartko, J. J. (1966). The Intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19(1), 3–11. <https://doi.org/10.2466/pr0.1966.19.1.3>
- Baudrexel, S., Volz, S., Preibisch, C., Klein, J. C., Steinmetz, H., Hilker, R., & Deichmann, R. (2009). Rapid single-scan T2\*-mapping using exponential excitation pulses and image-based correction for linear background gradients. *Magnetic Resonance in Medicine*, 62(1), 263–268. <https://doi.org/10.1002/mrm.21971>
- Blakemore, S. J., & Lindenberger, U. (2020). Introduction to the section “I. Brain circuits over a lifetime”. In D. Poeppel, G. Mangun, & M. S. Gazzaniga (Eds.), *The cognitive neurosciences* (pp. 3–5). MIT Press.

- Brandmaier, A. M., von Oertzen, T., Ghisletta, P., Lindenberger, U., & Hertzog, C. (2018). Precision, reliability, and effect size of slope variance in latent growth curve models: Implications for statistical power analysis. *Frontiers in Psychology*, 9, 1–16. <https://doi.org/10.3389/fpsyg.2018.00294>
- Brandmaier, A. M., Wenger, E., Bodammer, N. C., Kühn, S., Raz, N., & Lindenberger, U. (2018). Assessing reliability in neuroimaging research through intra-class effect decomposition (ICED). *eLife*, 7, 1–19. <https://doi.org/10.7554/eLife.35718>
- Callaghan, M. F., Josephs, O., Herbst, M., Zaitsev, M., Todd, N., & Weiskopf, N. (2015). An evaluation of prospective motion correction (PMC) for high resolution quantitative MRI. *Frontiers in Neuroscience*, 9, 1–9. <https://doi.org/10.3389/fnins.2015.00097>
- Cherubini, A., Péran, P., Caltagirone, C., Sabatini, U., & Spalletta, G. (2009). Aging of subcortical nuclei: Microstructural, mineralization and atrophy modifications measured in vivo using MRI. *NeuroImage*, 48(1), 29–36. <https://doi.org/10.1016/j.neuroimage.2009.06.035>
- Corbin, N., & Callaghan, M. F. (2021). Imperfect spoiling in variable flip angle T1 mapping at 7T: Quantifying and minimizing impact. *Magnetic Resonance in Medicine*, 86(2), 693–708. <https://doi.org/10.1002/mrm.28720>
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3), 968–980. <https://doi.org/10.1016/j.neuroimage.2006.01.021>
- Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R. S. J., Helms, G., & Weiskopf, N. (2011). Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). *NeuroImage*, 55(4), 1423–1434. <https://doi.org/10.1016/j.neuroimage.2011.01.052>
- Duyn, J. H., Van Gelderen, P., Li, T. Q., De Zwart, J. A., Koretsky, A. P., & Fukunaga, M. (2007). High-field MRI of brain cortical substructure based on signal phase. *Proceedings of the National Academy of Sciences of the United States of America*, 104(28), 11796–11801. <https://doi.org/10.1073/pnas.0610821104>
- Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, 31(7), 792–806. <https://doi.org/10.1177/0956797620916786>
- Gracien, R. M., Maiworm, M., Brüche, N., Shrestha, M., Nöth, U., Hattingen, E., Wagner, M., & Deichmann, R. (2020). How stable is quantitative MRI? – Assessment of intra- and inter-scanner-model reproducibility using identical acquisition sequences and data analysis programs. *NeuroImage*, 207, 1–11. <https://doi.org/10.1016/j.neuroimage.2019.116364>
- Helms, G., Dathe, H., & Dechent, P. (2008). Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magnetic Resonance in Medicine*, 59(3), 667–672. <https://doi.org/10.1002/mrm.21542>
- Helms, G., Dathe, H., Kallenberg, K., & Dechent, P. (2008). High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magnetic Resonance in Medicine*, 60(6), 1396–1407. <https://doi.org/10.1002/mrm.21732>
- Helms, G., & Dechent, P. (2010). Dependence of R2\* bias on through-voxel frequency dispersion and gradient echo train in high-resolution 3D R2\* mapping. In *Proceedings of the 18th scientific meeting ISMRM* (p. 690). ISMRM.
- Helms, G., Draganski, B., Frackowiak, R., Ashburner, J., & Weiskopf, N. (2009). Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. *NeuroImage*, 47(1), 194–198. <https://doi.org/10.1016/j.neuroimage.2009.03.053>
- Hoyer, W. J., Stawski, R. S., Wasylshyn, C., & Verhaeghen, P. (2004). Adult age and digit symbol substitution performance: A meta-analysis. *Psychology and Aging*, 19(1), 211–214. <https://doi.org/10.1037/0882-7974.19.1.211>
- Karch, J. D., Filevich, E., Wenger, E., Lisofsky, N., Becker, M., Butler, O., Mårtensson, J., Lindenberger, U., Brandmaier, A. M., & Kühn, S. (2019). Identifying predictors of within-person variance in MRI-based brain volume estimates. *NeuroImage*, 200, 575–589. <https://doi.org/10.1016/j.neuroimage.2019.05.030>
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., Mac Iver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–490. <https://doi.org/10.1016/j.neuron.2016.12.041>
- Lee, Y., Callaghan, M. F., Acosta-Cabrero, J., Lutti, A., & Nagy, Z. (2019). Establishing intra- and inter-vendor reproducibility of T1 relaxation time measurements with 3T MRI. *Magnetic Resonance in Medicine*, 81(1), 454–465. <https://doi.org/10.1002/mrm.27421>
- Lehrl, S., Merz, J., Burkard, G., & Fischer, B. (1991). Manual zum MWT-a. Perimed.
- Leutritz, T., Seif, M., Helms, G., Samson, R. S., Curt, A., Freund, P., & Weiskopf, N. (2020). Multiparameter mapping of relaxation (R1, R2\*), proton density and magnetization transfer saturation at 3 T: A multi-center dual-vendor reproducibility and repeatability study. *Human Brain Mapping*, 41(15), 4232–4247. <https://doi.org/10.1002/hbm.25122>
- Lindenberger, U., Li, S. C., & Bäckman, L. (2006). Delineating brain-behavior mappings across the lifespan: Substantive and methodological advances in developmental neuroscience. *Neuroscience and Biobehavioral Reviews*, 30(6), 713–717. <https://doi.org/10.1016/j.neubiorev.2006.06.006>
- Lövdén, M., Schaefer, S., Noack, H., Bodammer, N. C., Kühn, S., Heinze, H.-J., Düzel, E., Bäckman, L., & Lindenberger, U. (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiology of Aging*, 33(3), 620.e9–620.e22. <https://doi.org/10.1016/j.neurobiolaging.2011.02.013>
- Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., & Weiskopf, N. (2010). Optimization and validation of methods for mapping of the radio-frequency transmit field at 3T. *Magnetic Resonance in Medicine*, 64(1), 229–238. <https://doi.org/10.1002/mrm.22421>
- Lutti, A., Stadler, J., Josephs, O., Windischberger, C., Speck, O., Bernarding, J., Hutton, C., & Weiskopf, N. (2012). Robust and fast whole brain mapping of the RF transmit field B1 at 7T. *PLoS One*, 7(3), 1–7. <https://doi.org/10.1371/journal.pone.0032379>
- Maclaren, J., Armstrong, B. S. R., Barrows, R. T., Danishad, K. A., Ernst, T., Foster, C. L., Gumus, K., Herbst, M., Kadashevich, I. Y., Kusik, T. P., Li, Q., Lovell-Smith, C., Prieto, T., Schulze, P., Speck, O., Stucht, D., & Zaitsev, M. (2012). Measurement and correction of microscopic head motion during magnetic resonance imaging of the brain. *PLoS One*, 7(11), 3–11. <https://doi.org/10.1371/journal.pone.0048088>
- Martin, W. R. W., Wieler, M., & Gee, M. (2008). Midbrain iron content in early Parkinson disease: A potential biomarker of disease status. *Neurology*, 70(16), 1411–1417. <https://doi.org/10.1212/01.wnl.0000286384.31050.b5>
- May, A. (2011). Experience-dependent structural plasticity in the adult human brain. *Trends in Cognitive Sciences*, 15(10), 475–482. <https://doi.org/10.1016/j.tics.2011.08.002>
- Neeb, H., Ermer, V., Stocker, T., & Shah, N. J. (2008). Fast quantitative mapping of absolute water content with full brain coverage. *NeuroImage*, 42(3), 1094–1109. <https://doi.org/10.1016/j.neuroimage.2008.03.060>
- Noble, S., Scheinost, D., & Constable, R. T. (2020). A guide to the measurement and interpretation of fMRI test-retest reliability. *Current Opinion in Behavioral Sciences*, 40, 27–32. <https://doi.org/10.1016/j.cobeha.2020.12.012>

- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Papazoglou, S., Streubel, T., Ashtarayeh, M., Pine, K. J., Edwards, L. J., Brammerloh, M., Kirilina, E., Morawski, M., Jäger, C., Geyer, S., Callaghan, M. F., Weiskopf, N., & Mohammadi, S. (2019). Biophysically motivated efficient estimation of the spatially isotropic  $R_2^*$  component from a single gradient-recalled echo measurement. *Magnetic Resonance in Medicine*, 82(5), 1804–1811. <https://doi.org/10.1002/mrm.27863>
- Péran, P., Cherubini, A., Luccichenti, G., Hagberg, G., Démonet, J. F., Rascol, O., Celsis, P., Caltagirone, C., Spalletta, G., & Sabatini, U. (2009). Volume and iron content in basal ganglia and thalamus. *Human Brain Mapping*, 30(8), 2667–2675. <https://doi.org/10.1002/hbm.20698>
- Péran, P., Hagberg, G., Luccichenti, G., Cherubini, A., Brainovich, V., Celsis, P., Caltagirone, C., & Sabatini, U. (2007). Voxel-based analysis of  $R_2^*$  maps in the healthy human brain. *Journal of Magnetic Resonance Imaging*, 26(6), 1413–1420. <https://doi.org/10.1002/jmri.21204>
- Preibisch, C., & Deichmann, R. (2009). Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles. *Magnetic Resonance in Medicine*, 61(1), 125–135. <https://doi.org/10.1002/mrm.21776>
- Raz, N., & Rodrigue, K. M. (2006). Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neuroscience and Biobehavioral Reviews*, 30(6), 730–748. <https://doi.org/10.1016/j.neubiorev.2006.07.001>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Rowley, C. D., Campbell, J. S. W., Wu, Z., Leppert, I. R., Rudko, D. A., Pike, G. B., & Tardif, C. L. (2021). A model-based framework for correcting B1+ inhomogeneity effects in magnetization transfer saturation and inhomogeneous magnetization transfer saturation maps. *Magnetic Resonance in Medicine*, 86(4), 2192–2207. <https://doi.org/10.1002/mrm.28831>
- Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, 2, 1–10. <https://doi.org/10.3389/fnagi.2010.00027>
- Schyboll, F., Jaekel, U., Petruccione, F., & Neeb, H. (2020). Origin of orientation-dependent  $R_1$  ( $=1/T_1$ ) relaxation in white matter. *Magnetic Resonance in Medicine*, 84(5), 2713–2723. <https://doi.org/10.1002/mrm.28277>
- Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., & Mohammadi, S. (2019). HMRI – A toolbox for quantitative MRI in neuroscience and clinical research. *NeuroImage*, 194(January), 191–210. <https://doi.org/10.1016/j.neuroimage.2019.01.029>
- Tofts, P. (2003). *Quantitative MRI of the brain: Measuring changes caused by disease*. John Wiley and Sons. <https://doi.org/10.1002/0470869526>
- Versluis, M. J., Peeters, J. M., van Rooden, S., van der Grond, J., van Buchem, M. A., Webb, A. G., & van Osch, M. J. P. (2010). Origin and reduction of motion and f0 artifacts in high resolution T2\*-weighted magnetic resonance imaging: Application in Alzheimer's disease patients. *NeuroImage*, 51(3), 1082–1088. <https://doi.org/10.1016/j.neuroimage.2010.03.048>
- Volz, S., Nöth, U., Jurcoane, A., Ziemann, U., Hattingen, E., & Deichmann, R. (2012). Quantitative proton density mapping: Correcting the receiver sensitivity bias via pseudo proton densities. *NeuroImage*, 63(1), 540–552. <https://doi.org/10.1016/j.neuroimage.2012.06.076>
- von Oertzen, T., Brandmaier, A. M., & Tsang, S. (2015). Structural equation modeling with  $\Omega$ nyx. *Structural Equation Modeling*, 22(1), 148–161. <https://doi.org/10.1080/10705511.2014.935842>
- Wechsler, D. (1981). *Manual for the Wechsler adults intelligence scale—Revised*. Psychological Corporation.
- Weiskopf, N., Callaghan, M. F., Josephs, O., Lutti, A., & Mohammadi, S. (2014). Estimating the apparent transverse relaxation time ( $R_2^*$ ) from images with different contrasts (ESTATICS) reduces motion artifacts. *Frontiers in Neuroscience*, 8, 1–10. <https://doi.org/10.3389/fnins.2014.00278>
- Weiskopf, N., Edwards, L. J., Helms, G., Mohammadi, S., & Kirilina, E. (2021). Quantitative magnetic resonance imaging of brain anatomy and in vivo histology. *Nature Reviews Physics*, 3(8), 570–588. <https://doi.org/10.1038/s42254-021-00326-1>
- Weiskopf, N., Hutton, C., Josephs, O., Turner, R., & Deichmann, R. (2007). Optimized EPI for fMRI studies of the orbitofrontal cortex: Compensation of susceptibility-induced gradients in the readout direction. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 20(1), 39–49. <https://doi.org/10.1007/s10334-006-0067-6>
- Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., & Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *NeuroImage*, 54(3), 2116–2124. <https://doi.org/10.1016/j.neuroimage.2010.10.023>
- Weiskopf, N., Suckling, J., Williams, G., Correia, M., Inkster, B., Tait, R., Ooi, C., Bullmore, E. T., & Lutti, A. (2013). Quantitative multiparameter mapping of  $R_1$ ,  $PD^*$ ,  $MT$ , and  $R_2^*$  at 3T: A multi-center validation. *Frontiers in Neuroscience*, 7, 1–11. <https://doi.org/10.3389/fnins.2013.00095>
- Wharton, S., & Bowtell, R. (2012). Fiber orientation-dependent white matter contrast in gradient echo MRI. *Proceedings of the National Academy of Sciences of the United States of America*, 109(45), 18559–18564. <https://doi.org/10.1073/pnas.1211075109>
- Wright, P. J., Mougín, O. E., Totman, J. J., Peters, A. M., Brookes, M. J., Coxon, R., Morris, P. E., Clemence, M., Francis, S. T., Bowtell, R. W., & Gowland, P. A. (2008). Water proton T1 measurements in brain tissue at 7, 3, and 1.5T using IR-EPI, IR-TSE, and MPRAGE: Results and optimization. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 21(1–2), 121–130. <https://doi.org/10.1007/s10334-008-0104-8>
- Yarnykh, V. L. (2010). Optimal radiofrequency and gradient spoiling for improved accuracy of T1 and B1 measurements using fast steady-state techniques. *Magnetic Resonance in Medicine*, 63(6), 1610–1626. <https://doi.org/10.1002/mrm.22394>

**How to cite this article:** Wenger, E., Polk, S. E., Kleemeyer, M. M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for magnetization transfer and proton density but attenuated for  $R_1$  and  $R_2^*$  in healthy young adults. *Human Brain Mapping*, 43(11), 3585–3603. <https://doi.org/10.1002/hbm.25870>



**Polk, S. E.**, Kleemeyer, M. M., Köhncke, Y., Brandmaier, A. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Wenger, E., & Düzel, S. (2022). Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2022.852737>





# Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise

## OPEN ACCESS

### Edited by:

Joshua Oon Soo Goh,  
National Taiwan University, Taiwan

### Reviewed by:

Yuka Kotozaki,  
Iwate Medical University, Japan  
Gerard Nisal Bischof,  
Julich Research Center, Helmholtz  
Association of German Research  
Centres (HZ), Germany  
Li-Wei Kuo,  
National Health Research Institutes,  
Taiwan

### \*Correspondence:

Sarah E. Polk  
spolk@mpib-berlin.mpg.de

†These authors share senior  
authorship

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 11 January 2022

**Accepted:** 15 March 2022

**Published:** 17 May 2022

### Citation:

Polk SE, Kleemeyer MM,  
Köhncke Y, Brandmaier AM,  
Bodammer NC, Misgeld C, Porst J,  
Wolffarth B, Kühn S, Lindenberger U,  
Wenger E and Düzel S (2022) Change  
in Latent Gray-Matter Structural  
Integrity Is Associated With Change in  
Cardiovascular Fitness in Older Adults  
Who Engage in At-Home Aerobic  
Exercise.  
*Front. Hum. Neurosci.* 16:852737.  
doi: 10.3389/fnhum.2022.852737

Sarah E. Polk<sup>1,2\*</sup>, Maike M. Kleemeyer<sup>1</sup>, Ylva Köhncke<sup>1</sup>, Andreas M. Brandmaier<sup>1,3,4</sup>,  
Nils C. Bodammer<sup>1</sup>, Carola Misgeld<sup>5</sup>, Johanna Porst<sup>5</sup>, Bernd Wolffarth<sup>5</sup>, Simone Kühn<sup>6</sup>,  
Ulman Lindenberger<sup>1,3</sup>, Elisabeth Wenger<sup>1†</sup> and Sandra Düzel<sup>1†</sup>

<sup>1</sup> Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany, <sup>2</sup> International Max Planck Research School on the Life Course (LIFE), Berlin, Germany, <sup>3</sup> Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany, <sup>4</sup> Department of Psychology, MSB Medical School Berlin, Berlin, Germany, <sup>5</sup> Department of Sports Medicine, Charité – Universitätsmedizin Berlin, Humboldt Universität zu Berlin, Berlin, Germany, <sup>6</sup> Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, Berlin, Germany

In aging humans, aerobic exercise interventions have been found to be associated with more positive or less negative changes in frontal and temporal brain areas, such as the anterior cingulate cortex (ACC) and hippocampus, relative to no-exercise control conditions. However, individual measures such as gray-matter (GM) probability may afford less reliable and valid conclusions about maintenance or losses in structural brain integrity than a latent construct based on multiple indicators. Here, we established a latent factor of GM structural integrity based on GM probability assessed by voxel-based morphometry, magnetization transfer saturation, and mean diffusivity. Based on this latent factor, we investigated changes in structural brain integrity during a six-month exercise intervention in brain regions previously reported in studies using volumetric approaches. Seventy-five healthy, previously sedentary older adults aged 63–76 years completed an at-home intervention study in either an exercise group (EG;  $n = 40$ ) or in an active control group (ACG;  $n = 35$ ). Measures of peak oxygen uptake ( $VO_{2peak}$ ) taken before and after the intervention revealed a time-by-group interaction, with positive average change in the EG and no reliable mean change in the ACG. Significant group differences in structural brain integrity changes were observed in the right and left ACC, right posterior cingulate cortex (PCC), and left juxtapositional lobule cortex (JLC). In all instances, average changes in the EG did not differ reliably from zero, whereas average changes in the ACG were negative, pointing to maintenance of structural brain integrity in the EG, and to losses in the ACG. Significant individual differences in change were observed for right ACC and left JLC. Following up on these differences, we found that exercising participants with greater fitness gains



also showed more positive changes in structural integrity. We discuss the benefits and limitations of a latent-factor approach to changes in structural brain integrity, and conclude that aerobic fitness interventions are likely to contribute to brain maintenance in old age.

**Keywords:** physical activity, fitness, brain structure integrity, aging, older adults, structural equation modeling

## INTRODUCTION

As a result of progress in global health and development, the worldwide population of individuals over 60 years old is projected to double by 2050 (HelpAge International, 2018). However, as the proportion of older adults in the population increases, so does the concern of health-related changes associated with advancing adult age. Brain health is of great societal importance, as the risk of cognitive impairments leading to dementia increases with age, and individual differences in the degree to which brain structure, function, and neurochemistry can be maintained into old age are hypothesized to predict individual differences in cognitive functioning among older adults (Nyberg et al., 2012; Lindenberger, 2014; Cabeza et al., 2018; Nyberg and Pudas, 2019; Nyberg and Lindenberger, 2020; Johansson et al., 2022).

Aerobic exercise shows promise as a modifiable lifestyle factor that may potentially promote brain maintenance in old age. Intervention studies focusing on volumetric characteristics of brain structure have found that gray-matter (GM) volume shows more positive changes in exercisers than control participants in frontal areas, such as the anterior cingulate cortex (ACC), and in temporal areas, such as the hippocampus (Colcombe et al., 2004; Erickson et al., 2011). Higher levels of and more positive changes in physical fitness have also been associated with greater GM volume and attenuation of volume loss in prefrontal, parietal, and temporal regions, including the hippocampus (Colcombe et al., 2003; Weinstein et al., 2012; Maass et al., 2015; Kleemeyer et al., 2016).

However, individual measures such as GM probability may afford less valid conclusions about maintenance or losses in structural brain integrity on a generalized level than a latent construct based on multiple indicators. Latent constructs express the variance shared by multiple measures, thereby separating common variance from specific variance and measurement error (Wansbeek and Meijer, 2001). Therefore, the use of latent factors can improve the estimation of associations among constructs of interest. Based on pioneering work by Kühn et al. (2017), Köhncke et al. (2021) introduced a multi-trait multi-method model using structural equation modeling (SEM) capturing the shared variance between GM volume, mean diffusivity (MD), and magnetization transfer (MT) ratio in a latent factor of GM structural integrity for several regions of the brain. The authors were able to show that, in a cross-sectional sample, older participants generally showed lower scores on these integrity factors, which they interpreted as a reflection of age-related deterioration of overall GM, in line with studies focusing on single indicators

(Raz et al., 2005; Fjell and Walhovd, 2010; Grydeland et al., 2013; Seiler et al., 2014). In addition, GM structural integrity correlated positively with episodic memory performance (Köhncke et al., 2021).

In the current intervention study, we adapted this cross-sectional model of GM structural integrity to a longitudinal context in order to measure change in structural integrity as a latent factor in brain regions of interest previously reported in studies using volumetric approaches. The three indicators used in the current model were all measured using magnetic resonance imaging (MRI) and captured different characteristics of GM structural integrity. GM volume was calculated using T<sub>1</sub>-weighted images and voxel-based morphometry (VBM; Ashburner and Friston, 2000; Good et al., 2001), which estimates GM concentrations at each voxel based on signal intensity. MT saturation, an improvement to the MT ratio measure, which is affected by spatial variations of the transmit field for excitation and the local T<sub>1</sub> relaxation (Helms et al., 2008b), was used in the current models. MT maps quantify the transfer of magnetization between tissue water and protons bound to macromolecules, and can be used to assess microstructural changes to GM, with lower MT values correlating with demyelination and axonal loss (see Seiler et al., 2014). MD, measured using diffusion-weighted imaging (DWI), estimates the rate of water diffusion in each voxel (Pierpaoli and Basser, 1996), and is commonly used as a measure of white matter integrity, but can also be used as an index of GM density, with greater MD values corresponding to lower tissue density, likely reflecting demyelination (Song et al., 2005) and lower axon fiber density (Beaulieu, 2002; Fukutomi et al., 2019). By combining these three imaging modalities, we established a latent factor of GM structural integrity representing the commonalities across these three measures of brain structure. We use this novel longitudinal modeling approach to look at exercise-induced changes in a latent factor that represents general structural integrity, focusing on the shared variance of GM volume, MT, and MD, rather than any one of these alone, thereby removing any modality-specific measurement error. Given this emphasis on the commonalities across the individual modalities and to acknowledge the level of abstractness of this latent factor, we refer to this factor as “integrity.”

## The AKTIV Study

Previous studies investigating the effects of exercise among older adults generally have been conducted in a laboratory setting (e.g., Colcombe et al., 2006; Erickson et al., 2011), which may not reflect the exercise opportunities accessible to most older adults in their everyday lives. To examine whether engaging in aerobic exercise



at home may also benefit older adults in terms of both fitness and brain health, the “Aktives Altern für Körper und Geist” (active aging for body and mind) study (AKTIV) implemented a personalized at-home physical exercise regime.

In the current analyses, we first investigated whether six months of moderate, at-home aerobic exercise could effectively boost cardiovascular fitness in older adults. We then validated the GM structural integrity model in a longitudinal manner in 12 regions of interest (ROIs) in the frontal, midline, and temporal areas selected based on previous publications (Colcombe et al., 2006; Erickson et al., 2011), and examined whether the group who exercised showed either gains (increase over controls) or maintenance (attenuation of loss relative to controls) in GM structural integrity within these regions. Finally, we investigated the association between changes in cardiovascular fitness and changes in structural brain integrity, with the hypothesis that greater increases in fitness should be positively associated with greater increases (or smaller losses) in GM structural integrity.

## MATERIALS AND METHODS

### Sample and Study Design

In the current analyses, we focused on the effects of physical training on GM structural integrity by comparing a group of individuals who regularly engaged in aerobic exercise with a group of sedentary individuals. These two groups were drawn from the larger AKTIV study, which investigated the effects of cognitive and physical training in comparison to an active control group (ACG). Here, we explored the effects of physical training alone vs. no physical training (see Interventions for details) with a sub-sample of 75 healthy, previously sedentary adults aged 63–76 years.

A full description of the AKTIV study design and methods can be found in Wenger et al. (2022); for convenience, we describe relevant materials and methods here. Volunteers for the AKTIV study were recruited through a participant data bank with participants from earlier, unrelated studies, and newspaper advertisements. A telephone screening was conducted to exclude individuals if they met any of the following criteria: MRI contraindications; inability to meet the time requirements of the study; not right-handed; younger than 63 or older than 78 years old at the start of the study; already engaging in aerobic exercise more than once every 2 weeks; fluent in a language other than German or English, or fluent in more than two languages; receiving medical treatment for Parkinson's, gout, rheumatism, heart attack, stroke, cancer, severe back problems, severe arrhythmia, severe chronic liver or kidney failure, severe disease of the hematopoietic system, mental illness (e.g., depression), or neurological disease (e.g., epilepsy, brain tumor). The 201 volunteers who did not meet the exclusion criteria were randomly assigned (with the exception of couples who were jointly assigned so that participants would remain blind to other groups) to one of four groups: (1) an ACG, (2) a language training group, (3) a physical exercise training group, or (4) a combined language and physical exercise

training group. Next, potential participants were invited to a physical assessment including cardiopulmonary exercise testing (CPET) with lactate diagnostics at the Department of Sports Medicine at the Charité – Berlin University of Medicine. Based on this exam, a further 22 volunteers dropped out or were excluded due to existing medical conditions. Finally, participants underwent an initial MRI session before beginning their assigned training [pre-intervention, time point 1 (T1)]. Nineteen participants dropped out before the training started due to disinterest and one additional participant was excluded due to claustrophobia. Thus, the effective initial sample consisted of 159 individuals.

After 3 months of training [mid-intervention, time point 2 (T2)], MRI acquisitions were repeated, consisting of the same scans. After a total of 6 months of training [post-intervention, time point 3 (T3)], MR measures were acquired once more, and participants again underwent CPET at the Charité.

During the intervention, 17 participants dropped out due to physical complaints (e.g., knee or back pain during training), disinterest, time constraints, or unspecified reasons.

The Ethics Committee of the German Psychological Society (DGPs) approved the study and written informed consent was collected from all participants.

### Interventions

As previously stated, we focused on the effects of physical exercise and therefore only used the data of participants who completed the intervention in the exercise-only group or the ACG.

Participants in the exercise group (EG;  $n = 40$  completed, mean age = 69.8 years, 50% females) engaged in moderate at-home exercise three to four times per week at any time of the day using a bicycle ergometer (DKN Ergometer AM-50) and a personalized interval training regime programmed onto tablets (Lenovo TB2-X30L TAB) that were synced to the ergometers *via* Bluetooth. Tablets were equipped with SIM cards so that data could be uploaded to the study server whenever an Internet connection was available. The training initially lasted 30 min at an individually set intensity (25–140 W, mean = 67.9, SD = 26.65). After completing each training session, participants could indicate if they found the training too easy or too difficult, and the intensity could be remotely adjusted accordingly. In this way, the training was highly personalized so that participants would not be discouraged by an exceedingly easy/difficult exercise program. Approximately every two weeks, difficulty was automatically increased by 3 min per interval (up to 56 min total) and 3–4 W. Participants in the EG were also instructed to read pre-selected literature on the tablet at a slow pace for an additional 15 min on days when they trained and for 45 min on the other days, so that participants in both groups would engage in approximately 45 min of study-related activity on at least six days per week. Finally, those in the EG participated in weekly 1-h group sessions (5–10 participants per session) at the institute consisting of toning and stretching, led by an external instructor.

Participants in the ACG ( $n = 35$  completed, mean age = 70.7 years, 40% females) also received a tablet and were instructed to read pre-selected literature for 45 min daily

for at least six days per week. These participants also attended weekly group sessions during which they discussed literary excerpts led by external facilitators.<sup>1</sup>

Compliance was defined as engaging in at least an average of 90 min of group-relevant activity (reading or exercise) per week over 21 weeks ( $\geq 1890$  min;  $n_{non-compliant}$  in ACG = 3,  $n_{non-compliant}$  in EG = 4) with no pauses of longer than 2 weeks ( $n_{non-compliant}$  in EG = 1). Participants in the EG also needed to exercise at a steady or slightly increasing intensity (based on Watts) over the duration of the intervention, meaning those with decreasing Wattage were counted as non-compliant ( $n_{non-compliant}$  in EG = 5).

Finally, regarding sample size, *post hoc* sensitivity analyses conducted in G\*Power (version 3.1.9.6; Faul et al., 2007, 2009) indicated that, with  $\alpha = 0.05$ ,  $1 - \beta = 0.95$ , and a study design with two groups and three time points, a sample size of  $N = 75$  could reliably capture time-by-group interaction effects with an effect size of  $f \geq 0.19$  and correlations with a coefficient of  $r \geq 0.367$ .

## Data Acquisition

### Cardiovascular Fitness

Cardiovascular fitness, indexed by peak oxygen uptake ( $VO_{2peak}$ ; measured at 30-s intervals, relativized to body weight in kg), was assessed at the physical assessments at pre- and post-intervention using CPET. This was conducted under the supervision of the overseeing physician using a bicycle ergometer (Ergoselect 100k, Ergoline GmbH, Bitz, Germany) using the Quark Clinical-based Metabolic Cart with the standard Breath-by-Breath setup, and the V2 Mask (Hans Rudolph, Inc.), which covers the mouth and nose, and is fastened to the back of the head. Participants were instructed to pedal at a constant rate of 60–70 rotations per minute during the entire protocol, which consisted of a 3-min rest phase, an exertion phase with a starting resistance of 20 W, which increased by 20 W every 3 min until participants reported they had reached maximum exertion, and a 5-min recovery phase with no resistance.

## Magnetic Resonance Imaging

### Acquisition

Participants were scanned pre-, mid-, and post-intervention. MR images were acquired using a 3T Magnetom Tim Trio MRI scanner system (Siemens Medical Systems, Erlangen) using a 32-channel radiofrequency head coil.  $T_1$ -weighted images were obtained using a 3D  $T_1$ -weighted magnetization prepared gradient-echo (MPRAGE) sequence. The multi-parameter mapping protocol used to acquire the MT maps comprised one static magnetic ( $B_0$ ) gradient echo (GRE)-field map, one radiofrequency (RF) transmit field map ( $B_1$ ), and three multi-echo 3D fast low angle shot (FLASH) scans (Helms et al., 2008b; Tabelow et al., 2019). Diffusion-weighted images were obtained with a single-shot diffusion-weighted spin-echo-refocused echo-planar imaging sequence. Further details regarding the acquisition parameters can be found in the **Supplementary Material**.

<sup>1</sup><http://shared-reading.de/>

## Preprocessing

Structural  $T_1$ -weighted images were preprocessed using the Computational Anatomy Toolbox 12 (CAT12, Structural Brain Mapping group, Jena University Hospital; Gaser and Dahnke, 2016) in Statistical Parametric Mapping (SPM12, Institute of Neurology<sup>2</sup>) using the default parameters of the longitudinal pipeline. Estimation of MT maps was conducted in SPM12 using an adapted longitudinal pipeline with the hMRI toolbox (Tabelow et al., 2019<sup>3</sup>). DW images were preprocessed using MRtrix (version 3.0\_RC3; Tournier et al., 2019), FSL (FMRIB's Software Library, version 6.0.2; Smith et al., 2004; Woolrich et al., 2009; Jenkinson et al., 2012), and ANTS (version 2.2.0; Avants et al., 2010, 2011), following the Basic and Advanced Tractography with MRtrix for All Neurophiles (B.A.T.M.A.N.) tutorial (Tahedl, 2018). Further details regarding preprocessing can be found in the **Supplementary Material**.

## Regions of Interest

Regions of interest were selected based on previous intervention studies on the effects of aerobic exercise on brain structure (Colcombe et al., 2006; Erickson et al., 2011). Mean values of VBM, MT, and MD were extracted from the left and right hippocampus, ACC, posterior cingulate cortex (PCC), precentral gyrus (PCG), juxtapositional lobule cortex [JLC, previously supplementary motor cortex (SMA)], and pars triangularis and pars opercularis combined as inferior frontal gyrus (IFG), as defined by the Harvard-Oxford cortical and subcortical structural atlases<sup>4,5</sup> (Desikan et al., 2006), for a total of 12 ROIs. VBM and MT maps were calculated such that each participant's resulting map was in MNI space, so ROI masks were simply resized using the *Coregister and reslice* module in SPM12 with the first participant's map at T1 as a reference, and *fslstats* in FSL was used to extract the mean and SD across all non-zero voxels within each ROI. As calculation of MD maps resulted in images in each participant's native space, ROI masks were first transformed into native space for each participant using *applywarp* in FSL with the transformations generated during preprocessing and used to transform the non-diffusion-weighted images into native space, then *fslstats* was used to extract means and SDs across all non-zero voxels.

VBM values from each ROI were adjusted for intracranial volume [ICV, calculated using the *Estimate TIV and global tissue volumes* module in CAT12] using the analysis of covariance formula from Raz et al. (2005): adjusted volume = raw volume -  $b \times (ICV - \text{mean ICV})$ , where  $b$  is the slope of regression of GM probability in the relevant ROI on ICV.

## Statistical Analyses

All statistical analyses were conducted using R (R Core Team, 2021), version 4.1.2 (2021-11-01), in RStudio (RStudio Team, 2021), version 2021.09.1.

<sup>2</sup>[www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)

<sup>3</sup><https://hmri-group.github.io/hMRI-toolbox/>

<sup>4</sup><https://neurovault.org/images/1705/>

<sup>5</sup><https://neurovault.org/images/1700/>

## Univariate and Multivariate Outlier Detection

Univariate outliers within each measure and time point were defined as data points further than 4 SD away from the mean, resulting in one data point being discarded (hippocampus MT at T3). Multivariate outlier detection was conducted within measures across all three time points (two time points for VO<sub>2</sub>peak) for complete cases using *robustMD* from the *faoutlier* R package (Chalmers and Flora, 2015) with the classical product-moment method (criterion = 0.001). In this way, we looked for abnormal patterns across time points within our sample (e.g., one data point having a much higher value than the other two), and removed all three data points of such outliers. This resulted in no cases being discarded for VO<sub>2</sub>peak, and 17 out of 2052 cases being discarded across the three MRI modalities and 12 ROIs in participants with all three observations.

## Structural Equation Modeling

SEM was used to investigate differences in group means of change in cardiovascular fitness indexed by VO<sub>2</sub>peak and latent GM integrity, as well as the relationship between these two changes using a multivariate, multigroup approach. Models were specified and estimated using the *OpenMx* R package (Pritikin et al., 2015; Neale et al., 2016; Hunter, 2018; Boker et al., 2021), version 2.19.8. Full information maximum likelihood (FIML) was used to account for missing data without the need for case-wise deletion. Observed variables were rescaled to have a mean of 0 and a SD of 1 longitudinally (i.e., data from all time points were first stacked then rescaled) to preserve the relative mean differences between the same indicators measured at different time points. The root mean square error of approximation (RMSEA) and the comparative fit index (CFI) were used to evaluate model fit, using rough thresholds of RMSEA < 0.08 and CFI > 0.90 to indicate acceptable model fit (Schermelleh-Engel et al., 2003).

Likelihood-ratio tests (LRTs) were used to determine the statistical significance of group differences on individual parameters, as well as of certain parameter estimates within a model. To conduct an LRT, a model in which the parameter(s) of interest is freely estimated is compared to a nested model in which this parameter is fixed (e.g., to 0 or equal across groups). The difference in  $\chi^2$  (i.e., the likelihood ratio) between the two models indicates the difference in fit, and if this difference is significant, the null hypothesis that the models fit equally well can be rejected (Kline, 2016). Given previous studies on the effects of aging and exercise on brain volume and structural integrity (Colcombe et al., 2006; Erickson et al., 2011; Köhncke et al., 2021), we had strong hypotheses that exercise would be beneficial to structural integrity of the brain. That is, those participants who exercised should show gains in or maintenance of structural integrity, while those who did not exercise would show declines in integrity. Therefore, unless otherwise indicated, one-sided hypothesis testing was conducted looking for changes in the positive direction within the EG and in the negative direction within the ACG.

Latent change score models (LCSMs) were used to evaluate group mean differences in change in VO<sub>2</sub>peak and GM integrity

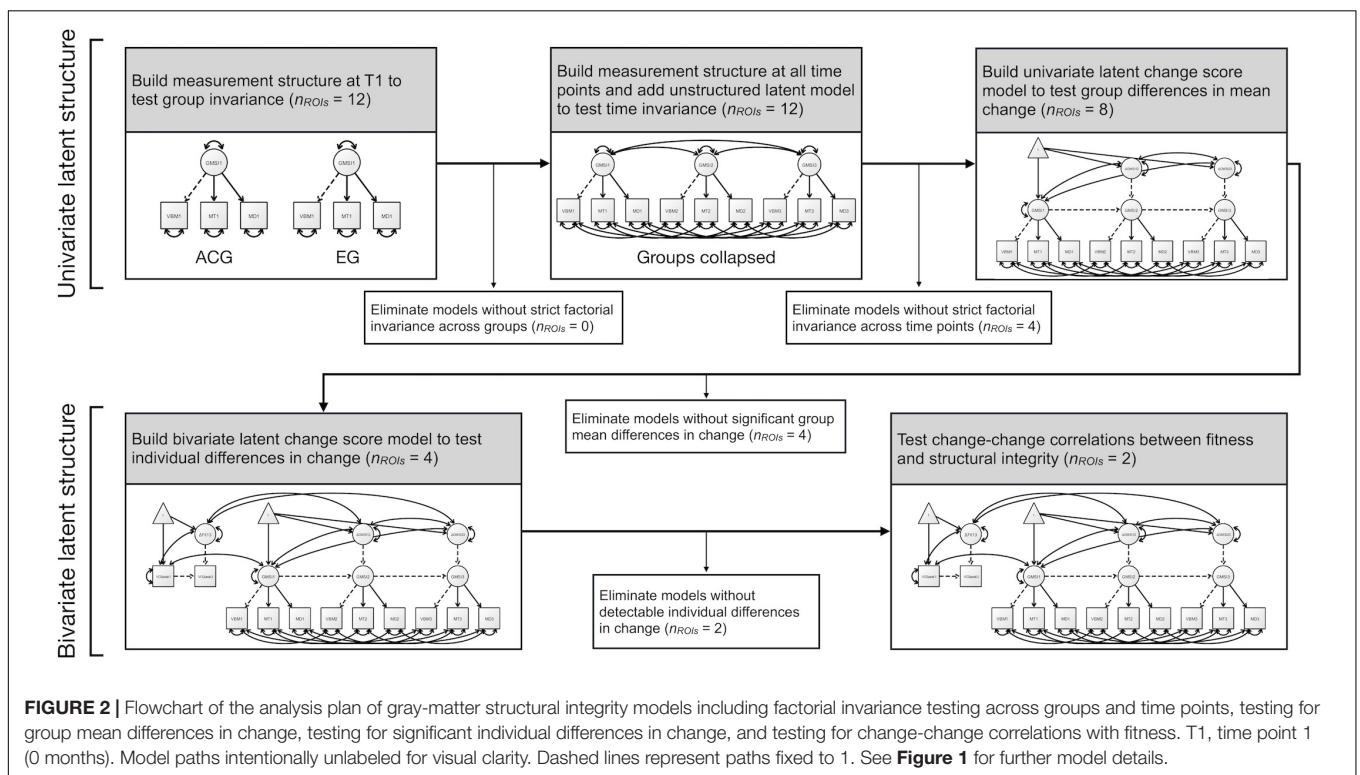
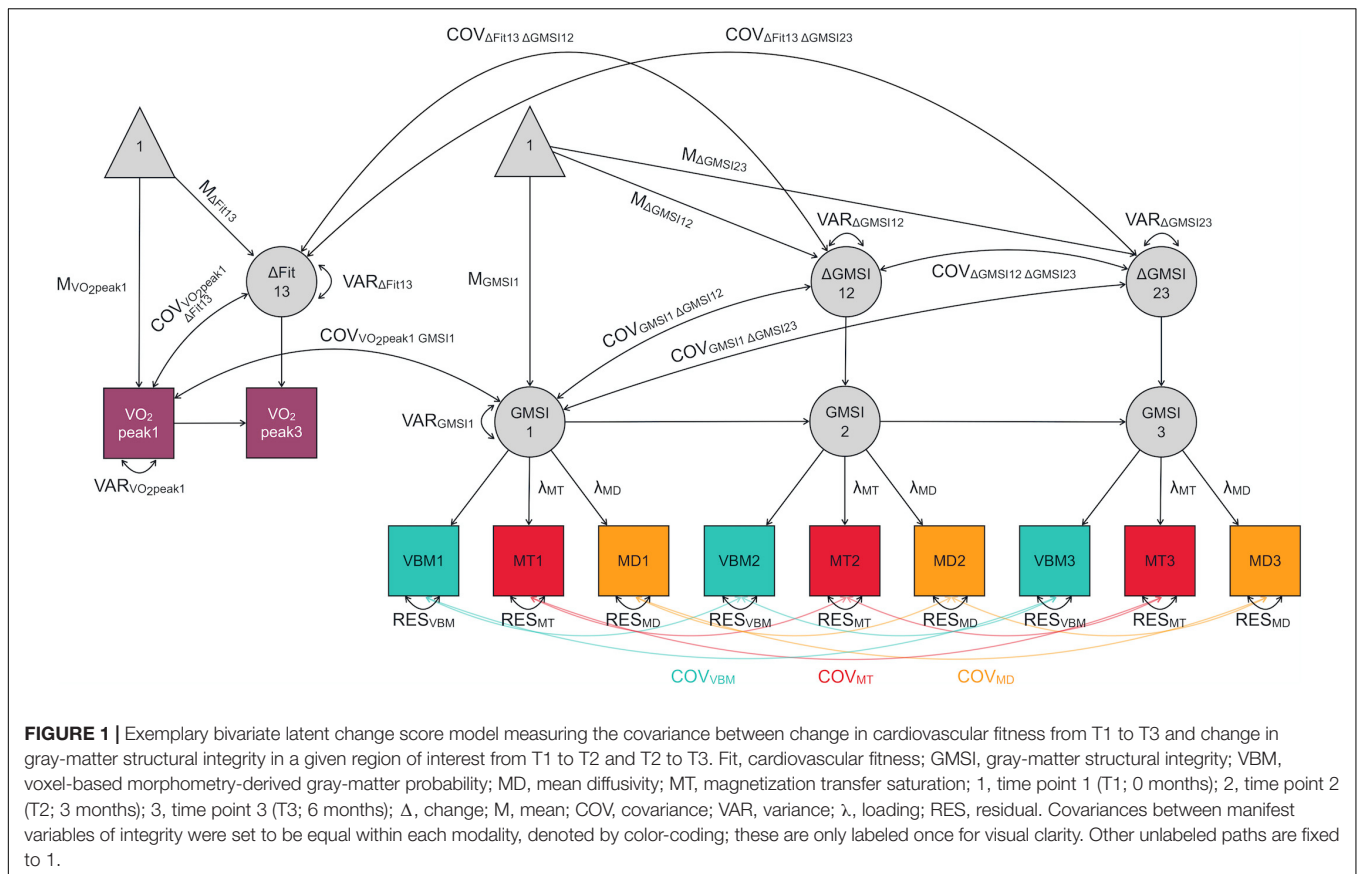
following the tutorial by Kievit et al. (2018). Full details of the model validation can be found in the **Supplementary Material**, but we describe the models briefly in the following. A univariate LCSM was built to measure mean change in VO<sub>2</sub>peak; a pseudo-latent factor,  $\Delta$ VO<sub>2</sub>peak, captured the difference in VO<sub>2</sub>peak between T1 and T3. Multigroup models were used to test for differences in change between the two groups in cardiovascular fitness. Multivariate LCSMs with three time points were built to measure mean change in GM structural integrity from T1 to T2 as well as from T2 to T3 in each ROI individually. As is common practice in SEM, factorial invariance testing was performed to establish whether the measurement structure of the structural integrity models held across groups and time points. In those regions that survived invariance testing, mean differences in change between the two groups were investigated. Finally, in those models showing detectable individual differences in change as well as significant mean differences in change between groups, change-change correlations between fitness and GM structural integrity were examined using bivariate LCSMs (see **Figure 1**) both in the whole sample as well as separately for the two groups. This strategy was chosen to add credibility to a causal interpretation of change-change associations in the exercise group (see Ghisletta and Lindenberger, 2004). For an overview of the analysis plan of GM structural integrity models, see **Figure 2**.

## RESULTS

A description of the sample can be found in **Table 1**. Participants who did not meet compliance criteria were excluded from T2 and T3 ( $n_{ACG} = 3$ ,  $n_{EG} = 10$ ) but were kept in at T1 under the assumption that they did not differ from other participants at baseline. One further exercise participant was excluded from T2 and T3 due to technical difficulties.

### Cardiovascular Fitness

A univariate LCSM with one pseudo-latent difference score is exactly identified, therefore fit indices are perfect by definition. Cardiovascular fitness as indexed by VO<sub>2</sub>peak showed a significantly positive mean change in the EG, unstandardized estimate of the mean ( $b_0$ ) = 0.419 (i.e., average in increase of 0.419 mL/kg/min from T1 to T3), standard error (SE) = 0.089,  $\Delta\chi^2(df = 1) = 18.67$ ,  $p < 0.001$ , whereas mean change in the ACG was not significant,  $b_0 = 0.123$ , SE = 0.088,  $\Delta\chi^2(df = 1) = 1.92$ ,  $p > 0.050$ . Furthermore, the EG showed significantly greater mean change than the ACG,  $\Delta\chi^2(df = 1) = 5.37$ ,  $p = 0.010$  (see **Figures 3, 4**). In terms of percent change, the EG showed a mean change of 12.8% (SE = 2.28), while the ACG showed a mean change of 3.7% (SE = 2.22). A significant difference in VO<sub>2</sub>peak between males and females was seen at baseline,  $t(70.3) = 3.95$ ,  $p < 0.001$ , with males showing higher VO<sub>2</sub>peak (mean = 25.5, SD = 5.99) than females ( $M = 20.7$ , SD = 4.46). Males and females did not differ in percent change in VO<sub>2</sub>peak, and no associations of age or years of education were seen with either baseline or percent change in VO<sub>2</sub>peak.

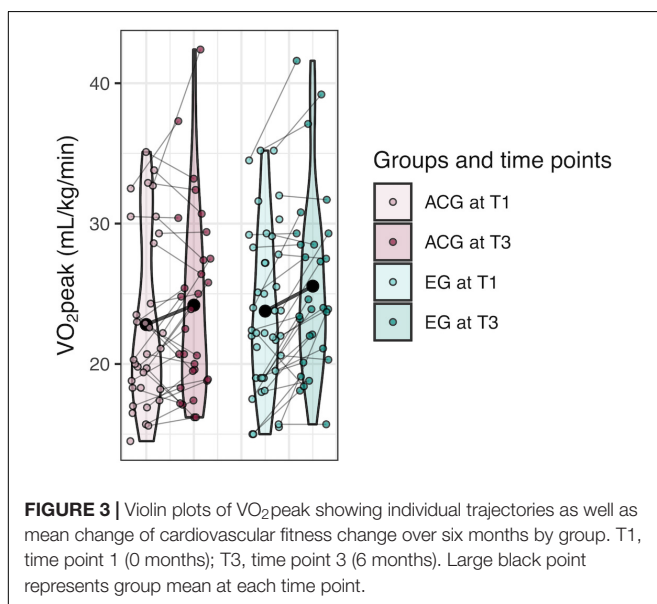




**TABLE 1** | Sample demographics and intervention specifics.

	Active control group	Exercise group
<i>n</i> completed intervention	35	40
<i>n</i> fully complied	32	29
Age at baseline, M/SD (range)	70.7/3.81 (64.0–76.0)	69.8/3.49 (63.9–76.9)
Sex, % of female participants	40.0	50.0
Years of education, M/SD (range)	13.4/3.27 (7–16)	13.2/3.02 (7–16)
DSST score at baseline, M/SD (range)	47.3/9.42 (21–68)	45.2/10.90 (21–75)
Total minutes spent in intervention, M/SD (range)	4772/1819.6 (2505–10,858)	6554/1222.9 (4098–9790)
Minutes spent reading, M/SD (range)	4772/1819.6 (2505–10,858)	3381/1143.3 (915–5855)
Minutes spent exercising, M/SD (range)	–	3173/410.4 (2556–3937)

*M*, mean; *SD*, standard deviation; *DSST*, Digit Symbol Substitution Test. *Ms* and *SDs* of age at baseline, sex, years of education, and *DSST* score at baseline are calculated within participants who completed the intervention. Total minutes spent in intervention are calculated within participants who fully complied to the intervention.



**FIGURE 3** | Violin plots of  $VO_{2peak}$  showing individual trajectories as well as mean change of cardiovascular fitness change over six months by group. T1, time point 1 (0 months); T3, time point 3 (6 months). Large black point represents group mean at each time point.

## Gray-Matter Structural Integrity

Out of the 12 initial GM structural integrity models, eight survived testing for factorial invariance, which implies that the factor structure did not vary across groups or time points: right and left hippocampus, right and left ACC, right PCC, right and left JLC, and right IFG. For these models, all standardized factor loadings were significant ( $ps < 0.050$ ). In general, the variable with the highest loading was MD (average standardized loading =  $-0.857$ ), revealing that this measure was the strongest indicator of latent GM structural integrity; next was VBM (0.527), followed by MT (0.463). See full model set-up results in **Table 2** for more details. Results for combined left and right hemispheres can be found in **Supplementary Material**.

## Group Differences in Change

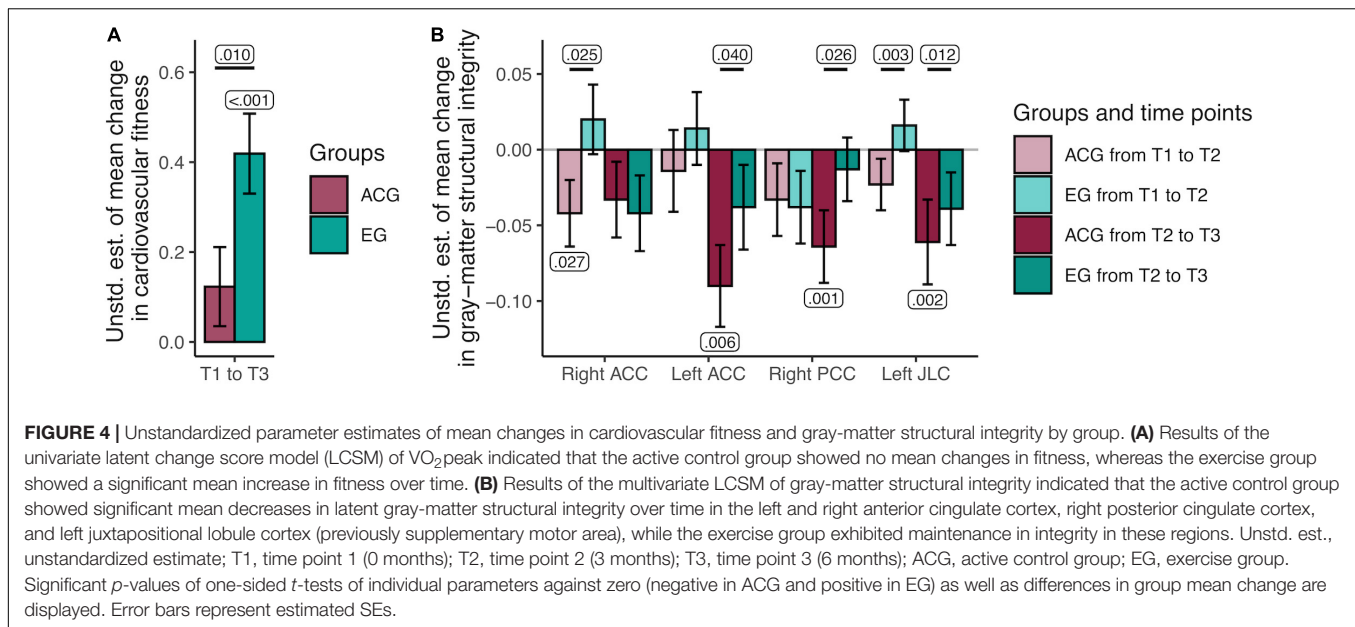
Group differences in mean change in structural integrity from T1 to T2 and from T2 to T3 were tested in those models that showed factorial invariance across groups and time points. All univariate LCSMs had acceptable fit indices, RMSEAs  $< 0.062$  (95% CIs =  $[0, <0.116]$ ), CFIs  $> 0.988$ . The EG showed significantly more positive change than the ACG in the right ACC from T1 to T2, in the left ACC from T2 to T3, in the right PCC from T2 to T3, and in the left JLC from T1 to T2 and from T2 to T3 (see **Table 2** and **Figure 4** for details). These group differences were primarily driven by mean decreases in GM integrity within the ACG: right ACC from T1 to T2,  $b_0 = -0.042$ , SE = 0.022,  $\Delta\chi^2(df = 1) = 3.68$ ,  $p = 0.027$ , left ACC from T2 to T3,  $b_0 = -0.090$ , SE = 0.027,  $\Delta\chi^2(df = 1) = 6.44$ ,  $p = 0.006$ , right PCC from T2 to T3,  $b_0 = -0.064$ , SE = 0.024,  $\Delta\chi^2(df = 1) = 9.05$ ,  $p = 0.001$ , left JLC from T2 to T3,  $b_0 = -0.061$ , SE = 0.028,  $\Delta\chi^2(df = 1) = 8.73$ ,  $p = 0.002$ . GM structural integrity did not significantly decline in the left JLC within the ACG from T1 to T2, nor did it significantly increase within the EG in any of the ROIs showing significant differences in group mean change. These time-by-group interaction effects were also detected when including age, sex, and years of education as indicator variables in the models, estimating means and residual variances of each, as well as regressions from each to baseline GM structural integrity, change from T1 to T2 and change from T2 to T3. In the following, we therefore discuss the simpler models, excluding these demographic factors, for the sake of parsimony.

## Correlations Between Change in Cardiovascular Fitness and Change in Gray-Matter Structural Integrity

Two models, the right ACC and the left JLC, showed group differences in mean changes and reliable variance in change, thereby allowing us to investigate whether individual differences in fitness changes and individual differences in integrity changes were correlated. Both models had acceptable fit indices, RMSEAs  $< 0.059$  (95% CIs =  $[0, <0.106]$ ), CFIs  $> 0.987$ . Change in cardiovascular fitness was positively correlated with change in GM structural integrity in the right ACC from T1 to T2, standardized estimate ( $\phi$ ) = 0.753,  $\Delta\chi^2(df = 1) = 11.60$ ,  $p < 0.001$ , and in the left JLC from T1 to T2,  $\phi = 0.469$ ,  $\Delta\chi^2(df = 1) = 6.42$ ,  $p = 0.006$ . Crucially, we observed a group difference in change-change correlation in the right ACC from T1 to T2,  $\Delta\chi^2(df = 1) = 4.52$ ,  $p = 0.017$ , with the ACG showing no significant correlation,  $\phi = 0.424$ ,  $\Delta\chi^2(df = 1) = 1.92$ ,  $p > 0.050$ , and the EG showing a significantly positive correlation,  $\phi = 1.110$ ,  $\Delta\chi^2(df = 1) = 10.54$ ,  $p = 0.001$ . These correlations were also detected when including age, sex, and years of education as indicator variables in the models, thus we discuss the results from the simpler models in the following.

## DISCUSSION

This study used a multivariate, multigroup approach in SEM to investigate the effects of six months of at-home aerobic exercise



on cardiovascular fitness and a latent measure of GM structural integrity comprising multiple MR imaging modalities, which may be a more reliable measure of structural integrity than individual MR measures, such as GM volume. Change-change relationships between fitness and GM structural integrity were also explored.

## Cardiovascular Fitness

Participants who engaged in interval training on a stationary bike at home for three to four days a week showed an increase in cardiovascular fitness, indexed by  $VO_2$ peak, over six months, and also improved more than an ACG who did not engage in regular aerobic exercise. This serves as a proof of concept for the current study, showing that the exercise intervention utilized in this sample was effective at improving cardiovascular fitness. Notably, the current design differs from previous exercise interventions that looked at exercise-induced changes in the brain in at least two dimensions: firstly, participants exercised in their own convenience in their homes, only coming to the lab once a week for a group stretching and toning session. This corroborates previous studies showing that regular at-home aerobic exercise can also improve cardiovascular fitness in older adults (King, 1991; Salvetti et al., 2008). This finding is important as regular exercise at home with only one supervised stretching and toning per week may be more accessible for an older population than personal training in a facility multiple times a week.

In addition to training taking place in participants' homes, the exercise regimes were also highly personalized and flexible. Initial difficulty for each participant was individually determined by a sports medicine physician, and during the intervention, participants could indicate the subjective perception of difficulty (i.e., too easy, too difficult) so that subsequent training could be modified accordingly. Participants were also not supervised during the exercise bouts. Still, this adaptive, at-home interval training regime was effective at increasing the cardiovascular

fitness of those in the EG over those in the ACG. This is also in line with research on fitness improvements in aging; for example, one study found that older adults who adhered to a six-month exercise program at home under no supervision had greater  $VO_2$ peak at post-test than those who did not adhere, while the groups did not differ at baseline (Morey et al., 2003). Taken together, an adaptive, at-home aerobic exercise regime seems to be an effective intervention for improving cardiovascular fitness in healthy, previously sedentary older adults.

## Gray-Matter Structural Integrity

The statistical analyses used in this study build on previous work (Kühn et al., 2017; Köhncke et al., 2021) in which a multimodal latent factor measuring GM structural integrity was established in a cross-sectional sample, and expanded this model for use in a longitudinal intervention study including three measurement time points to investigate patterns of change. Our results indicate that exercise promotes the maintenance (i.e., attenuated decrease) of structural integrity in regions of the brain that previously have been found to increase in volume in the course of an exercise intervention (Colcombe et al., 2006), namely the right and left ACC, the right PCC, and the left JLC (also termed supplementary motor area). These areas have been shown to undergo substantial age-related atrophy (e.g., Raz et al., 2005; Fjell et al., 2009), with exaggerated posterior atrophy found in patients with Alzheimer's disease (Lehmann et al., 2011). However, the vulnerability of brain structure to age-related effects in these regions seems to be accompanied by increased amenability to exercise- and physical fitness-induced maintenance and/or gains in older age (see also Colcombe et al., 2006; Ruscheweyh et al., 2011). It has been suggested before that intra-cortical myelin content may be one potentially important mechanism here (Walhovd et al., 2016), with high-myelin regions being more resistant to change, and regions with lower myelin content, such as the medial temporal

**TABLE 2** | Testing for invariance across groups and time points, and testing for equal vs. unequal mean change parameters across groups.

Region of interest	Group invariance (T1)		Time invariance (groups collapsed)		Standardized estimates of factor loadings			Group difference in mean change	
	$\Delta\chi^2_{df=2}$ baseline vs. metric	$\Delta\chi^2_{df=3}$ metric vs. strict	$\Delta\chi^2_{df=4}$ baseline vs. metric	$\Delta\chi^2_{df=6}$ metric vs. strict	VBM	MT	MD	$\Delta\chi^2_{df=1}$ $\Delta$ SI T1 to T2	$\Delta\chi^2_{df=1}$ $\Delta$ SI T2 to T3
Right HC*	0.33	4.34	0.69	4.67	0.792 <sup>†</sup>	0.576 <sup>†</sup>	-0.887 <sup>†</sup>	0.48	0.09
Left HC*	0.51	4.19	2.18	4.51	0.751 <sup>†</sup>	0.464 <sup>†</sup>	-0.987 <sup>†</sup>	1.33	0.42
Right ACC*	2.82	2.49	6.08	2.53	0.596 <sup>†</sup>	0.518 <sup>†</sup>	-0.895 <sup>†</sup>	3.87 <sup>†</sup>	0.38
Left ACC*	1.78	1.20	5.60	8.06	0.278 <sup>†</sup>	0.632 <sup>†</sup>	-0.256 <sup>†</sup>	1.85	3.07 <sup>†</sup>
Right PCC*	4.01	0.69	0.00	4.49	0.494 <sup>†</sup>	0.418 <sup>†</sup>	-1.000 <sup>†</sup>	0.75	3.75 <sup>†</sup>
Left PCC	0.00	6.21	17.12 <sup>†</sup>	-	-	-	-	-	-
Right PCG	0.00	1.28	10.16 <sup>†</sup>	-	-	-	-	-	-
Left PCG	4.64	2.33	10.85 <sup>†</sup>	-	-	-	-	-	-
Right JLC*	1.57	3.29	1.07	12.19	0.412 <sup>†</sup>	0.417 <sup>†</sup>	-0.942 <sup>†</sup>	0.95	1.87
Left JLC*	1.15	0.90	2.82	7.49	0.478 <sup>†</sup>	0.415 <sup>†</sup>	-0.931 <sup>†</sup>	7.45 <sup>†</sup>	5.05 <sup>†</sup>
Right IFG*	0.44	5.37	0.00	9.77	0.416 <sup>†</sup>	0.263 <sup>†</sup>	-0.957 <sup>†</sup>	2.34	0.15
Left IFG	5.02	1.44	19.10 <sup>†</sup>	-	-	-	-	-	-

HC, hippocampus; ACC, anterior cingulate cortex; PCC, posterior cingulate cortex; PCG, precentral gyrus; JLC, juxtapositional lobule cortex; IFG, inferior frontal gyrus;  $\Delta$ SI, change in structural integrity.

\*Model shows factorial invariance across groups and time points.

<sup>†</sup> $p < 0.050$ . Factorial invariance can be assumed if the measurement invariance test is non-significant (two-sided), meaning the model fit does not significantly worsen if parameters are set to be equal across groups/time points. Significance of standardized factor loadings tested using a Wald test (one-sided). A group difference in mean change can be inferred if the  $p$ -value is significant (one-sided), meaning the model fit significantly worsens when the parameter is constrained to be equal across groups.

lobe and cingulate cortices (Grydeland et al., 2013), being more prone to change in both the negative and positive direction.

The latent factor of GM structural integrity as established in this study captured the variance common to VBM, MT, and MD. Köhncke et al. (2021) reported that older individuals showed lower factor scores in the prefrontal cortex, hippocampus, and parahippocampal gyrus. This suggests that lower factor scores may be indicative of greater GM deterioration that occurs during normal aging, which could be caused by various and potentially correlated structural changes, including loss of dendritic spines and dendritic arbors, decreasing synaptic density, demyelination, and loss of glia and small blood vessels (Hof and Morrison, 2004; Morrison and Baxter, 2012; Zatorre et al., 2012; Raz and Daugherty, 2018). Considering the single indicators, the factor loadings indicate that lower factor scores result from a pattern of lower VBM, lower MT, and higher MD values, which are thought to reflect lower estimates of GM volume, myelination, and density, respectively. Therefore, a factor score capturing the shared variance between these indicators seems to represent general properties of GM structure that decline with age. Further supporting their interpretation that the latent factor reflects structural integrity, Köhncke et al. (2021) were also able to show a positive association between the latent factor and a latent factor comprising four episodic memory tasks, which is in line with the brain maintenance hypothesis that brain integrity across multiple levels is important for cognitive performance (Nyberg et al., 2012; Lindenberger, 2014; Cabeza et al., 2018; Nyberg and Pudas, 2019; Nyberg and Lindenberger, 2020; Johansson et al., 2022).

Here, we established the same latent factor of GM structural integrity in a longitudinal design encompassing three time points. The assumption of factorial invariance across two groups and three time points was found to be tenable for the right and

left hippocampus, right and left ACC, right PCC, left JLC, and right IFG. In the right and left ACC, right PCC, and left JLC, the ACG showed decreases in integrity, while the changes in the EG were significantly more positive, indicating that the exercise intervention had helped to maintain structural integrity in these areas. This supports the hypothesis that exercise has a neuroprotective effect on general structural integrity in older adults in areas of the brain where effects of exercise on volume have been observed before (e.g., Colcombe et al., 2006).

One mechanism hypothesized to underlie the relationship between aerobic exercise and brain structure is the increase in cerebral blood flow that occurs during bouts of exercise. In response to a complex combination of partial pressure of arterial carbon dioxide and oxygen, blood pressure, cerebral metabolism, and neurogenic regulation, acute physical exercise increases cerebral blood flow (Querido and Sheel, 2007; Smith and Ainslie, 2017), bringing with it oxygen and nutrients. With more resources available, both neurons and the surrounding cells may be better sustained, which would then be reflected in the latent factor of structural integrity. In contrast, individuals in the ACG, who did not engage in aerobic exercise, were less likely to experience this regular increase in cerebral blood flow, to the effect that GM structural integrity would be more likely to continue on a downward trajectory. Indeed, one study, a short-term exercise intervention (12 weeks) in older adults, even found an increase in cerebral blood flow at rest in the ACC within an exercise group vs. a control group (Chapman et al., 2013).

Notably, varying patterns of the timing of structural integrity changes were seen across ROIs. In the right ACC, the difference between groups in integrity change was seen in the first 3 months of the intervention, while in the left ACC and right PCC, this difference was seen in the second three months, and in the left

JLC, this difference was evident throughout the six months. To some degree, group differences in change may have emerged only later in the study because unspecific initial interventions effects may have been shared across both conditions. The active control participants, though not engaging in exercise training, also changed their daily and weekly routines to incorporate more interaction with technology (tablet use), as well as with new social partners (weekly group sessions), which might have constituted a departure from their usual routine with potentially beneficial effects, in line with work suggesting positive associations between brain maintenance, cognition, and an active lifestyle (Lövdén et al., 2005; Hertzog et al., 2008; Nyberg et al., 2012; Small et al., 2012; Mintzer et al., 2019). However, to the degree that they habituated to these new daily practices, the initial overall effect might have worn off, while the mechanisms conveying a positive effect of exercise on brain integrity continued to operate in EG participants. Conversely, participants in the EG may have experienced initial maintenance in structural integrity at the beginning of their new training regimes, but as their brain and vascular systems grew more accustomed to the impulse afforded by aerobic exercise, a normal trajectory of decline might have resumed. Given the small sample size, these considerations are clearly speculative. More research is needed to better understand the cascade of mechanisms that convey benefits of aerobic exercise on different areas of the aging human brain.

### **Positive Correlation Between Change in Cardiovascular Fitness and Structural Integrity in the Right Anterior Cingulate Cortex and Left Juxtapositional Lobule Cortex**

Finally, in the right ACC and the left JLC, we were able to reliably measure significant individual differences in change, allowing us to investigate change-change correlations with cardiovascular fitness. A positive correlation was found between change in cardiovascular fitness and change in right ACC structural integrity from T1 to T2 in exercisers but not controls, indicating that those exercisers who gained more cardiovascular fitness during the intervention also showed less decline in structural integrity in the right ACC during the first 3 months of the intervention. A positive correlation was also found between change in cardiovascular fitness and change in left JLC structural integrity from T1 to T2, but this correlation did not differ between groups.

Many cross-sectional studies investigating the relationship between cardiovascular fitness and brain structure (using a single indicator approach) in older adults have found positive associations between fitness and in frontal areas, temporal areas, or both, as well as parietal, posterior (e.g., precuneus), and sub-cortical (e.g., caudate) areas (see review by d'Arbeloff, 2020). Similarly, some non-intervention longitudinal studies have reported positive associations between fitness and brain structure in similar brain regions (see d'Arbeloff, 2020), and one study found that baseline cardiovascular fitness was related to the progression of dementia severity and brain atrophy in Alzheimer's patients (Vidoni et al., 2012). The current findings

extend this previous work by reporting a change in the right ACC that is likely to reflect a causal effect of a change in cardiovascular fitness.

The current study has a number of limitations that should be addressed in future studies. First, the sample size was relatively small, especially for SEM. This might help to explain why individual differences in change often failed to differ reliably from zero. The current sample is also relatively homogeneous; healthy, previously sedentary older adults from an area with relatively high socioeconomic status were recruited to participate and were further screened for health conditions before being allowed to participate in the study. Healthy sedentary adults might be equipped with a range of protective factors that keep them healthy in the presence of a lifestyle that might result in deteriorating health in most other individuals. Thus, the generalizability of the present results to other segments of the aging population is unclear.

## **CONCLUSION**

In this study, we introduced a multimodal modeling approach for investigating the effects of aerobic fitness interventions on regional GM structural integrity in human aging. Our findings corroborate and extend earlier results by showing that at-home aerobic exercise among healthy sedentary older adults results in improved cardiovascular fitness and helps to maintain GM structural integrity in areas that have been found to show exercise-induced volume changes.

## **DATA AVAILABILITY STATEMENT**

The original contributions presented in the study are publicly available. The data and relevant scripts for analysis can be found here: <https://osf.io/yw865/>.

## **ETHICS STATEMENT**

The studies involving human participants were reviewed and approved by the Ethics Committee of the German Psychological Society (DGPs). The participants provided their written informed consent to participate in this study.

## **AUTHOR CONTRIBUTIONS**

SP assisted with data acquisition, analyzed the data, interpreted the results, and wrote the manuscript. MK preprocessed imaging data and revised the manuscript. YK and AB assisted with SEM analysis and revised the manuscript. NB designed the neuroimaging protocol and revised the manuscript. CM and JP performed physical assessments including cardiopulmonary exercise testing and revised the manuscript. BW designed the physical assessment protocol and revised the manuscript. SK designed the study and revised the manuscript. UL and SD designed the study, interpreted the results,



and revised the manuscript. EW designed the study, preprocessed imaging data, interpreted the results, and revised the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Max Planck Society and the Max Planck Institute for Human Development and is part of the BMBF-funded Energi Consortium (01GQ1421B).

## ACKNOWLEDGMENTS

We are very grateful to the Neotiv team for providing the app for ergometer training as well as their technical support, to everyone at Shared Reading for organizing the book club for our active

## REFERENCES

- Ashburner, J., and Friston, K. J. (2000). Voxel-Based Morphometry—The Methods. *NeuroImage* 11, 805–821. doi: 10.1006/nimg.2000.0582
- Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., and Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* 54, 2033–2044. doi: 10.1016/j.neuroimage.2010.09.025
- Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., et al. (2010). The optimal template effect in hippocampus studies of diseased populations. *NeuroImage* 49, 2457–2466. doi: 10.1016/j.neuroimage.2009.09.062
- Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system—A technical review. *NMR Biomed.* 15, 435–455. doi: 10.1002/nbm.782
- Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., et al. (2021). *OpenMx 2.19.6 User Guide\**.
- Cabeza, R., Albert, M., Belleville, S., Craik, F. I. M., Duarte, A., Grady, C. L., et al. (2018). Maintenance, reserve and compensation: The cognitive neuroscience of healthy ageing. *Nat. Rev. Neurosci.* 19, 701–710. doi: 10.1038/s41583-018-0068-2
- Chalmers, R. P., and Flora, D. B. (2015). faoutlier: An R Package for Detecting Influential Cases in Exploratory and Confirmatory Factor Analysis. *Appl. Psychol. Measurement* 39, 573–574. doi: 10.1177/0146621615597894
- Chapman, S., Aslan, S., Spence, J., DeFina, L., Keebler, M., Didehbani, N., et al. (2013). Shorter term aerobic exercise improves brain, cognition, and cardiovascular fitness in aging. *Front. Aging Neurosci.* 5:75. doi: 10.3389/fnagi.2013.00075
- Colcombe, S. J., Erickson, K. I., Raz, N., Webb, A. G., Cohen, N. J., McAuley, E., et al. (2003). Aerobic Fitness Reduces Brain Tissue Loss in Aging Humans. *J. Gerontol. Series A: Biol. Sci. Med. Sci.* 58, M176–M180. doi: 10.1093/gerona/58.2.M176
- Colcombe, S. J., Erickson, K. I., Scaif, P. E., Kim, J. S., Prakash, R., McAuley, E., et al. (2006). Aerobic Exercise Training Increases Brain Volume in Aging Humans. *J. Gerontol. Series A Biol. Sci. Med. Sci.* 61, 1166–1170. doi: 10.1093/gerona/61.11.1166
- Colcombe, S. J., Kramer, A. F., McAuley, E., Erickson, K. I., and Scaif, P. (2004). Neurocognitive Aging and Cardiovascular Fitness: Recent Findings and Future Directions. *J. Mol. Neurosci.* 24, 009–014. doi: 10.1385/JMN:24:1:009
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- d'Arbeloff, T. (2020). Cardiovascular fitness and structural brain integrity: An update on current evidence. *GeroScience* 42, 1285–1306. doi: 10.1007/s11357-020-00244-7
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing

control group, to Michael Krause for his continuous assistance in the implementation of data preprocessing on the computing cluster, and to Steven M. Boker and Timo von Oertzen for their input on the structural equation modeling. We thank Sebastian Schröder and his student assistants for providing the technical infrastructure of the study, Kirsten Becker and Anke Schepers-Klingebiel for their organizational assistance, and the MRI team at the Max Planck Institute for Human Development (Sonali Beckmann, Nadine Taube, Thomas Feg, and Davide Santoro) and all the participants for their time and support.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2022.852737/full#supplementary-material>

- the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.10.1021
- Erickson, K. I., Voss, M. W., Prakash, R. S., Basak, C., Szabo, A., Chaddock, L., et al. (2011). Exercise training increases size of hippocampus and improves memory. *Proc. Natl. Acad. Sci.* 108, 3017–3022. doi: 10.1073/pnas.1015950108
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Beh. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146
- Fjell, A. M., and Walhovd, K. B. (2010). Structural Brain Changes in Aging: Courses. *Causes Cogn. Conseq. Rev. Neurosci.* 21, 187–221. doi: 10.1515/REVNEURO.2010.21.3.187
- Fjell, A. M., Walhovd, K. B., Fennema-Notestine, C., McEvoy, L. K., Hagler, D. J., Holland, D., et al. (2009). One-year brain atrophy evident in healthy aging. *J. Neurosci. J. Soc. Neurosci.* 29, 15223–15231. doi: 10.1523/JNEUROSCI.3252-09.2009
- Fukutomi, H., Glasser, M. F., Murata, K., Akasaka, T., Fujimoto, K., Yamamoto, T., et al. (2019). Diffusion Tensor Model links to Neurite Orientation Dispersion and Density Imaging at high b-value in Cerebral Cortical Gray Matter. *Sci. Rep.* 9:12246. doi: 10.1038/s41598-019-48671-7
- Gaser, C., and Dahnke, R. (2016). CAT – A computational anatomy toolbox for the analysis of structural MRI data. *Hum. Brain Mapp.* 2016, 336–348.
- Ghisletta, P., and Lindenberger, U. (2004). Static and Dynamic Longitudinal Structural Analyses of Cognitive Changes in Old Age. *Gerontology* 50, 12–16. doi: 10.1159/000074383
- Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N. A., Friston, K. J., and Frackowiak, R. S. J. (2001). A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains. *NeuroImage* 14, 21–36. doi: 10.1006/nimg.2001.0786
- Grydeland, H., Walhovd, K. B., Tamnes, C. K., Westlye, L. T., and Fjell, A. M. (2013). Intracortical Myelin Links with Performance Variability across the Human Lifespan: Results from T1- and T2-Weighted MRI Myelin Mapping and Diffusion Tensor Imaging. *J. Neurosci.* 33, 18618–18630. doi: 10.1523/JNEUROSCI.2811-13.2013
- Helms, G., Dathe, H., and Dechent, P. (2008a). Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation: Rational Approximation of the FLASH Signal. *Magnet. Resonan. Med.* 59, 667–672. doi: 10.1002/mrm.21542
- Helms, G., Dathe, H., Kallenberg, K., and Dechent, P. (2008b). High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity

- and T1 relaxation obtained from 3D FLASH MRI: Saturation and Relaxation in MT FLASH. *Magnet. Resonan. Med.* 60, 1396–1407. doi: 10.1002/mrm.21732
- HelpAge International. (2018). *Global AgeWatch Insights. The right to Health for Older People, the Right to be Counted*. Available online at: <http://www.globalagewatch.org/download/5c0e922bebfcd>
- Hertzog, C., Kramer, A. F., Wilson, R. S., and Lindenberger, U. (2008). Enrichment Effects on Adult Cognitive Development: Can the Functional Capacity of Older Adults Be Preserved and Enhanced? *Psychological science in the public interest. J. Am. Psychol. Soc.* 9, 1–65. doi: 10.1111/j.1539-6053.2009.01034.x
- Hof, P. R., and Morrison, J. H. (2004). The aging brain: Morphomolecular senescence of cortical circuits. *Trends Neurosci.* 27, 607–613. doi: 10.1016/j.tins.2004.07.013
- Hunter, M. D. (2018). State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment. *Structural Equation Modeling. Multidiscipl. J.* 25, 307–324. doi: 10.1080/10705511.2017.1369354
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *NeuroImage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Johansson, J., Wählin, A., Lundquist, A., Brandmaier, A. M., Lindenberger, U., and Nyberg, L. (2022). Model of brain maintenance reveals specific change-change association between medial-temporal lobe integrity and episodic memory. *Aging Brain* 2:100027. doi: 10.1016/j.nbas.2021.100027
- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., et al. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Dev. Cogn. Neurosci.* 33, 99–117. doi: 10.1016/j.dcn.2017.11.007
- King, A. C. (1991). Group- vs Home-Based Exercise Training in Healthy Older Men and Women: A Community-Based Clinical Trial. *JAMA* 266:1535. doi: 10.1001/jama.1991.03470110081037
- Kleemeyer, M. M., Kühn, S., Prindle, J., Bodammer, N. C., Brechtel, L., Garthe, A., et al. (2016). Changes in fitness are associated with changes in hippocampal microstructure and hippocampal volume among older adults. *NeuroImage* 131, 155–161. doi: 10.1016/j.neuroimage.2015.11.026
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling (Fourth edition)*. New York: The Guilford Press.
- Köhneke, Y., Düzel, S., Sander, M. C., Lindenberger, U., Kühn, S., and Brandmaier, A. M. (2021). Hippocampal and Parahippocampal Gray Matter Structural Integrity Assessed by Multimodal Imaging Is Associated with Episodic Memory in Old Age. *Cereb. Cortex* 31, 1464–1477. doi: 10.1093/cercor/bhaa287
- Kühn, S., Düzel, S., Eibich, P., Krekel, C., Wüstemann, H., Kolbe, J., et al. (2017). In search of features that constitute an “enriched environment” in humans: Associations between geographical properties and brain structure. *Sci. Rep.* 7:11920. doi: 10.1038/s41598-017-12046-7
- Lehmann, M., Crutch, S. J., Ridgway, G. R., Ridha, B. H., Barnes, J., Warrington, E. K., et al. (2011). Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical Alzheimer’s disease. *Neurobiol. Aging* 32, 1466–1476. doi: 10.1016/j.neurobiolaging.2009.08.017
- Lindenberger, U. (2014). Human cognitive aging: Corriger la fortune? *Science* 346, 572–578. doi: 10.1126/science.1254403
- Lövdén, M., Ghisletta, P., and Lindenberger, U. (2005). Social participation attenuates decline in perceptual speed in old and very old age. *Psychol. Aging* 20, 423–434. doi: 10.1037/0882-7974.20.3.423
- Maass, A., Düzel, S., Goerke, M., Becke, A., Sobieray, U., Neumann, K., et al. (2015). Vascular hippocampal plasticity after aerobic exercise in older adults. *Mol. Psychiatr.* 20, 585–593. doi: 10.1038/mp.2014.114
- Mintzer, J., Donovan, K. A., Kindy, A. Z., Lock, S. L., Chura, L. R., and Barracca, N. (2019). Lifestyle choices and brain health. *Front. Med.* 6:204. doi: 10.3389/fmed.2019.00204
- Morey, M. C., Dubbert, P. M., Doyle, M. E., MacAller, H., Crowley, G. M., Kuchibhatla, M., et al. (2003). From supervised to unsupervised exercise: factors associated with exercise adherence. *J. Aging. Phys. Act.* 11, 351–368. doi: 10.1123/japa.11.3.351
- Morrison, J. H., and Baxter, M. G. (2012). The ageing cortical synapse: Hallmarks and implications for cognitive decline. *Nat. Rev. Neurosci.* 13, 240–250. doi: 10.1038/nrn3200
- Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., et al. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika* 81, 535–549. doi: 10.1007/s11336-014-9435-8
- Nyberg, L., and Lindenberger, U. (2020). ““Brain maintenance and cognition in old age,”” in *The Cognitive Neurosciences*, 6th Edn, eds D. Poeppel, G. R. Mangun, and M. S. Gazzaniga (Cambridge: MIT Press), 81–89.
- Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., and Bäckman, L. (2012). Memory aging and brain maintenance. *Trends Cogn. Sci.* 16, 292–305. doi: 10.1016/j.tics.2012.04.005
- Nyberg, L., and Pudas, S. (2019). Successful Memory Aging. *Ann. Rev. Psychol.* 70, 219–243. doi: 10.1146/annurev-psych-010418-103052
- Pierpaoli, C., and Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magnet. Resonan. Med.* 36, 893–906. doi: 10.1002/mrm.1910360612
- Pritikin, J. N., Hunter, M. D., and Boker, S. M. (2015). Modular Open-Source Software for Item Factor Analysis. *Educ. Psychol. Measurement* 75, 458–474. doi: 10.1177/0013164414554615
- Querido, J. S., and Sheel, A. W. (2007). Regulation of Cerebral Blood Flow During Exercise. *Sports Med.* 37, 765–782. doi: 10.2165/00007256-200737090-00002
- Raz, N., and Daugherty, A. M. (2018). Pathways to Brain Aging and Their Modifiers: Free-Radical-Induced Energetic and Neural Decline in Senescence (FRIENDS) Model - A Mini-Review. *Gerontology* 64, 49–57. doi: 10.1159/000479508
- Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., et al. (2005). Regional Brain Changes in Aging Healthy Adults: General Trends. *Individ. Diff. Modif. Cereb. Cortex* 15, 1676–1689. doi: 10.1093/cercor/bhi044
- RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. Boston: RStudio, PBC.
- Ruscheweyh, R., Willemer, C., Krüger, K., Duning, T., Warnecke, T., Sommer, J., et al. (2011). Physical activity and memory functions: An interventional study. *Neurobiol. Aging* 32, 1304–1319. doi: 10.1016/j.neurobiolaging.2009.08.001
- Salvetti, X. M., Oliveira, J. A., Servantes, D. M., Vincenzo, and de Paola, A. A. (2008). How much do the benefits cost? Effects of a home-based training programme on cardiovascular fitness, quality of life, programme cost and adherence for patients with coronary disease. *Clin. Rehab.* 22, 987–996. doi: 10.1177/0269215508093331
- Schermelele-Engel, K., Moosbrugger, H., and Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods Psychol. Res.* 8, 23–74.
- Seiler, S., Ropele, S., and Schmidt, R. (2014). Magnetization Transfer Imaging for *in vivo* Detection of Microstructural Tissue Changes in Aging and Dementia: A Short Literature Review. *J. Alzheimer’s Dis.* 42, S229–S237. doi: 10.3233/JAD-132750
- Small, B. J., Dixon, R. A., McArdle, J. J., and Grimm, K. J. (2012). Do changes in lifestyle engagement moderate cognitive decline in normal aging? *Evid. Victoria Longitud. Stud. Neuropsychology* 26, 144–155. doi: 10.1037/a0026579
- Smith, K. J., and Ainslie, P. N. (2017). Regulation of cerebral blood flow and metabolism during exercise: Cerebral blood flow and metabolism during exercise. *Exp. Physiol.* 102, 1356–1371. doi: 10.1113/EP086249
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Song, S.-K., Yoshino, J., Le, T. Q., Lin, S.-J., Sun, S.-W., Cross, A. H., et al. (2005). Demyelination increases radial diffusivity in corpus callosum of mouse brain. *NeuroImage* 26, 132–140. doi: 10.1016/j.neuroimage.2005.01.028
- Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., et al. (2019). HMRI – A toolbox for quantitative MRI in neuroscience and clinical research. *NeuroImage* 194, 191–210. doi: 10.1016/j.neuroimage.2019.01.029
- Tahedi, M. (2018). *B.A.T.M.A.N.: Basic and Advanced Tractography with MRtrix for All Neurophiles*. OSF Home, doi: 10.17605/OSF.IO/FKYHT
- Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., et al. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* 202:116137. doi: 10.1016/j.neuroimage.2019.116137
- Vidoni, E. D., Honea, R. A., Billinger, S. A., Swerdlow, R. H., and Burns, J. M. (2012). Cardiorespiratory fitness is associated with atrophy in Alzheimer’s and aging over 2 years. *Neurobiol. Aging* 33, 1624–1632. doi: 10.1016/j.neurobiolaging.2011.03.016

- Walhovd, K. B., Westerhausen, R., de Lange, A.-M. G., Bråthen, A. C. S., Grydeland, H., Engvig, A., et al. (2016). Premises of plasticity — And the loneliness of the medial temporal lobe. *NeuroImage* 131, 48–54. doi: 10.1016/j.neuroimage.2015.10.060
- Wansbeek, T., and Meijer, E. (2001). “Measurement error and latent variables,” in *A Companion to Theoretical Econometrics*, ed. B. H. Baltagi (New Jersey: Wiley), 162–179.
- Weinstein, A. M., Voss, M. W., Prakash, R. S., Chaddock, L., Szabo, A., White, S. M., et al. (2012). The association between aerobic fitness and executive function is mediated by prefrontal cortex volume. *BrainBehav., Immun.* 26, 811–819. doi: 10.1016/j.bbi.2011.11.008
- Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., and Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *NeuroImage* 54, 2116–2124. doi: 10.1016/j.neuroimage.2010.10.023
- Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., et al. (2013). Quantitative multi-parameter mapping of R1, PD\*, MT, and R2\* at 3T: A multi-center validation. *Front. Neurosci.* 7: 95. doi: 10.3389/fnins.2013.0095
- Wenger, E., Düzel, S., Kleemeyer, M. M., Polk, S. E., Köhncke, Y., Bodammer, N. C., et al. (2022). Vamos en bici: Study protocol of an investigation of cognitive and neural changes following language training, physical exercise training, or a combination of both. *BioRxiv* [Preprint]. doi: 10.1101/2022.01.30.478181
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage* 45, S173–S186. doi: 10.1016/j.neuroimage.2008.10.055
- Zatorre, R. J., Fields, R. D., and Johansen-Berg, H. (2012). Plasticity in gray and white: Neuroimaging changes in brain structure during learning. *Nat. Neurosci.* 15, 528–536. doi: 10.1038/nn.3045

**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.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Polk, Kleemeyer, Köhncke, Brandmaier, Bodammer, Misgeld, Porst, Wolfarth, Kühn, Lindenberger, Wenger and Düzel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## **S1: Supplementary methods**

### **Statistical analyses**

#### ***Magnetic resonance imaging***

**Acquisition.**  $T_1$ -weighted images were obtained using a 3D  $T_1$ -weighted magnetization prepared gradient-echo (MPRAGE) sequence using the following parameters: repetition time (TR) = 2500 ms; echo time (TE) = 4.77 ms; inversion time (TI) = 1100 ms; flip angle =  $7^\circ$ ; acquisition matrix =  $256 \times 256 \times 192$ ;  $1 \text{ mm}^3$  isotropic voxels; with the prescan normalize option and a 3D distortion correction for non-linear gradients; acquisition time = 9:20 min.

The multi-parameter mapping protocol used to acquire the MT maps comprised one static magnetic ( $B_0$ ) gradient echo (GRE)-field map, one radiofrequency (RF) transmit field map ( $B_1$ ), and three multi-echo 3D fast low angle shot (FLASH) scans (Helms, Dathe, & Dechent, 2008; Tabelow et al., 2019). The  $B_0$  GRE-field mapping sequence was acquired with the following parameters: 64 transversal slices; slice thickness = 2 mm with 50% distance factor; TR = 1020 ms; TE1/TE2 = 10/12.46 ms; flip angle =  $90^\circ$ ; acquisition matrix =  $64 \times 64$ ; FOV =  $192 \times 192$  mm; right-left phase encoding direction; bandwidth (BW) = 260 Hz/pixel; flow compensation;  $3.0 \times 3.0 \times 2.0 \text{ mm}^3$  voxel size; acquisition time = 2:14 min.

Maps of the local RF transmit/ $B_1^+$  field were acquired following recommendations by Lutti and colleagues (2010) and were measured and estimated from a 3D EPI acquisition of spin and stimulated echoes (SE and STE) with different flip angles. The following parameters were used: 4 mm isotropic resolution, matrix =  $64 \times 48 \times 48$ , FOV =  $256 \times 192 \times 192$  mm, parallel imaging using GRAPPA factor  $2 \times 2$  in PE and partition directions, TR = 500 ms, TE<sub>SE/STE</sub>/mixing time = 39.06 ms/33.80 ms. Eleven pairs of SE/STE image volumes were measured successively employing decreasing flip angles  $\alpha$  from  $115^\circ$  to  $65^\circ$  in steps of  $-5^\circ$  (applied in a  $\alpha-2\alpha-\alpha$  series of RF pulses to produce SEs and STEs; see Akoka et al., 1993). Acquisition time was 3 min.

The three different multi-echo FLASH sequences were acquired with predominant  $T_1$  weighting ( $T_1w$ ), proton density weighting (PDw), or magnetization transfer weighting (MTw) by adjusting the repetition time (TR) and flip angle ( $\alpha$ ;  $T_1w$ : TR/ $\alpha$  = 24.5 ms/ $21^\circ$ ; PDw and MTw: TR/ $\alpha$  = 24.5 ms/ $6^\circ$ ) and applying an off-resonance Gaussian-shaped RF pulse (4 ms duration,  $220^\circ$  nominal flip angle, 2 kHz frequency offset from water resonance); for the MTw sequence this was applied prior to excitation. Multiple gradient echoes with alternating readout polarity were acquired at six equidistant echo times (TE) between 2.34 ms and 14.04 ms for the  $T_1w$  and MTw acquisitions, with two additional echoes at TE = 16.38

ms and 18.72 ms for the PDw acquisition. A high readout bandwidth (BW) of 465 Hz/pixel was used to minimize off-resonance artifacts. To reduce data acquisition time, GRAPPA parallel imaging with an acceleration factor of two was applied in the phase-encoding direction (anterior-posterior; outer/slow phase encoding loop) and 6/8 partial Fourier acquisitions in the partition direction (left-right; inner/fast phase encoding loop). The following additional acquisition parameters were used: 1 mm isotropic resolution, 176 slices per slab, FOV = 256 × 240 mm, acquisition time of each FLASH sequence = 7:03 min.

Diffusion-weighted images were obtained with a single-shot diffusion-weighted spin-echo-refocused echo-planar imaging sequence with the following parameters: TR = 9700 ms; TE = 120 ms; 62 slices; FOV = 224 × 224 mm; a two-shell scheme was used for diffusion-weighting applying two  $b$ -values: 710 s/mm<sup>2</sup> (30 directions) and 2850 s/mm<sup>2</sup> (60 directions), with directions distributed over a whole sphere for each shell, plus ten non-diffusion-weighted images; GRAPPA acceleration factor = 2; 2 mm<sup>3</sup> isotropic voxels; acquisition time = 16:41 min. Six additional images inverting the phase encoding direction were acquired without diffusion weighting; acquisition time = 1:29 min.

**Preprocessing.** Structural T<sub>1</sub>-weighted images were preprocessed using the Computational Anatomy Toolbox 12 (CAT12, Structural Brain Mapping group, Jena University Hospital; Gaser & Dahnke, 2016) in Statistical Parametric Mapping (SPM12, Institute of Neurology; [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) using the default parameters of the longitudinal pipeline, in which an individual participant's images first undergo an inverse-consistent realignment (including intra-subject bias correction) and a mean image is calculated. Spatial normalization parameters are then estimated using Dartel normalization based on the segmentations of the mean image. These normalization parameters are then applied to the segmentations of the images at all time points. Nonlinear-only modulation of gray and white matter segments was applied. These images were smoothed using an 8 mm full-width half-maximum (FWHM) standard Gaussian kernel.

Estimation of MT maps was conducted in SPM12 using the hMRI toolbox (Tabelow et al., 2019; <https://hmri-group.github.io/hMRI-toolbox/>). Within this toolbox, quantitative as well as semi-quantitative estimates of MT, PD, R1, and R2\* were computed from unprocessed multi-echo T<sub>1</sub>w, PDw, and MTw RF-spoiled gradient echo acquisitions using the *Create hMRI maps* module, which corrects the qMRI estimates for spatial receive and transmit field inhomogeneities. As has been described in more detail elsewhere (Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008; Weiskopf et al., 2011, 2013), the signal from the multi-echo T<sub>1</sub>w, PDw, and MTw echoes can be described by the

Ernst equation (Ernst & Anderson, 1966; Helms, Dathe, & Dechent, 2008; Helms, Dathe, Kallenberg, et al., 2008).

The effective transverse relaxation rate ( $R2^* = 1/T_2^*$ ) was derived from the TE dependence of the signal. The unified description of the multi-echo data from all three contrasts into a single signal model, or ESTATICS (Weiskopf et al., 2014), provides a more robust estimation of  $R2^*$  with a higher signal-to-noise ratio. Using approximations of the signal equations for small repetition time TR and small flip angle, the longitudinal relaxation rate (R1),  $A^*$  map (proportional to PD), and MT were estimated. PD maps were then computed by calibrating the mean PD value in white matter to 69 percent units, since the global mean PD cannot be estimated accurately.

A corrected MT saturation value was calculated by correcting the original MT value by the local RF transmit field. This semi-quantitative parameter is unaffected by R1 and RF transmit field variations, in contrast to the conventional MT ratio (Helms, Dathe, Kallenberg, et al., 2008). As implemented in the toolbox, *Unified Segmentation* was used to correct for RF sensitivity bias and applied the recommended 3D EPI B1 bias correction (Lutti et al., 2012).

A longitudinal processing pipeline of the data was adapted to achieve an improved within-subject coregistration of the created maps. In this pipeline, all MT and PD maps were first thresholded (MT: 0–5, PD: 0–200) to improve segmentation performance. Multichannel segmentations were then conducted using the thresholded MT and PD maps at all measurement time points (up to three per participant). The resulting gray and white matter segmentations from all three measurements per subject, formatted to be imported to DARTEL using the *DARTEL imported* option, were then fed into *SHOOT* to create an unbiased within-subject registration. The respective deformations were applied to the raw MT and PD maps to warp them into each subject's template space, and the median MT and PD maps across all measurements of a single individual were computed. These median maps were then subjected to a multichannel segmentation again and a group template was created from the resulting DARTEL-imported gray and white matter segmentations of all subjects using *SHOOT*. Subsequently, the two deformation fields (from native space to subject template space and from subject template space to group template space) were combined. For each measurement, this combined deformation field together with all four parameter maps (MT, PD, R1,  $R2^*$ ) was fed into *SHOOT normalize* to achieve normalization to MNI space. The gray and white matter segmentations derived by the multichannel segmentation of the median MT and PD maps were also spatially normalized to MNI space using Jacobian

modulation. Finally, from these normalized tissue class segmentations as well as the four parameter maps, smoothed tissue-specific MPMs were computed applying a 6mm FWHM smoothing kernel and weighted averaging.

DW images were preprocessed using MRtrix (version 3.0\_RC3; Tournier et al., 2019), FSL (FMRIB's Software Library, version 6.0.2; Jenkinson et al., 2012; Smith et al., 2004; Woolrich et al., 2009), and ANTS (version 2.2.0; Avants et al., 2010, 2011), following the Basic and Advanced Tractography with MRtrix for All Neurophiles (B.A.T.M.A.N.) tutorial (Tahedl, 2018). DW images were first denoised using *dwidenoise* with a  $5 \times 5 \times 5$  patch size, Gibb's ringing artifacts were removed using *mrdegibbs*, correction for EPI distortion, B<sub>0</sub>-field inhomogeneity, and eddy-current and movement distortion were applied using FSL's *topup* and *eddy\_cuda* with the default settings, adding outlier detection and replacement, and saving contrast-to-noise ratio and residual maps. Binary brain masks were created by running FSL's *bet* on each participant's mean b0 image. These were visually checked and manually fixed in cases where needed. Diffusion tensor maps were fit to each dataset using the lower *b*-value (710 s/mm<sup>2</sup>) with the MRtrix command *dwi2tensor* and afterwards MD values were derived via *tensor2metric*.

### ***Modeling changes in fitness***

**Group mean differences in change in VO<sub>2</sub>peak.** Latent change score models (LCSM) were used to evaluate group mean differences in change in VO<sub>2</sub>peak and GM integrity following the tutorial by Kievit and colleagues (2018). A univariate LCSM was built to measure mean change in VO<sub>2</sub>peak, in which a pseudo-latent factor,  $\Delta\text{VO}_2\text{peak}$ , captured the difference in VO<sub>2</sub>peak between T1 and T3, which was in this case simply a difference score, as VO<sub>2</sub>peak was directly measured. The means and variances of VO<sub>2</sub>peak at T1 and  $\Delta\text{VO}_2\text{peak}$  were estimated, as well as the correlation between the two. To test whether the EG showed significantly more positive mean change than the ACG, a multigroup model was used to test whether the change means could be fixed to equality across groups without significantly affecting the model fit.

### ***Modeling changes in GM structural integrity***

**Factorial invariance testing of GM integrity models.** GM integrity was modeled as a latent variable with freely estimated mean and variance in each of the pre-selected ROIs separately. This latent variable predicted VBM (loading fixed to 1 for model identification), MT (loading freely estimated), and MD (loading freely estimated), each with freely estimated residual variance. To ensure that the latent factor measured represented the same construct across groups as well as across time points, a series of measurement invariance tests were

conducted sequentially, first across groups at T1 then across time points, collapsing the two groups. First, the GM integrity factor structure at T1 was tested for metric (i.e., identical factor loadings) then strict factorial invariance (i.e., identical residual variances) across the two groups using LRTs (Cheung & Rensvold, 1999). In ROIs showing strict group invariance, the two groups were collapsed, and factorial invariance across the three time points was tested by including the same measurement structure of GM integrity at T2 and T3 in an unstructured model with freely estimated covariance between latent integrity at each time point (Widaman et al., 2010), as well as residual covariances of each image modality (T1 to T2, T2 to T3, and T1 to T3), set to be equal within each modality. Residual means were set to zero under the assumption that changes in the latent integrity factor would reflect changes in the three observed variables similarly; that is, the pattern of change over time was assumed to be similar across observed variables. Further, none of the models showed a worse fit when assuming residual means of zero versus freely estimated, thus we omit them for the sake of parsimony.

**Group differences in mean change in GM integrity.** For those regions showing strict factorial invariance across time (i.e., identical factor loadings and residual variances), to investigate group differences in mean change, the unstructured latent matrix was then replaced with a structured matrix modeling latent change scores from T1 to T2 and from T2 to T3. This allowed the models to have different means and variances in change between each pair of consecutive time points. Covariances between T1 and both change from T1 to T2 and from T2 to T3 and the covariance between the two latent change scores were freely estimated. Differences in group mean change were then tested using a multigroup model; an LRT was conducted between a model in which the means were freely estimated within each group and a model in which the means were restricted to be equal across groups.

**Change-change covariance between VO<sub>2</sub>peak and GM integrity.** Lastly, the relationship between change in VO<sub>2</sub>peak and change in latent GM integrity was investigated using a bivariate LCSM, in which the covariance between latent change score variables was estimated (VO<sub>2</sub>peak change to GM integrity change from T1 to T2, VO<sub>2</sub>peak change to GM integrity change from T2 to T3), accounting for the covariance between the two baseline variables (VO<sub>2</sub>peak and latent GM integrity at T1). See Figure 2 for the full bivariate LCSM. First, to confirm that reliable individual differences in change in GM integrity could be detected in the full multivariate model, which is necessary to investigate change-change relationships, two separate LRTs were conducted between the full model and 1) a model in which the variance of change in integrity from T1 to T2 and all covariance parameters



connected with change in integrity from T1 to T2 were fixed to zero ( $df = 4$ ), and 2) a model in which the variance of change in integrity from T2 to T3 and all covariance parameters connected with change in integrity from T2 to T3 were fixed to zero ( $df = 4$ ) following the logic of the generalized variance test (e.g., Brandmaier et al., 2018). To investigate whether the covariance between change in cardiovascular fitness and change in GM integrity was significant, the full model including both groups was compared to a constrained nested model in which one covariance path was fixed to zero. Finally, an LRT was run to investigate whether the covariance paths could be set to equal across groups without significantly affecting model fit, and if so, LRTs were used to test whether the covariances were greater than zero in each group separately.

### References

- Akoka, S., Franconi, F., Seguin, F., & Le Pape, A. (1993). Radiofrequency map of an NMR coil by imaging. *Magnetic Resonance Imaging*, *11*(3), 437–441. doi: 10.1016/0730-725X(93)90078-R
- Brandmaier, A. M., von Oertzen, T., Ghisletta, P., Lindenberger, U., & Hertzog, C. (2018). Precision, Reliability, and Effect Size of Slope Variance in Latent Growth Curve Models: Implications for Statistical Power Analysis. *Frontiers in Psychology*, *9*, 294. doi: 10.3389/fpsyg.2018.00294
- Cheung, G. W., & Rensvold, R. B. (1999). Testing Factorial Invariance across Groups: A Reconceptualization and Proposed New Method. *Journal of Management*, *25*(1), 1–27. doi: 10.1177/014920639902500101
- Ernst, R. R., & Anderson, W. A. (1966). Application of Fourier Transform Spectroscopy to Magnetic Resonance. *Review of Scientific Instruments*, *37*, 93–102.
- Gaser, C., & Dahnke, R. (2016). CAT – A computational anatomy toolbox for the analysis of structural MRI data. *Human brain mapping*, *2016*, 336–348.
- Helms, G., Dathe, H., & Dechent, P. (2008). Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation: Rational Approximation of the FLASH Signal. *Magnetic Resonance in Medicine*, *59*(3), 667–672. doi: 10.1002/mrm.21542
- Helms, G., Dathe, H., Kallenberg, K., & Dechent, P. (2008). High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and  $T_1$  relaxation obtained from 3D FLASH MRI: Saturation and Relaxation in MT FLASH. *Magnetic Resonance in Medicine*, *60*(6), 1396–1407. doi: 10.1002/mrm.21732
- Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. doi: 10.1016/j.neuroimage.2011.09.015
- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, *33*, 99–117. doi: 10.1016/j.dcn.2017.11.007
- Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., & Weiskopf, N. (2010). Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T: Optimized RF Transmit Field Mapping at 3T. *Magnetic Resonance in Medicine*, *64*(1), 229–238. doi: 10.1002/mrm.22421

- Lutti, A., Stadler, J., Josephs, O., Windischberger, C., Speck, O., Bernarding, J., Hutton, C., & Weiskopf, N. (2012). Robust and Fast Whole Brain Mapping of the RF Transmit Field B1 at 7T. *PLoS ONE*, 7(3), e32379. doi: 10.1371/journal.pone.0032379
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051
- Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M. F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., & Mohammadi, S. (2019). HMRI – A toolbox for quantitative MRI in neuroscience and clinical research. *NeuroImage*, 194, 191–210. doi: 10.1016/j.neuroimage.2019.01.029
- Tahedi, M. (2018). *B.A.T.M.A.N.: Basic and Advanced Tractography with MRtrix for All Neurophiles*. doi: 10.17605/OSF.IO/FKYHT
- Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., & Connelly, A. (2019). MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202, 116137. doi: 10.1016/j.neuroimage.2019.116137
- Weiskopf, N., Callaghan, M. F., Josephs, O., Lutti, A., & Mohammadi, S. (2014). Estimating the apparent transverse relaxation time (R2\*) from images with different contrasts (ESTATICS) reduces motion artifacts. *Frontiers in Neuroscience*, 8. doi: 10.3389/fnins.2014.00278
- Weiskopf, N., Lutti, A., Helms, G., Novak, M., Ashburner, J., & Hutton, C. (2011). Unified segmentation based correction of R1 brain maps for RF transmit field inhomogeneities (UNICORT). *NeuroImage*, 54(3), 2116–2124. doi: 10.1016/j.neuroimage.2010.10.023
- Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., Ooi, C., Bullmore, E. T., & Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD\*, MT, and R2\* at 3T: A multi-center validation. *Frontiers in Neuroscience*, 7. doi: 10.3389/fnins.2013.00095
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial Invariance Within Longitudinal Structural Equation Models: Measuring the Same Construct Across Time. *Child Development Perspectives*, 4(1), 10–18. doi: 10.1111/j.1750-8606.2009.00110.x
- Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45(1), S173–S186. doi: 10.1016/j.neuroimage.2008.10.055

**S2: Supplementary results****Bilateral regions of interest**

ROIs were also averaged across left and right hemispheres. Out of the six bilateral GM structural integrity models, two survived testing for factorial invariance: hippocampus and PCC. For these models, all standardized factor loadings were significant ( $p < .050$ ; hippocampus:  $\lambda_{MD} = -0.941$ ,  $\lambda_{MT} = 0.532$ ; PCC:  $\lambda_{MD} = -0.931$ ,  $\lambda_{MT} = 0.471$ ). No differences in mean group change were detected in the hippocampus. A significant maintenance effect of exercise in the bilateral PCC from T2 to T3 was found,  $\Delta\chi^2(1) = 4.80$ ,  $p = .014$ , with the EG showing no mean change in PCC integrity,  $\beta = -0.145$ ,  $SE = 0.357$ ,  $\Delta\chi^2(1) = 0.17$ ,  $p = .339$ , while the ACG showed a significant mean decrease,  $\beta = -1.161$ ,  $SE = 0.597$ ,  $\Delta\chi^2(1) = 10.09$ ,  $p = .001$ . Neither model showed adequate variance in change to test for change-change correlations with cardiovascular fitness.

**S3: Means of observed variables**

Observed variable	ACG			EG		
	T1	T2	T3	T1	T2	T3
VO <sub>2</sub> peak	22.8	—	24.2	23.8	—	25.5
VBM HC right	0.433	0.432	0.434	0.429	0.432	0.427
VBM HC left	0.444	0.445	0.446	0.445	0.450	0.448
VBM ACC right	0.391	0.387	0.385	0.392	0.392	0.391
VBM ACC left	0.404	0.401	0.398	0.389	0.390	0.389
VBM PCC right	0.390	0.387	0.385	0.369	0.367	0.367
VBM PCC left	0.405	0.404	0.401	0.386	0.383	0.382
VBM PCG right	0.317	0.314	0.308	0.308	0.307	0.304
VBM PCG left	0.317	0.315	0.308	0.315	0.313	0.309
VBM JLC right	0.366	0.363	0.360	0.361	0.355	0.355
VBM JLC left	0.366	0.362	0.359	0.356	0.353	0.353
VBM IFG right	0.350	0.347	0.347	0.348	0.357	0.356
VBM IFG left	0.349	0.351	0.348	0.349	0.355	0.354
MT HC right	0.805	0.815	0.811	0.817	0.824	0.803
MT HC left	0.804	0.807	0.811	0.813	0.822	0.798
MT ACC right	0.822	0.827	0.812	0.824	0.831	0.816
MT ACC left	0.819	0.828	0.811	0.819	0.826	0.810
MT PCC right	0.862	0.864	0.851	0.864	0.873	0.867
MT PCC left	0.865	0.871	0.857	0.870	0.878	0.876
MT PCG right	0.879	0.880	0.868	0.877	0.882	0.876
MT PCG left	0.869	0.875	0.860	0.873	0.869	0.864
MT JLC right	0.824	0.836	0.823	0.822	0.821	0.812
MT JLC left	0.828	0.845	0.830	0.831	0.832	0.822
MT IFG right	0.828	0.834	0.825	0.826	0.830	0.829
MT IFG left	0.821	0.836	0.823	0.831	0.828	0.821
MD HC right	0.00170	0.00169	0.00171	0.00165	0.00159	0.00160
MD HC left	0.00175	0.00174	0.00174	0.00169	0.00162	0.00164
MD ACC right	0.00139	0.00140	0.00140	0.00132	0.00128	0.00130
MD ACC left	0.00157	0.00158	0.00159	0.00153	0.00152	0.00152
MD PCC right	0.00130	0.00132	0.00134	0.00130	0.00132	0.00132
MD PCC left	0.00144	0.00146	0.00146	0.00140	0.00142	0.00142
MD PCG right	0.00164	0.00167	0.00169	0.00164	0.00166	0.00166
MD PCG left	0.00163	0.00167	0.00168	0.00163	0.00164	0.00165
MD JLC right	0.00155	0.00159	0.00160	0.00158	0.00161	0.00161
MD JLC left	0.00173	0.00176	0.00178	0.00174	0.00177	0.00179
MD IFG right	0.00184	0.00188	0.00187	0.00182	0.00179	0.00181
MD IFG left	0.00180	0.00181	0.00182	0.00177	0.00176	0.00177

*Note.* ACG = active control group, EG = exercise group, T1 = time point 1 (0 months), T2 = time point 2 (3 months), T3 = time point 3 (6 months), VO<sub>2</sub>peak = peak oxygen uptake, VBM = voxel-based morphometry, MT = magnetization transfer, MD = mean diffusivity, HC = hippocampus, ACC = anterior cingulate cortex, PCC = posterior cingulate cortex, PCG = precentral gyrus, JLC = juxtapositional lobule cortex, IFG = inferior frontal gyrus.

## S4: Pearson correlation coefficients between observed variables at T1

Observed variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.	30.	31.	32.	33.	34.	35.	36.			
1. VO <sub>2</sub> peak																																							
2. VBM HC right	-.11																																						
3. MT HC right	.13	<b>.35*</b>																																					
4. MD HC right	.01	<b>-.71*</b>	<b>-.43*</b>																																				
5. VBM HC left	-.01	.82*	.19	<b>-.64*</b>																																			
6. MT HC left	.17	.38*	.84*	<b>-.44*</b>	<b>.32*</b>																																		
7. MD HC left	-.05	<b>-.68*</b>	<b>-.32*</b>	.89*	<b>-.75*</b>	<b>-.40*</b>																																	
8. VBM ACC right	-.12	.32*	.05	<b>-.36*</b>	.37*	-.09	<b>-.35*</b>																																
9. MT ACC right	.11	.31*	<b>.50*</b>	<b>-.30*</b>	.20	.41*	<b>-.27*</b>	<b>.18</b>																															
10. MD ACC right	-.01	<b>-.43*</b>	<b>-.21</b>	.41*	<b>-.36*</b>	-.16	.46*	<b>-.52*</b>	<b>-.51*</b>																														
11. VBM ACC left	-.01	.37*	-.03	<b>-.30*</b>	.45*	-.10	<b>-.36*</b>	.68*	.04	<b>-.27*</b>																													
12. MT ACC left	.12	.29*	.44*	<b>-.28*</b>	.18	.37*	<b>-.24</b>	.11	.94*	<b>-.42*</b>	<b>.06</b>																												
13. MD ACC left	-.02	<b>-.22</b>	<b>-.17</b>	.29*	-.21	-.11	.32*	-.21	<b>-.41*</b>	.65*	<b>-.26*</b>	<b>-.44*</b>																											
14. VBM PCC right	.08	.36*	.08	<b>-.35*</b>	.33*	-.08	<b>-.31*</b>	.42*	.13	<b>-.27*</b>	.55*	.11	<b>-.23*</b>																										
15. MT PCC right	.08	.25*	.77*	<b>-.30*</b>	.09	.59*	-.21	.19	.73*	<b>-.34*</b>	.03	.67*	<b>-.31*</b>	<b>.12</b>																									
16. MD PCC right	.16	<b>-.39*</b>	-.18	.28*	<b>-.32*</b>	-.06	.28*	<b>-.46*</b>	<b>-.45*</b>	.71*	<b>-.35*</b>	<b>-.40*</b>	.67*	<b>-.45*</b>	<b>-.39*</b>																								
17. VBM PCC left	.17	.28*	.08	<b>-.37*</b>	.33*	-.02	<b>-.32*</b>	.48*	.04	-.21	.51*	-.01	-.10	.86*	.08	<b>-.34*</b>																							
18. MT PCC left	.13	.23	.75*	<b>-.28*</b>	.07	.60*	-.18	.13	.72*	<b>-.32*</b>	-.04	.68*	<b>-.29*</b>	.09	.98*	<b>-.34*</b>	<b>.06</b>																						
19. MD PCC left	-.01	<b>-.34*</b>	<b>-.26*</b>	.28*	<b>-.33*</b>	-.19	.33*	<b>-.40*</b>	<b>-.52*</b>	.71*	<b>-.26*</b>	<b>-.44*</b>	.67*	<b>-.34*</b>	<b>-.41*</b>	.87*	<b>-.33*</b>	<b>-.40*</b>																					
20. VBM PCG right	-.02	.26*	-.17	-.09	.25*	<b>-.25*</b>	-.08	.26*	-.07	-.11	.34*	-.03	-.16	.43*	-.08	<b>-.33*</b>	.33*	-.10	-.21																				
21. MT PCG right	.15	.35*	.68*	<b>-.28*</b>	.18	.62*	-.20	.05	.78*	<b>-.31*</b>	.03	.76*	-.22	.05	.78*	-.23	-.05	.78*	<b>-.30*</b>	<b>-.04</b>																			
22. MD PCG right	.21	-.12	-.12	-.11	.03	.03	-.16	-.13	<b>-.40*</b>	.40*	-.14	<b>-.37*</b>	.49*	-.14	<b>-.38*</b>	.65*	-.08	<b>-.36*</b>	.59*	<b>-.22</b>	<b>-.30*</b>																		
23. VBM PCG left	.00	.31*	-.22	-.15	.33*	<b>-.29*</b>	-.15	.35*	-.10	-.18	.41*	-.06	-.13	.38*	-.10	<b>-.31*</b>	.30*	-.13	-.22	.80*	-.08	-.20																	
24. MT PCG left	.20	.37*	.65*	<b>-.31*</b>	.20	.64*	-.23	.04	.78*	<b>-.33*</b>	.03	.79*	-.24	.03	.70*	-.21	-.07	.73*	<b>-.30*</b>	-.06	.95*	<b>-.25*</b>	<b>-.11</b>																
25. MD PCG left	.25*	-.09	-.08	-.10	.05	.07	-.15	-.18	<b>-.43*</b>	.44*	-.16	<b>-.41*</b>	.52*	-.13	<b>-.35*</b>	.67*	-.03	<b>-.32*</b>	.61*	-.20	<b>-.29*</b>	.95*	<b>-.21</b>	<b>-.27*</b>															
26. VBM JLC right	-.13	.04	-.11	.09	.09	-.22	.10	.27*	.06	-.08	.36*	.05	-.13	.31*	.06	<b>-.35*</b>	.28*	.02	<b>-.28*</b>	.41*	.11	<b>-.38*</b>	.40*	.04	<b>-.36*</b>														
27. MT JLC right	.10	.29*	.57*	<b>-.28*</b>	.12	.50*	-.19	.09	.88*	<b>-.42*</b>	.04	.84*	<b>-.40*</b>	.16	.70*	<b>-.39*</b>	.04	.70*	<b>-.48*</b>	-.04	.83*	<b>-.39*</b>	-.15	.86*	<b>-.43*</b>	<b>.09</b>													
28. MD JLC right	.21	-.01	-.05	-.13	.10	.05	-.15	-.20	<b>-.34*</b>	.43*	-.09	<b>-.32*</b>	.46*	-.09	<b>-.27*</b>	.64*	-.07	<b>-.24*</b>	.60*	<b>-.26*</b>	-.21	.81*	-.22	-.17	.83*	<b>-.37*</b>	<b>-.34*</b>												
29. VBM JLC left	-.17	.18	-.04	-.08	.21	-.14	-.08	.33*	.11	-.10	.36*	.09	-.09	.43*	.12	<b>-.43*</b>	.44*	.06	<b>-.34*</b>	.42*	.09	<b>-.32*</b>	.42*	.02	<b>-.29*</b>	.80*	.08	<b>-.34*</b>											
30. MT JLC left	.08	.26*	.55*	<b>-.24*</b>	.09	.47*	-.16	.10	.86*	<b>-.37*</b>	-.03	.87*	<b>-.36*</b>	.04	.74*	<b>-.32*</b>	-.04	.76*	<b>-.39*</b>	-.07	.85*	<b>-.31*</b>	-.16	.89*	<b>-.34*</b>	.06	.91*	<b>-.28*</b>	<b>.09</b>										
31. MD JLC left	.12	.01	-.08	-.19	.08	.04	-.18	-.20	<b>-.42*</b>	.41*	-.08	<b>-.39*</b>	.46*	-.18	<b>-.37*</b>	.66*	-.10	<b>-.35*</b>	.62*	<b>-.27*</b>	<b>-.28*</b>	.80*	-.20	-.23	.85*	<b>-.42*</b>	<b>-.43*</b>	.83*	<b>-.42*</b>	<b>-.36*</b>									
32. VBM IFG right	.29*	.20	-.22	-.21	.26*	-.21	<b>-.28*</b>	.25*	-.02	<b>-.23*</b>	.41*	-.04	-.15	.44*	-.14	-.15	.36*	-.14	-.10	.48*	-.07	-.02	.48*	-.07	.01	-.03	.09	.22	-.07	.00									
33. MT IFG right	.22	.32*	.68*	<b>-.34*</b>	.21	.67*	<b>-.27*</b>	.04	.70*	-.23	.08	.67*	-.17	.07	.72*	-.16	-.03	.71*	-.23	-.10	.87*	-.14	-.13	.82*	-.16	.03	.74*	-.08	.03	.72*	-.14	<b>-.02</b>							
34. MD IFG right	-.05	<b>-.28*</b>	-.08	.37*	<b>-.26*</b>	-.04	.35*	<b>-.30*</b>	<b>-.40*</b>	.46*	<b>-.29*</b>	<b>-.35*</b>	.50*	<b>-.28*</b>	-.23	.35*	<b>-.25*</b>	-.23	.47*	-.04	<b>-.26*</b>	.41*	-.05	<b>-.25*</b>	.41*	-.15	<b>-.40*</b>	.26*	-.13	<b>-.33*</b>	.30*	<b>-.37*</b>	<b>-.28*</b>						
35. VBM IFG left	.01	.15	-.20	-.15	.28*	-.08	<b>-.25*</b>	.34*	<b>-.27*</b>	-.11	.28*	<b>-.33*</b>	.00	.38*	<b>-.24*</b>	-.14	.39*	<b>-.25*</b>	-.13	.33*	<b>-.28*</b>	.21	.27*	-.24	.20	.08	<b>-.25*</b>	.09	.23*	-.23	.10	.47*	-.16	-.09					
36. MT IFG left	.26*	.32*	.55*	<b>-.28*</b>	.25*	.61*	<b>-.27*</b>	.01	.71*	<b>-.29*</b>	.03	.71*	-.22	.01	.46*	-.13	-.11	.48*	<b>-.24*</b>	-.07	.81*	-.07	-.16	.89*	-.09	-.02	.74*	-.03	-.05	.74*	-.08	-.01	.76*	-.17	<b>-.13</b>				
37. MD IFG left	.13	-.17	-.02	.22	-.14	-.04	.25*	-.21	<b>-.42*</b>	.50*	-.22	<b>-.35*</b>	.56*	-.18	-.18	.44*	-.06	-.18	.47*	.01	-.18	.45*	.07	-.24	.50*	-.03	<b>-.41*</b>	.39*	-.05	<b>-.32*</b>	.36*	-.14	-.20	.64*	<b>-.14</b>	<b>-.26*</b>			

Note. T1 = time point 1 (0 months), VO<sub>2</sub>peak = peak oxygen uptake, VBM = voxel-based morphometry, MT = magnetization transfer, MD = mean diffusivity, HC = hippocampus, ACC = anterior cingulate cortex, PCC = posterior cingulate cortex, PCG = precentral gyrus, JLC = juxtapositional lobule cortex, IFG = inferior frontal gyrus. Correlation coefficients across imaging modalities within regions of interest are in bold for visual clarity.

\*  $p < .050$



**Polk, S. E.**, Kleemeyer, M. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Düzel, S., & Wenger, E. (2022). *Aerobic exercise is associated with region-specific changes in volumetric, tensor-based, and fixel-based measures of white matter integrity in healthy older adults* [Manuscript under revision].





1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24

Aerobic exercise is associated with region-specific changes in volumetric, tensor-based, and  
fixel-based measures of white matter integrity in healthy older adults

Sarah E. Polk<sup>1,2\*</sup>, Maike M. Kleemeyer<sup>1</sup>, Nils C. Bodammer<sup>1</sup>, Carola Misgeld<sup>3</sup>, Johanna  
Porst<sup>3</sup>, Bernd Wolfarth<sup>3</sup>, Simone Kühn<sup>4</sup>, Ulman Lindenberger<sup>1,5</sup>, Sandra Düzel<sup>1</sup> & Elisabeth  
Wenger<sup>1</sup>

<sup>1</sup>Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin,  
Germany

<sup>2</sup>International Max Planck Research School on the Life Course (LIFE), Berlin, Germany

<sup>3</sup>Department of Sports Medicine, Charité – Universitätsmedizin Berlin and Humboldt  
Universität zu Berlin, Berlin, Germany

<sup>4</sup>Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human  
Development, Berlin, Germany

<sup>5</sup>Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin,  
Germany, and London, U.K.

\* Corresponding author: spolk@mpib-berlin.mpg.de  
S.D. and E.W. share senior authorship

## 25 Abstract

26 White matter integrity and cognition have been found to decline with advancing adult age.  
27 Aerobic exercise may be effective in counteracting these declines. Generally, white matter  
28 integrity has been quantified using a volumetric measure (WMV) and with tensor-based  
29 parameters such as fractional anisotropy (FA) and mean diffusivity (MD), the validity of  
30 which appears to be compromised in the presence of crossing fibers. Fixel-based analysis  
31 techniques claim to overcome this problem by yielding estimates of fiber density (FD), cross-  
32 section (FC), and their product (FDC) in multiple directions per voxel. In a sample of 61  
33 healthy older adults aged 63 to 76 years, we quantified changes in white matter integrity  
34 following an aerobic exercise intervention with the commonly used volumetric and tensor-  
35 based metrics (WMV, FA, MD) and with fixel-based metrics (FD, FC, FDC), and  
36 investigated the associations of changes in these white matter parameters to changes in  
37 cardiovascular fitness and Digit Symbol Substitution task (DSST) performance, a marker of  
38 perceptual speed. In line with previous findings, we observed maintained WMV in the corpus  
39 callosum of exercisers, and positive change-change correlations between WMV and fitness  
40 and between WMV and DSST score. For FA and MD, group differences in change opposite  
41 to those hypothesized were found in the corpus callosum, posterior corona radiata, and  
42 superior longitudinal fasciculus. Likewise, regions in the prefrontal cortex showed group  
43 differences in FD and FDC change, but again with more positive change in controls and more  
44 negative change in exercisers. Finally, changes in FD and FDC in superficial WM were found  
45 to be inversely correlated to changes in fitness and DSST performance. The present results  
46 corroborate previous findings of WMV changes, but cast doubt on current physiological  
47 interpretations of both tensor-based and fixel-based indicators of white matter properties in  
48 the context of exercise intervention studies.

49

50 Keywords: aerobic exercise, cardiovascular fitness, aging, white matter integrity, diffusion-  
51 weighted imaging, perceptual speed

52

53

## 54 Highlights

- 55 ● At-home aerobic exercise improves cardiovascular fitness in older adults
- 56 ● Maintenance of corpus callosum volume is related to improved fitness and cognition
- 57 ● Indicators of frontal fiber density and cross-section decrease with exercise
- 58 ● Changes in fiber density indicator and cognition are inversely related
- 59 ● Results raise doubts about content validity of white matter parameter changes

60

61 Data availability statement: The original contributions presented in the study are publicly  
62 available. The data and relevant scripts for analysis can be found here: <https://osf.io/y5u24/>

63

64 Author contribution statement: SEP assisted with data acquisition, preprocessed imaging  
65 data, analyzed the data, interpreted the results, and wrote the manuscript. MMK preprocessed  
66 imaging data and revised the manuscript. NCB designed the neuroimaging protocol and  
67 revised the manuscript. CM and JP performed physical assessments including  
68 cardiopulmonary exercise testing and revised the manuscript. BW designed the physical  
69 assessment protocol and revised the manuscript. SK designed the study and revised the  
70 manuscript. UL designed the study, interpreted the results, and revised the manuscript. SD  
71 designed the study, interpreted the results, and revised the manuscript. EW designed the  
72 study, preprocessed imaging data, interpreted the results, and revised the manuscript.

73

74

## 75 **1. Introduction**

76 The human brain undergoes a range of structural changes as senescence progresses. The  
77 deterioration of white matter integrity, especially in frontal areas of the brain, has been  
78 widely documented using various methods of quantification and both cross-sectional and  
79 longitudinal data (e.g., Beck et al., 2021; Damoiseaux et al., 2009; Gunning-Dixon et al.,  
80 2009; Kelley et al., 2021; Liu et al., 2017; Sexton et al., 2014). This degradation is thought to  
81 contribute to age-related cognitive decline, particularly in the domains of executive function  
82 and processing speed (e.g., Bendlin et al., 2010; Bennett & Madden, 2014; Kennedy & Raz,  
83 2009; Madden et al., 2008).

84 Aerobic exercise and increased cardiovascular fitness have been proposed as a  
85 lifestyle factor intervention to slow or even reverse the deterioration of white matter integrity  
86 and subsequently to improve cognitive function in older adults. A meta-analysis by Sexton  
87 and colleagues (2016) found overall small but significant effects of physical fitness or  
88 activity on white matter volume (WMV). Several cross-sectional studies have found effects  
89 on both global (Benedict et al., 2013; Gow et al., 2012) and local WMV in frontal, temporal  
90 and parietal regions (Erickson et al., 2007; Ho et al., 2011; Tseng, Uh, et al., 2013). A  
91 longitudinal study employing an intervention design found preserved white matter volume in  
92 anterior WMV in older adults who participated in aerobic exercise for six months as  
93 compared to a stretching and toning control group (Colcombe et al., 2006).

94 Regarding measures of white matter micro-structural integrity, one study included in  
95 Sexton et al. (2016) that used whole-brain fractional anisotropy (FA), a diffusion tensor-  
96 derived metric of white matter integrity, found a positive association between levels of  
97 physical activity and global FA values cross-sectionally (Gow et al., 2012). Other cross-  
98 sectional studies investigating the effects of physical fitness or activity on local (i.e., region  
99 of interest-based) FA and mean diffusivity (MD) have found mixed results, with some studies  
100 reporting positive associations with local FA and negative associations with local MD in a  
101 number of regions, including the corpus callosum, superior longitudinal fasciculus, and  
102 corona radiata (Johnson et al., 2012; Z. Liu et al., 2012; Tseng, Gundapuneedi, et al., 2013),  
103 while a number of others reported no significant association (Burzynska et al., 2015; Marks et  
104 al., 2011; Tian et al., 2014). An intervention study comparing an aerobic exercise group to a  
105 stretching and toning group also found no group-level differences in whole-brain or in  
106 regional FA in the prefrontal, temporal, parietal, or occipital cortices after one year (Voss et  
107 al., 2013). However, they did find positive associations between percent change in  
108 cardiovascular fitness and prefrontal and temporal FA, as well as a positive association

109 between change in fitness and change in performance on a short-term memory task. Another  
110 intervention study found global decreases in FA and increase in MD after six months of  
111 exercise, although they did not compare these changes with a control group (Clark et al.,  
112 2019).

113         Recently, the use of the diffusion-tensor metrics of FA and MD as indicators of  
114 greater white matter integrity has been challenged, however. The diffusion-tensor model,  
115 with which these metrics are derived, is not fiber-specific, and is therefore not able to  
116 adequately distinguish between contributions to voxel-wise white matter integrity from  
117 multiple fibers in voxels with more complex multi-fiber geometry, such as crossing fibers  
118 (Raffelt et al., 2012, 2015, 2017) — and 60–90% of voxels are estimated to contain crossing  
119 fibers (Jeurissen et al., 2013). In order to make up for this shortcoming, newer metrics have  
120 been developed using “fixels,” which represent specific fiber bundles within a voxel (Raffelt  
121 et al., 2017). Within each fixel, a fiber orientation distribution (FOD) is computed using  
122 constrained spherical decomposition (Dhollander et al., 2016; Jeurissen et al., 2014). These  
123 FODs are then used to calculate apparent fiber density (FD) and fiber cross-section (FC),  
124 commonly transformed to  $\log(\text{FC})$ , as well as the product of FD and FC, fiber density and  
125 cross-section (FDC). The effects of aerobic exercise on these metrics in the aging brain is yet  
126 unknown, though a number of studies have investigated their association with age. Choy and  
127 colleagues (2020) found that all three metrics showed widespread negative associations with  
128 age in a cross-sectional sample, particularly in anterior regions of the brain. Similarly, Kelley  
129 and colleagues (2021) found that reduced FD, FC, and FDC in older adults versus younger  
130 adults was most prominent in fronto-limbic areas cross-sectionally. Interestingly, they also  
131 found regions in which older adults exhibited greater values than younger adults, including  
132 the superior longitudinal fasciculus, forceps major, and the body of the corpus callosum.

133         A link has also been proposed between aerobic exercise, white matter integrity, and  
134 cognitive performance in aging, given the potential effects of exercise on white matter  
135 integrity as well as the associations between white matter and cognitive outcomes. In a recent  
136 review by Erickson and colleagues (2019), the authors conclude that there is moderate  
137 evidence supporting an association between aerobic exercise and cognition in a number of  
138 domains in older adults (also see Barha et al., 2017), although an earlier review including  
139 twelve randomized controlled trials comparing aerobic exercise and a variety of control  
140 conditions found no evidence for an effect of aerobic exercise on cognition (Young et al.,  
141 2015). The role of white matter integrity in the relationship between aerobic exercise and  
142 cognition is still unclear, as relatively few intervention studies have investigated this question

143 (see Stillman et al., 2020) and even fewer have produced positive results. For example, in the  
144 study by Voss and colleagues (2013), where they found no group effect of exercise on white  
145 matter integrity, researchers also found no evidence for a group effect on short-term memory.

146 The current study aims to investigate the relationship between aerobic exercise and  
147 fitness, a number of white matter metrics, including WMV, diffusion tensor-derived FA and  
148 MD, and fixel-based FD, FC, and FDC, and cognition. The current sample of older adults  
149 participated in a six-month intervention, either in an aerobic exercise group or an active  
150 control group. We expected to find positive changes in WMV in exercisers as compared to  
151 controls. Given previous findings, we also expected to find increased FA values and  
152 decreased MD values as an effect of exercise, or overall decreases in FA and increases in  
153 MD, with no significant exercise effect. Finally, given the inverse association between age  
154 and fixel-based metrics, we expected to see an amelioration of these negative changes as a  
155 result of aerobic exercise. We also explored correlations between cardiovascular fitness,  
156 white matter metrics, and performance on a task indexing perceptual speed and executive  
157 function. Given previous findings of the associations between white matter integrity and  
158 perceptual speed in both younger (Magistro et al., 2015) and older adults (Papp et al., 2015),  
159 as well as previously documented links between cardiovascular fitness and perceptual speed  
160 (Prakash et al., 2010), we expected to see positive relationships between greater  
161 cardiovascular fitness, indicators of greater white matter integrity, and better performance on  
162 a task requiring perceptual speed.

## 163 **2. Materials and Methods**

### 164 **2.1 Sample and Study Design**

165 In the current analyses, we investigated the effects of aerobic exercise on white matter  
166 integrity in previously sedentary older adults by comparing individuals who participated in a  
167 physical training intervention group to those who did not engage in exercise. We used a  
168 subset of data from the AKTIV study, which investigated cognitive and physical exercise  
169 intervention effects in older adults. A full description of subject recruitment, intervention  
170 design, and all acquired measures can be found in Wenger et al. (2021). We repeat the  
171 relevant details for the current analyses here.

172 Healthy older adults from 63 to 78 years old were recruited if they met none of the  
173 following exclusion criteria: magnetic resonance imaging (MRI) contraindications; they  
174 could not meet the time requirements of the study; not right-handed; younger than 63 or older  
175 than 78 years old at the start of the study; engaging in aerobic exercise more than once every  
176 two weeks; fluent in a language other than German or English, or fluent in more than two

177 languages; or receiving medical treatment for Parkinson's, gout, rheumatism, heart attack,  
178 stroke, cancer, severe back problems, severe arrhythmia, severe chronic liver or kidney  
179 failure, severe disease of the hematopoietic system, mental illness (e.g., depression), or  
180 neurological disease (e.g., epilepsy, brain tumor).

181 Before the start of the intervention, participants first underwent a physical assessment  
182 including cardiopulmonary exercise testing (CPET) at the Charité – Universitätsmedizin  
183 Berlin, then came to the Max Planck Institute for Human Development, Berlin for a baseline  
184 MRI session and cognitive testing (T1). Out of the 201 recruits invited to participate, 41  
185 dropped out or were excluded due to existing medical conditions or claustrophobia in the  
186 scanner before the training. Participants trained at home in one of four intervention groups  
187 (active control, language, aerobic exercise, or combined language and aerobic exercise) for  
188 three months before being scanned a second time with the same sequences and completing  
189 the same cognitive battery at T2, and after a total of six months of at-home training,  
190 participants underwent MRI, cognitive testing, and a physical assessment under the same  
191 physician a final time (T3). A further 17 participants dropped out during the training citing  
192 physical complaints (i.e., pain during exercise), disinterest, time constraints, or unspecified  
193 reasons.

194 The ethics committee of the German Psychological Society (DGPs) approved the  
195 study and written informed consent was collected from all participants.

## 196 **2.2 Interventions**

197 We focus on the effects of an aerobic exercise intervention here, comparing the  
198 exercise-only group (EG) to the active control group (ACG).

199 Forty participants completed the study in the EG (mean age = 69.8 years, 50%  
200 females). Aerobic exercise was implemented with a stationary bicycle (DKN Ergometer AM-  
201 50) which was synchronized with a tablet (Lenovo TB2-X30L TAB) via Bluetooth. Using  
202 this tablet, participants could access their personalized interval training program, the initial  
203 level of which was determined by the sports medicine physician (30 minutes at 25–140  
204 Watts,  $M = 67.8$ ,  $SD = 26.65$ ). Participants were asked to exercise three to four times a week  
205 with no restrictions as to time of day. After each session, participants indicated their  
206 perceived exertion via the Borg Rating of Perceived Exertion Scale, which includes ratings  
207 from 6 (no exertion at all) to 20 (maximal exertion). If participants indicated a rating below  
208 12 (too easy) or above 15 (too difficult), the intensity of the training could be adjusted  
209 remotely. Training intensity increased automatically approximately every two weeks by three  
210 minutes and three to four Watts. A collection of pre-selected literature was also available on



211 the tablet, and participants were instructed to read at a slow pace for 15 minutes on days  
212 when they completed an exercise session or for 45 minutes on days when they did not. In  
213 total, participants engaged in some study-related activity for approximately 45 minutes a day  
214 for at least six days each week. Finally, in-person group sessions of five to ten individuals  
215 each were conducted once a week, during which participants in the EG engaged in a  
216 stretching and toning course, led by an external instructor. Adherence to the aerobic exercise  
217 intervention was defined as engaging in an average of 90 minutes of exercise a week for at  
218 least 21 weeks ( $\geq 1890$  minutes total) with no pauses of longer than two weeks, as well as a  
219 slight steady increase in training difficulty over the course of the study, as was automatically  
220 implemented by the interval training application.

221 Thirty-five participants completed the study in the ACG (mean age = 70.7 years, 40%  
222 females). They also received a tablet and were asked to read the selected literature for 45  
223 minutes a day on at least six days out of the week. In-person group sessions for participants in  
224 the ACG consisted of a book club, where groups discussed short stories led by external  
225 facilitators (<http://shared-reading.de/>). Adherence in the ACG was defined as at least 1890  
226 total minutes of reading during the study.

## 227 **2.3 Data Acquisition**

### 228 **2.3.1 Cardiovascular fitness**

229 Cardiovascular fitness was measured as peak oxygen uptake, or  $VO_{2peak}$ , relativized  
230 by body weight in kilograms, using CPET with a bicycle ergometer (Ergoselect 100k,  
231 Ergoline GmbH, Bitz, Germany) and the Quark Clinical-based Metabolic Cart using the  
232 standard Breath-by-Breath setup and the V2Mask (Hans Rudolph, Inc.).

### 233 **2.3.2 Magnetic resonance imaging**

#### 234 **2.3.2.1 Acquisition**

235 MR images were acquired on a 3T Magnetom Tim Trio MRI scanner system  
236 (Siemens Medical Systems, Erlangen) using a 32-channel radiofrequency head coil.  $T_1$ -  
237 weighted images were obtained using a 3D  $T_1$ -weighted magnetization prepared gradient-  
238 echo (MPRAGE) sequence with the following parameters: repetition time (TR) = 2500 ms;  
239 echo time (TE) = 4.77 ms; inversion time (TI) = 1100 ms; flip angle =  $7^\circ$ ; acquisition  
240 matrix =  $256 \times 256 \times 192$ ;  $1 \text{ mm}^3$  isotropic voxels; with the prescan normalize option and a  
241 3D distortion correction for non-linear gradients; acquisition time = 9:20 min. Diffusion-  
242 weighted images were obtained with a single-shot diffusion-weighted spin-echo-refocused  
243 echo-planar imaging sequence with the following parameters: TR = 9700 ms; TE = 120 ms;  
244 62 slices; FOV =  $224 \times 224 \text{ mm}$ ; a two-shell scheme was used for diffusion-weighting



245 applying two  $b$ -values: 710 s/mm<sup>2</sup> (30 directions) and 2850 s/mm<sup>2</sup> (60 directions), with  
246 directions distributed over a whole sphere for each shell, plus ten non-diffusion-weighted  
247 images; GRAPPA acceleration factor = 2; 2 mm<sup>3</sup> isotropic voxels; acquisition time = 16:41  
248 min. Six additional images inverting the phase encoding direction were acquired without  
249 diffusion weighting; acquisition time = 1:29 min.

#### 250 2.3.2.2 Preprocessing and calculation of voxel-wise values

251 T<sub>1</sub>-weighted images were preprocessed using the Computational Anatomy Toolbox 12  
252 (CAT12, Structural Brain Mapping group, Jena University Hospital) in Statistical Parametric  
253 Mapping (SPM12, Institute of Neurology). The longitudinal preprocessing pipeline with  
254 default parameters was used, in which within-subject inverse-consistent realignment with  
255 intra-subject bias correction is first applied to calculate a mean image for each participant  
256 before data segmentation. DARTel normalization was then used to estimate spatial  
257 normalization parameters based on the segmented images of the within-subject mean image,  
258 which were applied to the segmented images at each time point. Nonlinear-only modulation  
259 of gray and white matter segmentations was applied. Images were smoothed using an 8-mm  
260 full-width half-maximum (FWHM) standard Gaussian kernel.

261 Diffusion-weighted images were preprocessed using MRtrix (version 3.0\_RC3;  
262 Tournier et al., 2019), FSL (FMRIB's Software Library, version 6.0.2; Jenkinson et al., 2012;  
263 Smith et al., 2004; Woolrich et al., 2009), ANTS (version 2.2.0; Avants et al., 2010, 2011),  
264 following the Basic and Advanced Tractography with MRtrix for All Neurophiles  
265 (B.A.T.M.A.N.) tutorial (Tahedl, 2018). Images were denoised with a 5 × 5 × 5 patch and  
266 Gibb's ringing artifacts were removed using *dwidenoise* and *mrdegibbs* in MRtrix.  
267 Corrections for EPI distortion, B<sub>0</sub>-field inhomogeneity, and eddy-current and movement  
268 distortion using *topup* and *eddy\_cuda* in FSL with a quadratic spatial model for *eddy*,  
269 replacing outliers and saving contrast-to-noise ratio and residual maps. Binary brain masks  
270 were generated from each individual's mean b0 image using *bet* in FSL; these maps were  
271 visually inspected and manually adjusted where necessary.

272 To create FA and MD maps, we followed the TBSS User Guide from FSL  
273 (<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TBSS/UserGuide>; Smith et al., 2004, 2006). First, FA  
274 images were slightly eroded and end slices were zeroed with *tbss\_1\_preproc*. Nonlinear  
275 registration was then run and FA images were aligned to a 1mm<sup>3</sup> standard space using the  
276 JHU ICBM FA 1mm<sup>3</sup> atlas provided by FSL (<https://identifiers.org/neurovault.image:1402>)  
277 with *tbss\_2\_reg*. Following this, nonlinear transforms generated by the previous step were  
278 applied to all subjects to bring them into standard space with *tbss\_3\_postreg*, resulting in 4D

279 FA data for all subjects. MD maps were generated in a similar manner using *tbss\_non\_FA*:  
280 the nonlinear registration from the FA pipeline was also applied to the MD images and the  
281 resulting data were merged into a 4D image. The resulting FA and MD images were then  
282 smoothed in SPM using a 4mm FWHM standard Gaussian kernel for analysis in SPM.

283 Following preprocessing, FD, log(FC), and FDC were calculated following the “Fibre  
284 density and cross-section – Multi-tissue CSD” tutorial from the MRtrix3 documentation  
285 ([https://mrtrix.readthedocs.io/en/latest/fixel\\_based\\_analysis/mt\\_fibre\\_density\\_cross-](https://mrtrix.readthedocs.io/en/latest/fixel_based_analysis/mt_fibre_density_cross-section.html)  
286 [section.html](https://mrtrix.readthedocs.io/en/latest/fixel_based_analysis/mt_fibre_density_cross-section.html); Tournier et al., 2019). First, average tissue response functions were computed  
287 using *dwi2response* and *responsemean*. Multi-shell, multi-tissue CSD was then performed to  
288 estimate the FOD using *dwi2fod msmt\_csd*. Joint bias field correction and global intensity  
289 normalization were run using *mtnormalise*. Intra-subject templates were created from the  
290 FOD images from all time points using *population\_template*, to which all images within each  
291 subject were registered with *mrregister*. An unbiased, study-specific FOD template was then  
292 created using 40 randomly selected participants (ten from each intervention group) with  
293 *population\_template*, to which each within-subject FOD template was registered with  
294 *mrregister*. A template mask of voxels containing data in all images was created using  
295 *mrtransform* and *mrmath*, on which fixel segmentation was performed with *fod2fixel* and the  
296 default threshold of 0.06 to compute a WM template fixel mask. Intra-subject and study-  
297 specific warps were then combined using *transformcompose*, and were subsequently applied  
298 to the individual unwarped FOD images with *mrtransform* to bring them into common space  
299 without FOD reorientation. Each subject’s FOD image was then segmented to estimate fixels  
300 and their FD using *fod2fixel*, fixels of the subject FOD images were reoriented in template  
301 space using *fixelreorient*, and subject fixels were assigned to template fixels using  
302 *fixelcorrespondence* to establish which fixels correspond across subjects, as well as between  
303 subject and template fixels. Next, the fiber cross-section metric, FC, was calculated for each  
304 subject, as well as the log(FC) metric using *warp2metric* and *mrcalc*, and the combined  
305 measure of FDC was calculated by multiplying the FD and FC values from the previous steps  
306 using *mrcalc*. Finally, in order to conduct repeated measures ANOVA on the fixel-based  
307 metrics in SPM, voxel-wise metrics were calculated for FD, log(FC), and FDC using  
308 *fixel2voxel*. For FD and FDC, this was achieved by summing the fixel-wise values across  
309 directions to calculate total FD and FDC per voxel; for log(FC), a weighted mean based on  
310 the FD values was calculated across directions, resulting in an average voxel-wise log(FC)  
311 where the log(FC) value of the direction with the greatest FD was weighted most heavily.  
312 These voxel-wise metrics were converted to Nifti format for longitudinal analysis in SPM

313 using *mrconvert*. Finally, the FD, log(FC), and FDC maps were smoothed in SPM using a  
314 10mm FWHM standard Gaussian kernel.

### 315 **2.3.2 Digit Symbol Substitution Task**

316 Cognitive function, specifically perceptual speed was assessed with the Digit Symbol  
317 Substitution task (DSST; Wechsler, 1981). The DSST consists of a key of nine unique digit-  
318 symbol pairs, and rows of unpaired digits. Participants are asked to complete as many pairs as  
319 possible with the corresponding symbol within 90 seconds. Each correct answer is scored as  
320 1, one incorrect answer is counted as 0, and after two consecutive incorrect answers,  
321 responses are no longer counted.

## 322 **2.4 Statistical Analyses**

### 323 **2.4.1 Repeated measures ANOVA**

324 To investigate group differences in change in VO<sub>2</sub>peak and DSST score, ANOVAs  
325 with time point as a within-subject factor (T1, T2, T3) and group as a between-subject factor  
326 (ACG, EG) were conducted, with age, sex, and education included as covariates. These were  
327 run using commands from the *rstatix* package (Kassambara, 2021) in R (R Core Team,  
328 2021), version 4.1.2 (2021-11-01), using RStudio (RStudio Team, 2021), version  
329 2021.09.2+382. Post-hoc t-tests were run using base R *stats* commands.

### 330 **2.4.2 Flexible factorial analysis investigating time-by-group interactions**

331 To investigate group differences in change in the white matter metrics, flexible  
332 factorial models in SPM12 were used to compute voxel-wise statistics. This model, in  
333 contrast to the permutation-based models typically used in TBSS or FBA, can account for the  
334 fact that an individual's scans at different time points are not independent of one another.  
335 Smoothed white matter volume maps, smoothed FA and MD maps, and smoothed voxel-wise  
336 FD, log(FC), and FDC maps were entered into flexible factorial models with subject as a  
337 within-subject factor, time point as a within-subject factor (T1, T2, T3), and group as a  
338 between-subject factor (ACG, EG). Age, sex, and education were entered into the model as  
339 covariates of no interest. We tested for a time-by-group interaction to investigate whether  
340 changes across time points differed between groups. A threshold of  $p < .050$  with correction  
341 for false discovery rate (FDR) at the peak-level was applied first, and if no significant clusters  
342 were revealed, a more liberal threshold of  $p < .001$ , uncorrected, was applied. In all cases,  
343 correction for non-isotropic smoothness was applied in CAT12 with a cluster extent threshold  
344 of  $k > 100$ . Missing data was excluded case-wise: no white matter volume data were missing,  
345 four cases were excluded from the diffusion tensor-derived and fixel-based metrics due to  
346 missing scans at one time point.

347 To investigate the directions of effects found, within-subject mean values were  
 348 extracted from significant clusters using the REX: Response Exploration for Neuroimaging  
 349 Datasets toolkit in MATLAB (Duff et al., 2007). Paired t-tests were conducted in R on these  
 350 within-subject means to inspect within-group changes.

#### 351 **2.4.4 Correlations at baseline and change-change correlations**

352 Finally, the relationships between VO<sub>2</sub>peak, white matter integrity metrics, and DSST  
 353 score were investigated using Pearson correlations with the *Hmisc* R package (Harrell, 2021)  
 354 and differences between correlations calculated within-group were examined with the *cocor*  
 355 R package (Diedenhofen & Musch, 2015). Baseline correlations were calculated, as well as  
 356 correlations between percent change from T1 to T3 in VO<sub>2</sub>peak, extracted white matter  
 357 metrics from clusters showing group differences in change, and DSST score. Missing data  
 358 points were excluded pair-wise. Correction for FDR was applied to baseline correlations and  
 359 change-change correlations separately to correct for multiple comparisons.

### 360 **3. Results**

361 A description of the sample can be found in Table 1. Participants who did not meet  
 362 compliance criteria were excluded and one further exercise participant was excluded due to  
 363 technical difficulties. This resulted in  $n_{ACG} = 32$  and  $n_{EG} = 29$  included in the analyses.

364 **Table 1**

*Sample demographics and intervention specifics.*

	Active control group	Exercise group
<i>n</i> completed intervention	35	40
<i>n</i> fully adhered	32	29
Age at baseline, M/SD (range)	70.8 ± 3.93 (64.0–76.0)	70.2 ± 3.59 (63.9–76.9)
Sex, % of female participants	40.6	58.6
Years of education, M/SD (range)	13.4/3.15 (7–16)	13.0/3.14 (7–16)
Total minutes spent in intervention, M/SD (range)	4772/1819.6 (2505–10858)	6554/1222.9 (4098–9790)
Minutes spent reading, M/SD (range)	4772/1819.6 (2505–10858)	3381/1143.3 (915–5855)
Minutes spent exercising, M/SD (range)	–	3173/410.4 (2556–3937)

*Note.* M = mean; SD = standard deviation. Age, sex, years of education, and total minutes spent in intervention were calculated among those participants who fully adhered to the intervention and were included in the current analyses.

365

366

**Table 2**

Means and standard deviations of variables of VO<sub>2</sub>peak, Digit Symbol Substitution task score, and white matter metrics extracted from clusters showing significant time-by-group interactions.

Measure	Active control group			Exercise group		
	T1	T2	T3	T1	T2	T3
VO <sub>2</sub> peak (mL/kg/min)	23.5 ± 6.01	—	24.4 ± 6.52	22.9 ± 6.03	—	25.5 ± 6.21
DSST score	46.8 ± 8.41	47.3 ± 7.76	47.5 ± 8.83	46.1 ± 10.76	44.7 ± 10.65	49.1 ± 11.27

Metric	Peak voxel			Active control group			Exercise group		
	x	y	z	T1	T2	T3	T1	T2	T3
WMV	16	28	2	0.568 ± 0.0948	0.564 ± 0.0945	0.554 ± 0.0968	0.536 ± 0.0892	0.539 ± 0.0893	0.534 ± 0.0896
WMV	-21	-45	16	0.533 ± 0.0836	0.529 ± 0.0843	0.521 ± 0.0837	0.511 ± 0.0759	0.512 ± 0.0750	0.511 ± 0.0751
FA	-10	27	8	0.334 ± 0.0315	0.343 ± 0.0388	0.346 ± 0.0365	0.352 ± 0.0366	0.353 ± 0.0394	0.344 ± 0.0364
MD	26	-38	28	9.12×10 <sup>-04</sup> ± 9.06×10 <sup>-05</sup>	9.03×10 <sup>-04</sup> ± 8.19×10 <sup>-05</sup>	9.01×10 <sup>-04</sup> ± 8.55×10 <sup>-05</sup>	8.92×10 <sup>-04</sup> ± 5.24×10 <sup>-05</sup>	8.98×10 <sup>-04</sup> ± 5.42×10 <sup>-05</sup>	9.04×10 <sup>-04</sup> ± 5.44×10 <sup>-05</sup>
MD	34	-25	28	8.32×10 <sup>-04</sup> ± 5.73×10 <sup>-05</sup>	8.21×10 <sup>-04</sup> ± 5.58×10 <sup>-05</sup>	8.17×10 <sup>-04</sup> ± 6.31×10 <sup>-05</sup>	8.13×10 <sup>-04</sup> ± 4.93×10 <sup>-05</sup>	8.21×10 <sup>-04</sup> ± 5.09×10 <sup>-05</sup>	8.23×10 <sup>-04</sup> ± 5.00×10 <sup>-05</sup>
FD	3	65	26	0.0361 ± 0.00356	0.0365 ± 0.00332	0.0383 ± 0.00388	0.0368 ± 0.00381	0.0358 ± 0.00367	0.0350 ± 0.00319
FD	39	78	25	0.0100 ± 0.00280	0.0105 ± 0.00302	0.0115 ± 0.00358	0.0099 ± 0.00219	0.0097 ± 0.00211	0.0095 ± 0.00246
FDC	3	65	26	0.0480 ± 0.00875	0.0484 ± 0.00732	0.0503 ± 0.00843	0.0481 ± 0.00836	0.0468 ± 0.00864	0.0457 ± 0.00685
FDC	38	78	25	0.0117 ± 0.00390	0.0123 ± 0.00426	0.0135 ± 0.00489	0.0111 ± 0.00272	0.0109 ± 0.00264	0.0106 ± 0.00283

Note. DSST = Digit Symbol Substitution task; WMV = white matter volume; FA = fractional anisotropy; MD = mean diffusivity (mm<sup>2</sup>/s); FD = fiber density; FDC = fiber density and cross-section. FD and FDC are registered to a study-specific template.

367

### 3.1 Group differences in cardiovascular fitness change

Means and standard deviations of  $VO_2$ peak at T1 and T3 for each group are reported in Table 2. A repeated measures ANOVA including age, sex, and education as covariates revealed a significant time-by-group interaction in  $VO_2$ peak,  $F(1, 53) = 6.091, p = .017$ , Hedge's  $g = 0.009$ . Post-hoc pairwise t-tests indicated a significant increase in  $VO_2$ peak within exercisers,  $t(28) = 4.959, p < .001$ , with a mean percent change of 12.8% (SE = 2.28), but not within controls,  $t(29) = 1.279, p = .211$ , with a mean percent change of 3.7% (SE = 2.22).

### 3.2 Group differences in white matter integrity changes

#### 3.2.1 Whole-brain white matter volume

The contrast investigating group differences in change in WMV, controlling for age, sex, and education, revealed two significant clusters at  $p_{FDR} < .050$  with a cluster size threshold of  $k > 100$  (see Figure 1). One cluster was found in the right anterior corona radiata extending into the genu of the corpus callosum (1000 voxels); the EG showed no significant change in mean WMV within this cluster from T1 to T3,  $t(28) = -1.655, p = .109$ , while controls decreased significantly,  $t(31) = -7.682, p < .001$ . The other cluster was localized in the splenium of the corpus callosum (542 voxels); in this cluster, the EG again showed no significant change,  $t(28) = -0.170, p = .866$ , while the ACG showed a significant decrease,  $t(31) = -6.581, p < .001$ .

#### 3.2.2 Whole-brain fractional anisotropy

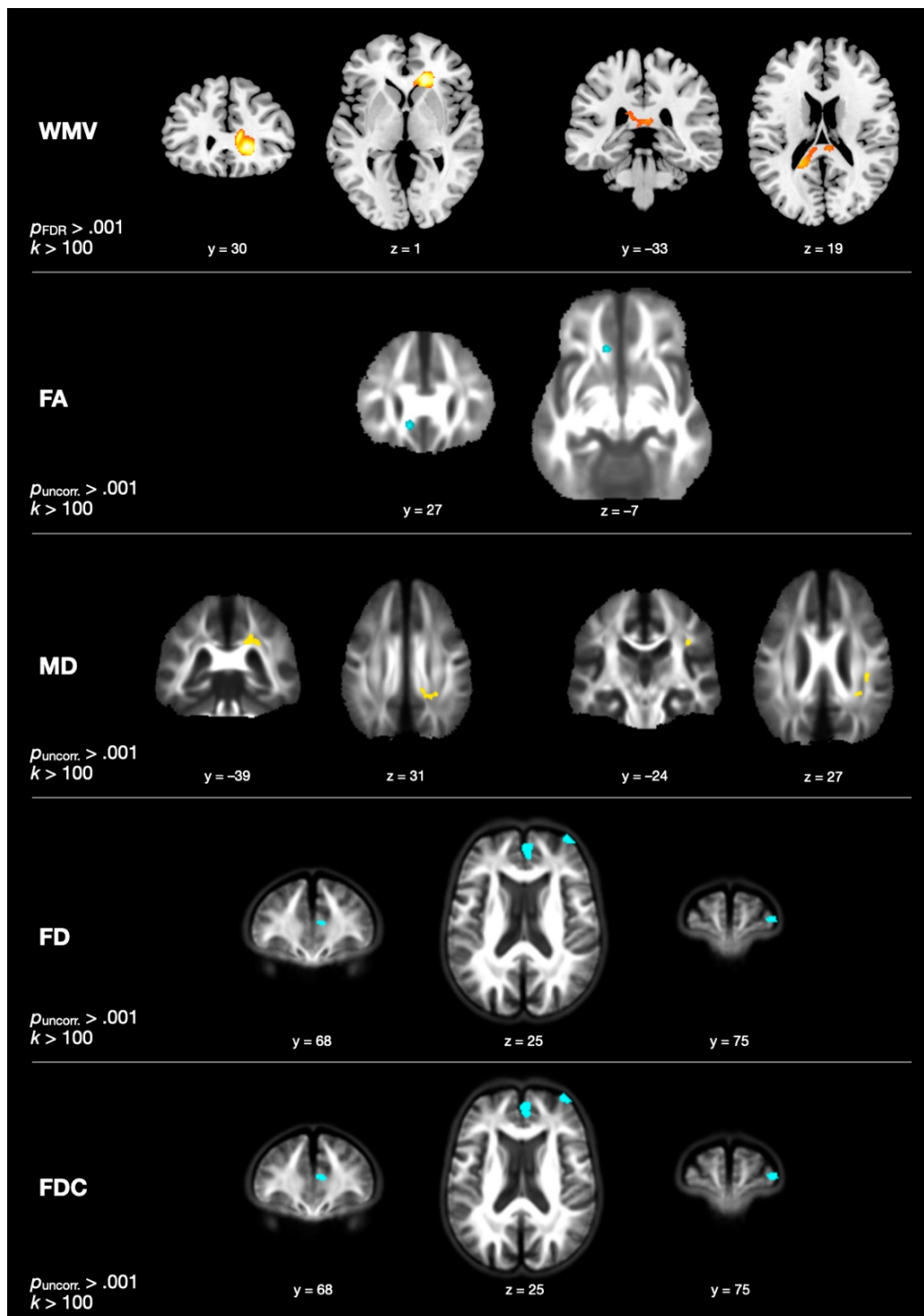
No clusters of significant time-by-group differences in FA were found at a threshold of  $p_{FDR} < .050, k > 100$ . At  $p_{uncorrected} < .001$ , one cluster was revealed (see Figure 1). In this cluster in the left part of the genu of the corpus callosum (136 voxels), the EG showed a significant decrease,  $t(28) = -4.2289, p < 0.001$ , while the ACG showed a significant increase in FA,  $t(27) = 5.149, p < 0.001$ .

#### 3.2.3 Whole-brain mean diffusivity

Regarding changes in MD, no clusters survived the threshold of  $p_{FDR} < .050$ , but two clusters showing a significant time-by-group interaction were revealed at the more lenient threshold of  $p_{uncorrected} < .001$  (see Figure 1). One cluster was found in the right posterior corona radiata extending into the splenium of the corpus callosum (358 voxels), in which the EG showed a significant increase in mean MD,  $t(28) = 4.530, p < .001$ , and the ACG showed a significant decrease,  $t(27) = -2.663, p = .013$ . The other cluster was found in the right superior longitudinal fasciculus (100 voxels), in which the EG again showed a significant



401 increase in mean MD,  $t(28) = 2.450, p = .021$ , and the ACG showed a significant decrease,  
 402  $t(27) = -3.458, p = .002$ .



403

404 **Figure 1**

405 Group differences in change in white matter volume (WMV), fractional anisotropy (FA), mean  
 406 diffusivity (MD), fiber density (FD), fiber density and cross-section (FDC). Yellow/orange-colored  
 407 clusters represent more positive changes in exercisers than controls, cyan-colored clusters represent  
 408 more negative changes in exercisers than controls. WMV, FA, and MD are calculated and displayed  
 409 in MNI space, whereas FD and FDC are calculated and displayed in a sample-specific space.

### 410 **3.2.4 Whole-brain fiber density**

411 Testing for time-by-group interactions in FD revealed two significant clusters at  
412  $p_{\text{uncorrected}} < .001$  larger than  $k = 100$ , neither of which survived FDR correction (see Figure 1).  
413 One cluster was localized in the right dorsomedial prefrontal cortex (dmPFC, 338 voxels),  
414 and another was seen in the right dorsolateral prefrontal cortex (dlPFC, 105 voxels). In the  
415 dmPFC, the EG showed a significant decrease in mean FD from T1 to T3,  $t(28) = -3.472$ ,  $p =$   
416  $.002$ , while the ACG showed a significant increase,  $t(27) = 4.255$ ,  $p < .001$ . In the dlPFC, the  
417 EG showed no change in mean FD,  $t(28) = -1.556$ ,  $p = .131$ , and the ACG showed a  
418 significant increase,  $t(27) = 3.545$ ,  $p = .001$ .

### 419 **3.2.5 Whole-brain fiber cross-section**

420 No significant clusters were revealed when testing for group differences in change in  
421  $\log(\text{FC})$ , either at the initial threshold of  $p_{\text{FDR}} < .050$  or the more liberal threshold of  $p_{\text{uncorrected}}$   
422  $< .001$ .

### 423 **3.2.6 Whole-brain fiber density and cross-section**

424 Finally, two clusters were found in which there were significant group differences in  
425 change in FDC at  $p_{\text{uncorrected}} < .001$  (see Figure 1), both of which almost entirely overlapped  
426 with those found in FD: one in the right dorsomedial prefrontal cortex (dmPFC, 386 voxels)  
427 and one in the right dorsolateral prefrontal cortex (dlPFC, 130 voxels). These clusters did not  
428 survive FDR correction either. The pattern of within-group change mirror that in FD: in the  
429 dmPFC, the EG decreased significantly from T1 to T3,  $t(28) = -3.208$ ,  $p = .003$ , while the  
430 ACG increased,  $t(27) = 3.931$ ,  $p < .001$ . In the dlPFC, the EG showed no significant change,  
431  $t(28) = -1.705$ ,  $p = .099$ , and the ACG showed a significant increase,  $t(27) = 3.399$ ,  $p = .002$ .  
432 Means and standard deviations of extracted mean values from each cluster showing a  
433 significant time-by-group interaction as well as peak voxel coordinates can be found in Table  
434 2.

## 435 **3.3 Cognition**

436 DSST score means and standard deviations are reported in Table 2. No significant  
437 time-by-group effects were found in DSST when controlling for age, sex, and education,  
438  $F(2,102) = 2.696$ ,  $p = .072$ , Hedge's  $g = 0.010$ .

## 439 **3.4 Correlations**

440 Baseline correlations between  $\text{VO}_2\text{peak}$ , DSST, and extracted mean values within  
441 each of the clusters showing significant time-by-group differences, as well as correlations  
442 between percent change in white matter metrics, all corrected for FDR, can be found in Table  
443 3.



444

445 **Table 3**

*Baseline and percent change correlations between VO<sub>2</sub>peak, Digit Symbol Substitution task score, and white matter metrics extracted from clusters showing significant time-by-group interactions.*

Baseline	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. VO <sub>2</sub> peak										
2. DSST	.11									
3. WMV rACR	.40*	.17								
4. WMV splenium	.33*	.01	.77*							
5. FA l genu	-.01	.14	.20	.20						
6. MD rPCR/splenium	-.05	-.03	-.16	-.18	-.53*					
7. MD rSLF	-.32*	-.11	-.34*	-.21	-.38*	.40*				
8. FD dmPFC	-.07	.05	.28	.03	.43*	-.22	-.32*			
9. FD dlPFC	.25	.09	.13	.10	.14	-.08	-.26	.30		
10. FDC mvPFC	.07	.01	.70*	.48*	.31*	-.22	-.31*	.78*	.25	
11. FDC dlPFC	.31*	.04	.46*	.38*	.18	-.09	-.27	.40*	.88*	.51*
Percent change	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. VO <sub>2</sub> peak										
2. DSST	.02									
3. WMV rACR	.18	.34*								
4. WMV splenium	.33*	.26	.65*							
5. FA l genu	-.23	.01	-.21	-.26						
6. MD rPCR/splenium	.17	.20	.18	.32*	-.39*					
7. MD rSLF	.06	-.14	.13	.10	-.42*	.47*				
8. FD dmPFC	-.32*	-.12	-.38*	-.41*	.46*	-.41*	-.28			
9. FD dlPFC	-.09	-.31	-.17	-.24	.37*	-.52*	-.46*	.31*		
10. FDC mvPFC	-.30*	-.12	-.36*	-.40*	.41*	-.38*	-.26	.98*	.26	
11. FDC dlPFC	-.09	-.30	-.19	-.26	.36*	-.50*	-.46*	.30*	1.00*	.25

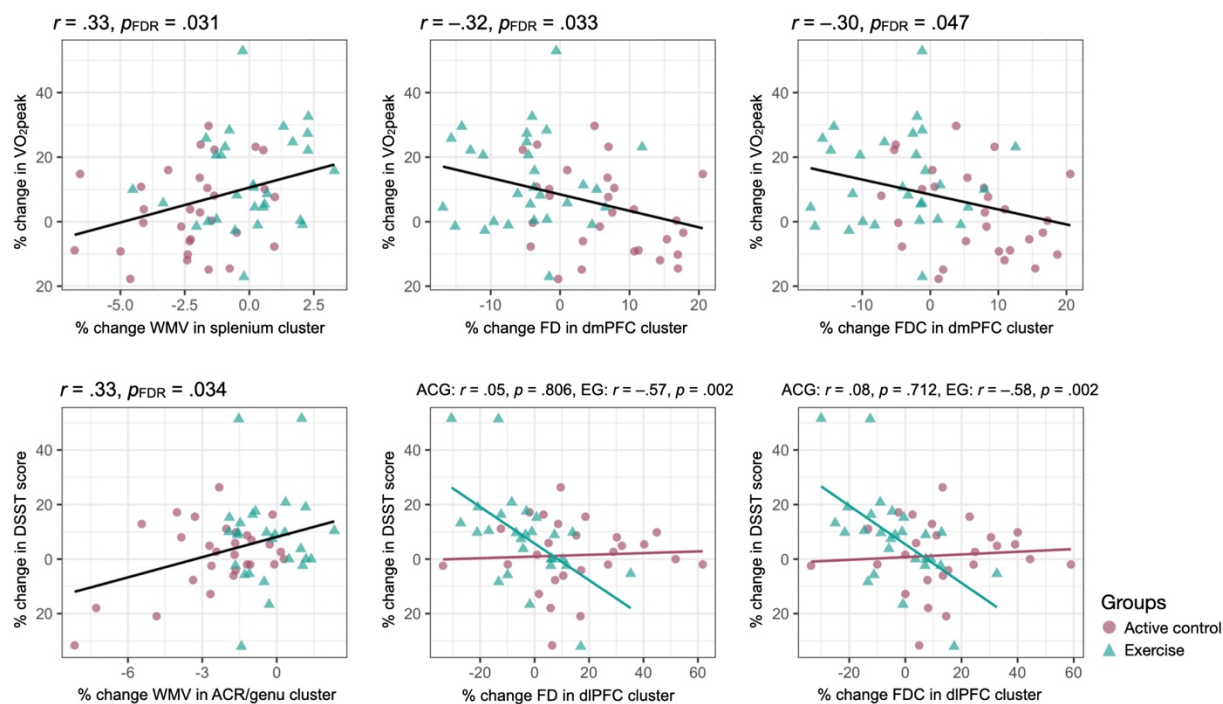
*Note.* DSST = Digit Symbol Substitution task; WMV = white matter volume; FA = fractional anisotropy; MD = mean diffusivity; FD = fiber density; FDC = fiber density and cross-section; r = right; l = left; ACR = anterior corona radiata; PCR = posterior corona radiata; SLF = superior longitudinal fasciculus; dmPFC = dorsomedial prefrontal cortex; dlPFC = dorsolateral prefrontal cortex.

\* significant at  $p_{FDR} < .05$

446

447           Regarding correlations with percent change in cardiovascular fitness, a positive  
448 correlation was found between percent change in VO<sub>2</sub>peak and percent change in WMV in  
449 the splenium of the corpus callosum,  $r(57) = .33$ ,  $p_{FDR} = .029$ . Negative correlations were  
450 found between percent change in VO<sub>2</sub>peak and percent change in FD in the dmPFC,  $r(55) = -$   
451  $.32$ ,  $p_{FDR} = .034$ , and percent change in FDC in the dmPFC,  $r(55) = -.30$ ,  $p_{FDR} = .047$ . No  
452 differences between correlation coefficients calculated within each group were found.

453



454

455 **Figure 2**

456 Significant correlations between percent change in  $VO_{2peak}$ , white matter integrity metrics extracted  
 457 from clusters with significant time-by-group interactions, and Digit Symbol Substitution task score.

458 Overall correlations are shown in black for those correlations that do not show significant group  
 459 differences. Correlations that differ significantly between groups are represented by color-coded lines.

460 WMV = white matter volume; FD = fiber density; FDC = fiber density and cross-section; dmPFC =  
 461 dorsomedial prefrontal cortex; DSST = Digit Symbol Substitution task; ACR = anterior corona radiata;

462 dlPFC = dorsolateral prefrontal cortex. Removing the two visual outliers in percent change in DSST  
 463 score results in a correlation of  $r = .33$ ,  $p_{uncorrected} = .015$  between percent change in WMV in the

464 cluster found in the rACR and percent change in DSST score, a within-controls correlation of  $r = .05$ ,  
 465  $p_{uncorrected} = .806$ , and a within-exercisers correlation of  $r = -.48$ ,  $p_{uncorrected} = .015$ , between percent

466 change in FD in the dlPFC cluster and percent change in DSST score, and a within-controls

467 correlation of  $r = .08$ ,  $p_{uncorrected} = .712$ , and a within-exercisers correlation of  $r = -.50$ ,  $p_{uncorrected} = .011$ ,  
 468 between percent change in FDC in the dlPFC cluster and percent change in DSST score.

469

470 Correlations with percent change in DSST score were also detected. A positive  
 471 correlation was found with percent change in WMV in the anterior corona radiata/genu of the

472 corpus callosum,  $r(51) = .35$ ,  $p_{FDR} = .026$ . Weak negative correlations with percent change in

473 FD and FDC in the dlPFC were also found,  $r(49) = -.31$ ,  $p_{uncorrected} = .014$  and  $r(49) = -.31$ ,

474  $p_{uncorrected} = .023$ , though these did not survive FDR correction,  $p_{FDR} = .051$  and  $p_{FDR} = .053$ ,

475 respectively. No group difference was found in the change-change correlation between DSST

476 and WMV, Fisher's  $z = 1.481$ ,  $p = .139$ . The correlation between percent change in DSST

477 score and percent change in FD was significantly different between groups, Fisher's  $z =$

478 2.399,  $p = .016$ ; the EG showed a significant negative correlation,  $r(25) = -.57$ ,  $p = .002$ ,  
479 while no significant association was found within the ACG,  $r(24) = .05$ ,  $p = .806$ . The  
480 correlation between percent change in DSST score and percent change in FDC was also  
481 significantly different between groups, Fisher's  $z = 2.537$ ,  $p = .011$ ; the EG showed a  
482 significant negative correlation,  $r(25) = -.58$ ,  $p = .002$ , while no significant association was  
483 found within the ACG,  $r(24) = .08$ ,  $p = .712$ . See Figure 2 for visualization of significant  
484 correlations between percent changes in variables of interest.

485 No significant correlation was found between change in  $VO_{2peak}$  and change in  
486 DSST score,  $r(51) = .02$ ,  $p_{FDR} = .882$ .

#### 487 **4. Discussion**

488 This study investigated the effects of aerobic exercise on several white matter  
489 integrity metrics including WMV, derived from voxel-based morphometry, FA and MD,  
490 derived using diffusion-tensor models, and FD,  $\log(FC)$ , and FDC, derived using fixel-based  
491 analyses. In particular, given the known weaknesses of diffusion-tensor modeling, we were  
492 interested in whether fixel-based analysis would be better suited to capturing exercise-  
493 induced changes in white matter integrity in a sample of healthy older adults. We also looked  
494 at correlations with change in cardiovascular fitness and change in a cognitive task indexing  
495 perceptual speed.

496 Participants in the aerobic exercise group engaged in at-home interval training on a  
497 stationary bike for three to four days a week for six months, leading to an increase in  
498 cardiovascular fitness ( $VO_{2peak}$ ) compared to active control participants. This indicates that  
499 at-home aerobic exercise that is personalized to the individual is an effective intervention for  
500 cardiovascular fitness in older adults. This finding is discussed in greater detail in Polk et al.  
501 (in press).

502 We found evidence of exercise-induced maintenance of WMV in the current sample.  
503 Namely, we found two clusters, one in the anterior corona radiata extending into the genu of  
504 the corpus callosum and one in the splenium of the corpus callosum, in which change over  
505 six months was significantly different between a group of individuals engaging in regular  
506 aerobic exercise and a group of sedentary individuals. Moreover, change in WMV in the  
507 splenium was correlated with change in cardiovascular fitness; more positive change in  
508  $VO_{2peak}$  was associated with reduced loss of WMV. This effect seemed to be general, as no  
509 difference between within-group correlation coefficients was detected, indicating that this  
510 relationship was specifically not exercise-induced. However, the EG exhibited a significant  
511 increase in  $VO_{2peak}$  over the course of the intervention, which may have then contributed to

512 the maintenance of WMV in the splenium. This is consistent with a number of cross-sectional  
513 (e.g., Erickson et al., 2007; Ho et al., 2011) and longitudinal (e.g., Colcombe et al., 2006)  
514 studies finding associations of physical activity, aerobic exercise, and cardiovascular fitness  
515 with WMV in both frontal and parietal areas. Furthermore, change in WMV in the right  
516 anterior corona radiata/genu of the corpus callosum was correlated with change in DSST  
517 score, with reduced loss of volume being associated with a greater increase in score from T1  
518 to T3. This corroborates cross-sectional findings that anterior corpus callosum size is  
519 associated with performance on the DSST (Fling et al., 2011). Again, this effect seemed to be  
520 general, with no group difference in correlation, but given the preservation of WMV in the  
521 EG compared to the ACG, as well as the significant increase in DSST score among  
522 exercisers, a causal relationship between aerobic exercise, anterior corpus callosum WMV,  
523 and performance on the DSST seems plausible.

524 The findings of the diffusion-tensor model-derived metrics, FA and MD, were  
525 surprising given our hypothesis that exercise should mitigate decreasing FA and increasing  
526 MD. In the current sample, we found evidence of the opposite: in the left part of the genu of  
527 the corpus callosum, FA decreased within exercisers and increased within controls. In the  
528 right posterior corona radiata and splenium as well as in the right superior longitudinal  
529 fasciculus, MD values in the EG increased while they decreased in the ACG. In part, this  
530 finding corroborates earlier work by Clark and colleagues (2019), who found widespread  
531 decreases in FA and increases in MD in a group of 57- to 86-year-old individuals who  
532 participated in aerobic exercise for six months, although this report did not include  
533 comparisons with a control group. Additionally, Voss and colleagues (2013) found no group-  
534 level effects of one year of aerobic exercise on white matter integrity as measured by whole-  
535 brain FA in older adults aged 55 to 80 years.

536 Regarding this finding within the scope of aging, in a study comparing whole-brain  
537 FA values between younger and older adults, a number of regions, including in the cingulum  
538 bundle, were found to have greater FA values in older adults as compared to younger adults  
539 (Kelley et al., 2021). The cluster found in the anterior cingulate area in the current study, in  
540 which exercisers showed decreased FA values as compared to controls, seems to fall in a  
541 similar area to those found in the analyses by Kelley and colleagues (2021). Perhaps the  
542 increases in FA found in this area within controls are therefore consistent with age-related  
543 decline, and the decrease in FA induced by aerobic exercise in older adults is indeed  
544 indicative of a protective effect of exercise. Similarly, although Kelley and colleagues (2021)  
545 did not directly assess age differences in MD, they did find greater FD values in older adults

546 than younger adults in the posterior part of the superior longitudinal fasciculus in a similar  
547 area to the cluster of group differences in MD change found in the current study. To further  
548 investigate this cluster and whether changes in FD could indeed help explain this result, we  
549 extracted the within-subject mean values of FD from this cluster using REX. However, a  
550 post-hoc repeated measures ANOVA did not indicate any significant time-by-group  
551 interaction effect on FD, nor did we see significant changes within-group or overall. Percent  
552 changes in MD and FD values in this region were also not significantly correlated ( $r = .00$ ).  
553 Additional research should be conducted to more deeply understand the longitudinal  
554 relationship between MD and FD. Of note, higher FA in the superior longitudinal fasciculus  
555 has been associated with worse scores on a measure of visual spatial abilities in a sample of  
556 young adults with Williams Syndrome (Hoeft et al., 2007).

557         The cluster of MD in the right posterior corona radiata/splenium of the corpus  
558 callosum, in which the EG increased while the ACG increased, is somewhat more  
559 complicated to interpret. It may be important to consider both the biological underpinnings of  
560 MD here, as well as the specific location of this cluster. Within regions where white matter  
561 tracts are highly unidirectional, such as within the body of the corpus callosum, FA and MD  
562 values may accurately map onto a number of factors indicating white matter integrity,  
563 including fiber coherence, fiber diameter and density, and myelination (Basser & Pierpaoli,  
564 1996; Beaulieu, 2002; Pipersalin & Basser, 1996). However, the diffusion-tensor model has  
565 known shortcomings in the face of crossing fibers (Raffelt et al., 2012, 2015, 2017). For  
566 example, if one voxel contains fibers running perpendicularly to one another, and both these  
567 fibers have a high degree of myelination, i.e., have a high level of integrity, and another voxel  
568 contains fibers running in parallel but with a slightly lower degree of myelination, the  
569 diffusion-tensor model would still assign a higher FA value or lower MD value to the voxel  
570 in which the fibers run parallel to one another. Indeed, Kelley and colleagues (2021) found  
571 strong negative voxel-wise correlations between FA and a measure of multi-fiber complexity,  
572 an index of the number of crossing fibers, throughout the brain.

573         Thus, it is necessary to take the location of MD changes into account in order to  
574 interpret findings in this context; one of the clusters in which MD showed a time-by-group  
575 effect was located at the intersection of two major white matter tracts: the corpus callosum  
576 and the corona radiata. The increase of MD in this area may thus be representative of an  
577 increase in the prominence of crossing fibers in this area, which could even perhaps be  
578 beneficial at the convergence of major tracts. This could also explain an increase in WMV  
579 (notably not in an overlapping cluster, but located in a similar area in the opposite

580 hemisphere), as an increase in myelination in multiple directions would increase the amount  
581 of white matter found within a voxel. In support of this interpretation, we found a positive  
582 correlation between percent change in WMV in the splenium and percent change in MD in  
583 the right posterior corona radiata/splenium of the corpus callosum (see Table 3).

584 In order to further investigate potential underlying changes to fibers in this region, we  
585 ran additional post-hoc repeated measures ANOVAs. Within-subject mean values of FD,  
586 log(FC), and FDC from the cluster in the right posterior corona radiata/splenium of the  
587 corpus callosum were extracted using REX, with the hypothesis that increasing average fiber  
588 density and/or cross-section would indicate an increase in density and/or diameter of  
589 perpendicularly running fibers in this area within the EG but not the ACG. However, we  
590 found no significant differences in change in these metrics between the groups, nor did we  
591 see significant change within groups in either direction with paired t-tests comparing T1 to  
592 T3 in this cluster. Overall, FD, log(FC), and FDC showed decreases in this cluster  
593 descriptively, but changes from T1 to T3 were not significant when pooling both groups.  
594 Correlations between percent change in MD and percent change in the fixel-based metrics in  
595 these clusters were also non-significant ( $r_s \leq .15$ ). Again, further research is necessary to  
596 fully understand the relationship between changes in MD and these fixel-based metrics within  
597 the same regions, as well as the seemingly complex relationship between the directionality of  
598 FA and MD changes and aerobic exercise in aging

599 Finally, FD and FDC findings were quite similar, which is unsurprising as FDC is  
600 simply a linear combination (i.e., the product) of FD and FC. Two clusters in the PFC were  
601 found, one in the dmPFC and one in the dlPFC, however, the direction of the effects again  
602 went in the opposite direction of that which we hypothesized. Namely, significant decreases  
603 in both FD and FDC were observed in the dmPFC within exercisers, while controls showed  
604 increases. In the dlPFC, the EG showed no change, however the ACG still showed  
605 significantly more positive change than the EG. This seems to indicate, perhaps counter-  
606 intuitively, that the density as well as combined density and cross-section of fiber bundles  
607 decrease as an effect of aerobic exercise.

608 Interestingly, in the dmPFC, change in both FD and FDC values were negatively  
609 correlated with change in  $VO_2$ peak, with decreases in fitness being associated with greater  
610 increases in FD and FDC. This effect seemed to be unrelated to the exercise intervention, as  
611 the group-wise correlations were not significantly different. Altogether, this seems to suggest  
612 that increases in FD and FDC in this specific region are associated with age-related decline,  
613 and that aerobic exercise and improved cardiovascular fitness may ameliorate this decline in



614 the form of decreased FD and FDC. Notably, the correlation with change in VO<sub>2</sub>peak was  
615 also not significant within either group, which could indicate a lack of power, and the overall  
616 correlation should be interpreted with caution, given the group differences in both FD and  
617 FDC changes, as well as in VO<sub>2</sub>peak change. Furthermore, in the dlPFC, changes in FD and  
618 FDC were found to be weakly negatively correlated with change in DSST score, and these  
619 correlations were significantly different between groups, with the EG showing a significant  
620 negative correlation and the ACG showing no association. This suggests that, on a functional  
621 level, lower levels of FD and FDC in this area in the dlPFC could be beneficial in aging, with  
622 greater exercise-induced declines being associated with greater improvement in performance  
623 on a task requiring a range of cognitive processes, including perceptual speed and executive  
624 function.

625         The purported advantage of FD and FDC over diffusion tensor-derived metrics is that  
626 they are thought to estimate the intra-axonal volume even within voxels with crossing fibers,  
627 with FD representing the density of fibers within axons and FDC representing a combined  
628 measure of both the density and diameter (Raffelt et al., 2017). Because density and cross-  
629 section are modeled in multiple directions within each fixel, they are supposedly not as  
630 susceptible to the issue of lower values being falsely assigned to a voxel due to increasing  
631 myelination in perpendicular directions, for example. This is supported by findings from a  
632 combined histological and *ex vivo* MRI study in rodents, which found that axonal density in  
633 both the optic nerve and optic chiasm measured histologically correlated with FD estimated  
634 with CSD (Rojas-Vite et al., 2019). Thus, these metrics are currently thought to be a more  
635 reliable measure of the physiological properties of white matter fibers than FA and MD.

636         Although some cross-sectional studies have established an inverse relationship  
637 between age and fixel-based metrics (e.g., Choy et al., 2020; Kelley et al., 2021), the  
638 relationship with functional outcomes such as cognitive abilities are not as clear. Research in  
639 patients has found lower FD and FDC throughout the brain in individuals with Alzheimer's  
640 disease (e.g., Mito et al., 2018) and Parkinson's disease (Zarkali et al., 2020), but the  
641 association between these metrics and cognition in healthy older adults has not yet been  
642 explored. The current study suggests that it may not be a universal truth that higher FD and  
643 FDC are beneficial for functional outcomes in aging, and furthermore, that the directionality  
644 of age-related changes as well as exercise-induced changes may be region-specific. Of note,  
645 both clusters found in the current analyses are localized at the border between white matter  
646 and gray matter, in what is known as superficial white matter. Studies using FA and MD have  
647 found an inverse relationship between superficial white matter integrity and age (Nazeri et

648 al., 2015), as well as cross-sectional associations with cognitive function (Reginold et al.,  
649 2016), but again, there has been little to no investigation of region-specific associations  
650 between age, cognition, and fixel-based metrics in these superficial white matter areas.

651 Altogether, the findings of the current study provide evidence of an effect of aerobic  
652 exercise on WMV, tensor-based FA and MD, and fixel-based FD and FDC in older adults,  
653 and that increases in WMV in the corpus callosum and reductions FD and FDC in superficial  
654 WM in the PFC are related to increased fitness as well as improvement on a cognitive task  
655 indexing perceptual speed and executive function.

#### 656 **4.1 Limitations**

657 The current study has a number of limitations that warrant mentioning. First, given the  
658 interventional nature of the study, the sample was only moderately sized. This may have  
659 reduced our power to find robust effects that would survive correction for multiple  
660 comparisons. Future studies should aim to replicate the current findings with larger sample  
661 sizes. Second, the sample in the current study is relatively homogeneous: participants were  
662 recruited from an area with relatively high socioeconomic status, they were well-educated on  
663 average, and indicated no major health problems, despite not engaging in physical activity on  
664 a regular basis. Thus, there could be other protective factors at play apart from aerobic  
665 exercise. Future studies should aim to generalize the findings of the current study in more  
666 heterogeneous samples. Finally, regarding the white matter findings using diffusion-tensor  
667 model- and fixel-based metrics, the clusters in which groups differed in change that were  
668 found at threshold of  $p_{\text{uncorrected}} < .001$  did not survive correction for false discovery rate, so  
669 these results should be interpreted with caution. Furthermore, both clusters of FD and FDC  
670 change in the dmPFC and dlPFC were localized at the edge of the cortex in superficial white  
671 matter. These fiber bundles are more complicated to study than deep white matter, given their  
672 proximity to gray matter, and may also be more susceptible to noise during MR acquisition,  
673 leading to difficulties in the estimation of certain WM tracts (Guevara et al., 2020; Kirilina et  
674 al., 2020; Reveley et al., 2015). Additionally, the current study used  $b$ -values of  $710 \text{ s/mm}^2$   
675 and  $2850 \text{ s/mm}^2$  for multi-shell analyses. However, a recent study investigating the impact of  
676 different  $b$ -values on the estimation of FD in a sample of children and adolescents (8–18  
677 years old) found that multi-shell schemes including higher  $b$ -values of  $4000 \text{ s/mm}^2$  or  $6000$   
678  $\text{s/mm}^2$  improved the sensitivity of tract-specific FD to age associations (Genc et al., 2020).  
679 Future research interested in associations between WM integrity measured with fixel-based  
680 metrics may therefore consider acquiring MR data with sequences that specifically target  
681 superficial WM, or with diffusion-weighted imaging using higher  $b$ -values.



## 682 **4.2 Conclusion**

683         The current study investigated the effects of aerobic exercise on white matter structure  
684 in older adults using voxel-based morphometry, diffusion-tensor model-derived metrics, and  
685 fixel-based metrics. We found strong evidence that aerobic exercise is protective of WMV in  
686 the corpus callosum, replicating previous findings. We also found that these changes  
687 positively correlated with both changes in cardiovascular fitness and changes in performance  
688 on a cognitive task indexing executive function and perceptual speed. We found a decrease in  
689 FA and an increase in MD within exercisers. While this was contrary to our expectations,  
690 there have been previous reports of similar findings. Moreover, we found that controls  
691 showed greater increases in FA and decreases in MD than exercisers, adding to the existing  
692 skepticism of the interpretation that increased FA and decreased MD are always beneficial in  
693 aging. Finally, we found decreased FD and FDC in exercisers as compared to controls in  
694 frontal regions near the cortex. These decreases were negatively correlated with increases in  
695  $VO_2$ peak overall, as well as increases in DSST scores within exercisers. This suggests that  
696 the generalized interpretation that higher density and cross-section of white matter fibers are  
697 always associated with better outcomes in aging may not be precise enough. Specifically,  
698 changes in FD and FDC in deep versus superficial WM may have different functional  
699 implications, and these metrics should be investigated in a region-specific manner to  
700 understand potentially differentiated biological underpinnings of FD and FDC changes in  
701 different brain areas.

702

703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714

### **Acknowledgements**

This work was supported by the Max Planck Society and the Max Planck Institute for Human Development and is part of the BMBF-funded Energi Consortium (01GQ1421B). We are very grateful to the Neotiv team for providing the app for ergometer training as well as their technical support, and to everyone at Shared Reading for organizing the book club for our active control group. We thank Michael Krause for his continuous assistance in the implementation of data preprocessing on the computing cluster, Sebastian Schröder and his student assistants for providing the technical infrastructure of the study, Kirsten Becker and Anke Schepers-Klingebiel for their organizational assistance, and the MRI team at the Max Planck Institute for Human Development (Sonali Beckmann, Nadine Taube, Thomas Feg, and Davide Santoro) and all the participants for their time and support.

715

## References

- 716 Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A  
717 reproducible evaluation of ANTs similarity metric performance in brain image  
718 registration. *NeuroImage*, *54*(3), 2033–2044.  
719 <https://doi.org/10.1016/j.neuroimage.2010.09.025>
- 720 Avants, B. B., Yushkevich, P., Pluta, J., Minkoff, D., Korczykowski, M., Detre, J., & Gee, J. C.  
721 (2010). The optimal template effect in hippocampus studies of diseased  
722 populations. *NeuroImage*, *49*(3), 2457–2466.  
723 <https://doi.org/10.1016/j.neuroimage.2009.09.062>
- 724 Barha, C. K., Davis, J. C., Falck, R. S., Nagamatsu, L. S., & Liu-Ambrose, T. (2017). Sex  
725 differences in exercise efficacy to improve cognition: A systematic review and meta-  
726 analysis of randomized controlled trials in older humans. *Frontiers in*  
727 *Neuroendocrinology*, *46*, 71–85. <https://doi.org/10.1016/j.yfrne.2017.04.002>
- 728 Bassler, P. J., & Pierpaoli, C. (1996). Microstructural and Physiological Features of Tissues  
729 Elucidated by Quantitative-Diffusion-Tensor MRI. *Journal of Magnetic Resonance,*  
730 *Series B*, *111*(3), 209–219. <https://doi.org/10.1006/jmrb.1996.0086>
- 731 Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system—A  
732 technical review. *NMR in Biomedicine*, *15*(7–8), 435–455.  
733 <https://doi.org/10.1002/nbm.782>
- 734 Beck, D., de Lange, A.-M. G., Maximov, I. I., Richard, G., Andreassen, O. A., Nordvik, J. E.,  
735 & Westlye, L. T. (2021). White matter microstructure across the adult lifespan: A  
736 mixed longitudinal and cross-sectional study using advanced diffusion models and  
737 brain-age prediction. *NeuroImage*, *224*, 117441.  
738 <https://doi.org/10.1016/j.neuroimage.2020.117441>
- 739 Bendlin, B. B., Fitzgerald, M. E., Ries, M. L., Xu, G., Kastman, E. K., Thiel, B. W., Rowley,  
740 H. A., Lazar, M., Alexander, A. L., & Johnson, S. C. (2010). White Matter in Aging  
741 and Cognition: A Cross-Sectional Study of Microstructure in Adults Aged Eighteen to  
742 Eighty-Three. *Developmental Neuropsychology*, *35*(3), 257–277.  
743 <https://doi.org/10.1080/87565641003696775>
- 744 Benedict, C., Brooks, S. J., Kullberg, J., Nordenskjöld, R., Burgos, J., Le Grevès, M.,  
745 Kilander, L., Larsson, E.-M., Johansson, L., Ahlström, H., Lind, L., & Schiöth, H. B.  
746 (2013). Association between physical activity and brain health in older adults.  
747 *Neurobiology of Aging*, *34*(1), 83–90.  
748 <https://doi.org/10.1016/j.neurobiolaging.2012.04.013>
- 749 Bennett, I. J., & Madden, D. J. (2014). Disconnected aging: Cerebral white matter integrity  
750 and age-related differences in cognition. *Neuroscience*, *276*, 187–205.  
751 <https://doi.org/10.1016/j.neuroscience.2013.11.026>
- 752 Burzynska, A. Z., Wong, C. N., Voss, M. W., Cooke, G. E., Gothe, N. P., Fanning, J.,  
753 McAuley, E., & Kramer, A. F. (2015). Physical activity is linked to greater moment-  
754 to-moment variability in spontaneous brain activity in older adults. *PLoS ONE*, *10*(8),  
755 1–18. <https://doi.org/10.1371/journal.pone.0134819>
- 756 Choy, S. W., Bagarinao, E., Watanabe, H., Ho, E. T. W., Maesawa, S., Mori, D., Hara, K.,  
757 Kawabata, K., Yoneyama, N., Ohdake, R., Imai, K., Masuda, M., Yokoi, T., Ogura,  
758 A., Taoka, T., Koyama, S., Tanabe, H. C., Katsuno, M., Wakabayashi, T., ... Sobue,  
759 G. (2020). Changes in white matter fiber density and morphology across the adult  
760 lifespan: A cross-sectional fixel-based analysis. *Human Brain Mapping*, *41*(12),  
761 3198–3211. <https://doi.org/10.1002/hbm.25008>
- 762 Clark, C. M., Guadagni, V., Mazerolle, E. L., Hill, M., Hogan, D. B., Pike, G. B., & Poulin,  
763 M. J. (2019). Effect of aerobic exercise on white matter microstructure in the aging  
764 brain. *Behavioural Brain Research*, *373*, 112042.

- 765 <https://doi.org/10.1016/j.bbr.2019.112042>
- 766 Colcombe, S. J., Erickson, K. I., Scalf, P. E., Kim, J. S., Prakash, R., McAuley, E., Elavsky,  
767 S., Marquez, D. X., Hu, L., & Kramer, A. F. (2006). Aerobic Exercise Training  
768 Increases Brain Volume in Aging Humans. *The Journals of Gerontology Series A:  
769 Biological Sciences and Medical Sciences*, 61(11), 1166–1170.  
770 <https://doi.org/10.1093/gerona/61.11.1166>
- 771 Damoiseaux, J. S., Smith, S. M., Witter, M. P., Sanz-Arigita, E. J., Barkhof, F., Scheltens, P.,  
772 Stam, C. J., Zarei, M., & Rombouts, S. A. R. B. (2009). White matter tract integrity in  
773 aging and Alzheimer's disease. *Human Brain Mapping*, 30(4), 1051–1059.  
774 <https://doi.org/10.1002/hbm.20563>
- 775 Dhollander, T., Raffelt, D., & Connelly, A. (2016). A novel iterative approach to reap the  
776 benefits of multi-tissue CSD from just single-shell (+b=0) diffusion MRI data.  
777 *Proceedings of the 24th Annual Meeting of the International Society of Magnetic  
778 Resonance in Medicine*, 3010.
- 779 Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical  
780 Comparison of Correlations. *PLOS ONE*, 10(4), e0121945.  
781 <https://doi.org/10.1371/journal.pone.0121945>
- 782 Duff, E. P., Cunnington, R., & Egan, G. F. (2007). REX: Response Exploration for  
783 Neuroimaging Datasets. *Neuroinformatics*, 5(4), 223–234.  
784 <https://doi.org/10.1007/s12021-007-9001-y>
- 785 Erickson, K. I., Colcombe, S. J., Wadhwa, R., Bherer, L., Peterson, M. S., Scalf, P. E., Kim,  
786 J. S., Alvarado, M., & Kramer, A. F. (2007). Training-induced plasticity in older  
787 adults: Effects of training on hemispheric asymmetry. *Neurobiology of Aging*, 28(2),  
788 272–283. <https://doi.org/10.1016/j.neurobiolaging.2005.12.012>
- 789 Erickson, K. I., Hillman, C., Stillman, C. M., Ballard, R. M., Bloodgood, B., Conroy, D. E.,  
790 Macko, R., Marquez, D. X., Petruzzello, S. J., & Powell, K. E. (2019). Physical  
791 Activity, Cognition, and Brain Outcomes: A Review of the 2018 Physical Activity  
792 Guidelines. *Medicine & Science in Sports & Exercise*, 51(6), 1242–1251.  
793 <https://doi.org/10.1249/MSS.0000000000001936>
- 794 Fling, B. W., Chapekis, M., Reuter-Lorenz, P. A., Anguera, J., Bo, J., Langan, J., Welsh, R.  
795 C., & Seidler, R. D. (2011). Age differences in callosal contributions to cognitive  
796 processes. *Neuropsychologia*, 49(9), 2564–2569.  
797 <https://doi.org/10.1016/j.neuropsychologia.2011.05.004>
- 798 Genc, S., Tax, C., Raven, E. P., Chamberland, M., Parker, G. D., & Jones, D. K. (2020).  
799 Impact of b-value on estimates of apparent fibre density. *Human brain  
800 mapping*, 41(10), 2583–2595. <https://doi.org/10.1002/hbm.24964>
- 801 Gow, A. J., Bastin, M. E., Munoz Maniega, S., Valdes Hernandez, M. C., Morris, Z., Murray,  
802 C., Royle, N. A., Starr, J. M., Deary, I. J., & Wardlaw, J. M. (2012). Neuroprotective  
803 lifestyles and the aging brain: Activity, atrophy, and white matter integrity.  
804 *Neurology*, 79(17), 1802–1808. <https://doi.org/10.1212/WNL.0b013e3182703fd2>
- 805 Guevara, M., Guevara, P., Román, C., & Mangin, J.-F. (2020). Superficial white matter: A  
806 review on the dMRI analysis methods and applications. *NeuroImage*, 212, 116673.  
807 <https://doi.org/10.1016/j.neuroimage.2020.116673>
- 808 Gunning-Dixon, F. M., Brickman, A. M., Cheng, J. C., & Alexopoulos, G. S. (2009). Aging  
809 of cerebral white matter: A review of MRI findings. *International Journal of  
810 Geriatric Psychiatry*, 24(2), 109–117. <https://doi.org/10.1002/gps.2087>
- 811 Harrell, F. E. J. (2021). *Hmisc: Harrell Miscellaneous* (4.6-0) [R]. [https://CRAN.R-  
812 project.org/package=Hmisc](https://CRAN.R-project.org/package=Hmisc)
- 813 Ho, A. J., Raji, C. A., Becker, J. T., Lopez, O. L., Kuller, L. H., Hua, X., Dinov, I. D., Stein,  
814 J. L., Rosano, C., Toga, A. W., & Thompson, P. M. (2011). The effects of physical

- 815 activity, education, and body mass index on the aging brain. *Human Brain Mapping*,  
816 32(9), 1371–1382. <https://doi.org/10.1002/hbm.21113>
- 817 Hoefl, F., Barnea-Goraly, N., Haas, B. W., Golarai, G., Ng, D., Mills, D., Korenberg, J.,  
818 Bellugi, U., Galaburda, A., & Reiss, A. L. (2007). More is not always better:  
819 increased fractional anisotropy of superior longitudinal fasciculus associated with  
820 poor visuospatial abilities in Williams syndrome. *The Journal of neuroscience : the*  
821 *official journal of the Society for Neuroscience*, 27(44), 11960–11965.  
822 <https://doi.org/10.1523/JNEUROSCI.3591-07.2007>
- 823 Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., & Smith, S. M. (2012).  
824 FSL. *NeuroImage*, 62(2), 782–790. <https://doi.org/10.1016/j.neuroimage.2011.09.015>
- 825 Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., & Sijbers, J. (2013). Investigating  
826 the prevalence of complex fiber configurations in white matter tissue with diffusion  
827 magnetic resonance imaging: Prevalence of Multifiber Voxels in WM. *Human Brain*  
828 *Mapping*, 34(11), 2747–2766. <https://doi.org/10.1002/hbm.22099>
- 829 Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., & Sijbers, J. (2014). Multi-tissue  
830 constrained spherical deconvolution for improved analysis of multi-shell diffusion  
831 MRI data. *NeuroImage*, 103, 411–426.  
832 <https://doi.org/10.1016/j.neuroimage.2014.07.061>
- 833 Johnson, N. F., Kim, C., Clasey, J. L., Bailey, A., & Gold, B. T. (2012). Cardiorespiratory  
834 fitness is positively correlated with cerebral white matter integrity in healthy seniors.  
835 *NeuroImage*, 59(2), 1514–1523. <https://doi.org/10.1016/j.neuroimage.2011.08.032>
- 836 Kassambara, A. (2021). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. (0.7.0)  
837 [R]. <https://CRAN.R-project.org/package=rstatix>
- 838 Kelley, S., Plass, J., Bender, A. R., & Polk, T. A. (2021). Age-Related Differences in White  
839 Matter: Understanding Tensor-Based Results Using Fixel-Based Analysis. *Cerebral*  
840 *Cortex (New York, N.Y.: 1991)*, 31(8), 3881–3898.  
841 <https://doi.org/10.1093/cercor/bhab056>
- 842 Kennedy, K. M., & Raz, N. (2009). Aging white matter and cognition: Differential effects of  
843 regional variations in diffusion properties on memory, executive functions, and speed.  
844 *Neuropsychologia*, 47(3), 916–927.  
845 <https://doi.org/10.1016/j.neuropsychologia.2009.01.001>
- 846 Kirilina, E., Helbling, S., Morawski, M., Pine, K., Reimann, K., Jankuhn, S., Dinse, J.,  
847 Deistung, A., Reichenbach, J. R., Trampel, R., Geyer, S., Müller, L., Jakubowski, N.,  
848 Arendt, T., Bazin, P. L., & Weiskopf, N. (2020). Superficial white matter imaging:  
849 Contrast mechanisms and whole-brain in vivo mapping. *Science advances*, 6(41),  
850 eaaz9281. <https://doi.org/10.1126/sciadv.aaz9281>
- 851 Liu, H., Yang, Y., Xia, Y., Zhu, W., Leak, R. K., Wei, Z., Wang, J., & Hu, X. (2017). Aging  
852 of cerebral white matter. *Ageing Research Reviews*, 34, 64–76.  
853 <https://doi.org/10.1016/j.arr.2016.11.006>
- 854 Liu, Z., Farzinfar, M., Katz, L. M., Zhu, H., Goodlett, C., Gerig, G., Styner, M., & Marks, B.  
855 L. (2012). Automated Voxel-Wise Brain DTI Analysis of Fitness and Aging. *The*  
856 *Open Medical Imaging Journal*, 6(1), 80–88.  
857 <https://doi.org/10.2174/1874347101206010080>
- 858 Madden, D. J., Spaniol, J., Costello, M. C., Bucur, B., White, L. E., Cabeza, R., Davis, S. W.,  
859 Dennis, N. A., Provenzale, J. M., & Huettel, S. A. (2008). Cerebral White Matter  
860 Integrity Mediates Adult Age Differences in Cognitive Performance. *Journal of*  
861 *Cognitive Neuroscience*, 21(2), 289–302. <https://doi.org/10.1162/jocn.2009.21047>
- 862 Magistro, D., Takeuchi, H., Nejad, K. K., Taki, Y., Sekiguchi, A., Nouchi, R., ...  
863 Kawashima, R. (2015) The Relationship between Processing Speed and Regional  
864 White Matter Volume in Healthy Young People. *PLoS ONE* 10(9): e0136386. doi:



- 865 10.1371/journal.pone.0136386
- 866 Marks, B., Katz, L., Styner, M., & Smith, J. (2011). Aerobic fitness and obesity: Relationship  
867 to cerebral white matter integrity in the brain of active and sedentary older adults.  
868 *British Journal of Sports Medicine*, 45(15), 1208–1215.  
869 <https://doi.org/10.1136/bjism.2009.068114>
- 870 Mito, R., Raffelt, D., Dhollander, T., Vaughan, D. N., Tournier, J.-D., Salvado, O.,  
871 Brodtmann, A., Rowe, C. C., Villemagne, V. L., & Connelly, A. (2018). Fibre-  
872 specific white matter reductions in Alzheimer’s disease and mild cognitive  
873 impairment. *Brain*, 141(3), 888–902. <https://doi.org/10.1093/brain/awx355>
- 874 Nazeri, A., Chakravarty, M. M., Rajji, T. K., Felsky, D., Rotenberg, D. J., Mason, M., Xu, L.  
875 N., Lobaugh, N. J., Mulsant, B. H., & Voineskos, A. N. (2015). Superficial white  
876 matter as a novel substrate of age-related cognitive decline. *Neurobiology of Aging*,  
877 36(6), 2094–2106. <https://doi.org/10.1016/j.neurobiolaging.2015.02.022>
- 878 Papp, K. V., Kaplan, R. F., Springate, B., Moscufo, N., Wakefield, D. B., Guttmann, C. R., &  
879 Wolfson, L. (2014). Processing speed in normal aging: effects of white matter  
880 hyperintensities and hippocampal volume loss. *Neuropsychology, development, and*  
881 *cognition. Section B, Aging, neuropsychology and cognition*, 21(2), 197–213. doi:  
882 10.1080/13825585.2013.795513
- 883 Pierpaoli, C., & Basser, P. J. (1996). Toward a quantitative assessment of diffusion  
884 anisotropy. *Magnetic Resonance in Medicine*, 36(6), 893–906.  
885 <https://doi.org/10.1002/mrm.1910360612>
- 886 Polk, S. E., Kleemeyer, M. M., Köhncke, Y., Brandmaier, A. M., Bodammer, N. C., Misgeld,  
887 C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Wenger, E., & Düzel, S. (in  
888 press). Change in latent gray-matter structural integrity is associated with change in  
889 cardiovascular fitness in older adults who engage in at-home aerobic exercise.  
890 *Frontiers in Human Neuroscience*.
- 891 Prakash, R. S., Snook, E. M., Motl, R. W., & Kramer, A. F. (2010). Aerobic fitness is  
892 associated with gray matter volume and white matter integrity in multiple  
893 sclerosis. *Brain research*, 1341, 41–51. doi: 10.1016/j.brainres.2009.06.063
- 894 R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R  
895 Foundation for Statistical Computing. <https://www.R-project.org/>
- 896 Raffelt, D. A., Smith, R. E., Ridgway, G. R., Tournier, J.-D., Vaughan, D. N., Rose, S.,  
897 Henderson, R., & Connelly, A. (2015). Connectivity-based fixel enhancement:  
898 Whole-brain statistical analysis of diffusion MRI measures in the presence of crossing  
899 fibres. *NeuroImage*, 117, 40–55. <https://doi.org/10.1016/j.neuroimage.2015.05.039>
- 900 Raffelt, D. A., Tournier, J.-D., Rose, S., Ridgway, G. R., Henderson, R., Crozier, S., Salvado,  
901 O., & Connelly, A. (2012). Apparent Fibre Density: A novel measure for the analysis  
902 of diffusion-weighted magnetic resonance images. *NeuroImage*, 59(4), 3976–3994.  
903 <https://doi.org/10.1016/j.neuroimage.2011.10.045>
- 904 Raffelt, D. A., Tournier, J.-D., Smith, R. E., Vaughan, D. N., Jackson, G., Ridgway, G. R., &  
905 Connelly, A. (2017). Investigating white matter fibre density and morphology using  
906 fixel-based analysis. *NeuroImage*, 144, 58–73.  
907 <https://doi.org/10.1016/j.neuroimage.2016.09.029>
- 908 Reginold, W., Luedke, A. C., Itorralba, J., Fernandez-Ruiz, J., Islam, O., & Garcia, A.  
909 (2016). Altered Superficial White Matter on Tractography MRI in Alzheimer’s  
910 Disease. *Dementia and Geriatric Cognitive Disorders Extra*, 6(2), 233–241.  
911 <https://doi.org/10.1159/000446770>
- 912 Reveley, C., Seth, A. K., Pierpaoli, C., Silva, A. C., Yu, D., Saunders, R. C., Leopold, D. A.,  
913 & Ye, F. Q. (2015). Superficial white matter fiber systems impede detection of long-  
914 range cortical connections in diffusion MR tractography. *Proceedings of the National*

- 915 *Academy of Sciences of the United States of America*, 112(21), E2820-8.  
 916 <https://doi.org/10.1073/pnas.1418198112>
- 917 Rojas-Vite, G., Coronado-Leija, R., Narvaez-Delgado, O., Ramírez-Manzanares, A.,  
 918 Marroquín, J. L., Noguez-Imm, R., Aranda, M. L., Scherrer, B., Larriva-Sahd, J., &  
 919 Concha, L. (2019). Histological validation of per-bundle water diffusion metrics  
 920 within a region of fiber crossing following axonal degeneration. *NeuroImage*, 201,  
 921 116013. <https://doi.org/10.1016/j.neuroimage.2019.116013>
- 922 RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC.  
 923 <http://www.rstudio.com/>
- 924 Sexton, C. E., Betts, J. F., Demnitz, N., Dawes, H., Ebmeier, K. P., & Johansen-Berg, H.  
 925 (2016). A systematic review of MRI studies examining the relationship between  
 926 physical fitness and activity and the white matter of the ageing brain. *NeuroImage*,  
 927 131, 81–90. <https://doi.org/10.1016/j.neuroimage.2015.09.071>
- 928 Sexton, C. E., Walhovd, K. B., Storsve, A. B., Tamnes, C. K., Westlye, L. T., Johansen-Berg,  
 929 H., & Fjell, A. M. (2014). Accelerated Changes in White Matter Microstructure  
 930 during Aging: A Longitudinal Diffusion Tensor Imaging Study. *Journal of*  
 931 *Neuroscience*, 34(46), 15425–15436. [https://doi.org/10.1523/JNEUROSCI.0203-](https://doi.org/10.1523/JNEUROSCI.0203-14.2014)  
 932 14.2014
- 933 Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E.,  
 934 Watkins, K. E., Ciccarelli, O., Cader, M. Z., Matthews, P. M., & Behrens, T. E. J.  
 935 (2006). Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion  
 936 data. *NeuroImage*, 31(4), 1487–1505.  
 937 <https://doi.org/10.1016/j.neuroimage.2006.02.024>
- 938 Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-  
 939 Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K.,  
 940 Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M.  
 941 (2004). Advances in functional and structural MR image analysis and implementation  
 942 as FSL. *NeuroImage*, 23, S208–S219.  
 943 <https://doi.org/10.1016/j.neuroimage.2004.07.051>
- 944 Stillman, C. M., Esteban-Cornejo, I., Brown, B., Bender, C. M., & Erickson, K. I. (2020).  
 945 Effects of Exercise on Brain and Cognition Across Age Groups and Health States.  
 946 *Trends in Neurosciences*, 43(7), 533–543. <https://doi.org/10.1016/j.tins.2020.04.010>
- 947 Tahedl, M. (2018). *B.A.T.M.A.N.: Basic and Advanced Tractography with MRtrix for All*  
 948 *Neurophiles*. <https://doi.org/10.17605/OSF.IO/FKYHT>
- 949 Tian, Q., Erickson, K. I., Simonsick, E. M., Aizenstein, H. J., Glynn, N. W., Boudreau, R.  
 950 M., Newman, A. B., Kritchevsky, S. B., Yaffe, K., Harris, T. B., & Rosano, C.  
 951 (2014). Physical Activity Predicts Microstructural Integrity in Memory-Related  
 952 Networks in Very Old Adults. *The Journals of Gerontology Series A: Biological*  
 953 *Sciences and Medical Sciences*, 69(10), 1284–1290.  
 954 <https://doi.org/10.1093/gerona/glt287>
- 955 Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens,  
 956 D., Jeurissen, B., Yeh, C.-H., & Connelly, A. (2019). MRtrix3: A fast, flexible and  
 957 open software framework for medical image processing and visualisation.  
 958 *NeuroImage*, 202, 116137. <https://doi.org/10.1016/j.neuroimage.2019.116137>
- 959 Tseng, B. Y., Gundapuneedi, T., Khan, M. A., Diaz-Arrastia, R., Levine, B. D., Lu, H.,  
 960 Huang, H., & Zhang, R. (2013). White matter integrity in physically fit older adults.  
 961 *NeuroImage*, 82, 510–516. <https://doi.org/10.1016/j.neuroimage.2013.06.011>
- 962 Tseng, B. Y., Uh, J., Rossetti, H. C., Cullum, C. M., Diaz-Arrastia, R. F., Levine, B. D., Lu,  
 963 H., & Zhang, R. (2013). Masters athletes exhibit larger regional brain volume and  
 964 better cognitive performance than sedentary older adults: MRI Study in Old Athletes’

- 965 Brains. *Journal of Magnetic Resonance Imaging*, 38(5), 1169–1176.  
966 <https://doi.org/10.1002/jmri.24085>
- 967 Voss, M. W., Heo, S., Prakash, R. S., Erickson, K. I., Alves, H., Chaddock, L., Szabo, A. N.,  
968 Mailey, E. L., Wójcicki, T. R., White, S. M., Gothe, N., Mcauley, E., Sutton, B. P., &  
969 Kramer, A. F. (2013). The influence of aerobic fitness on cerebral white matter  
970 integrity and cognitive function in older adults: Results of a one-year exercise  
971 intervention. *Human Brain Mapping*, 34(11), 2972–2985.  
972 <https://doi.org/10.1002/hbm.22119>
- 973 Wechsler, D. (1981). The psychometric tradition: Developing the wechsler adult intelligence  
974 scale. *Contemporary Educational Psychology*, 6(2), 82–85.  
975 [https://doi.org/10.1016/0361-476X\(81\)90035-7](https://doi.org/10.1016/0361-476X(81)90035-7)
- 976 Wenger, E., Düzel, S., Kleemeyer, M.M., Polk, S.E., Köhncke, Y., Bodammer, N.C.,  
977 Mohammadi, S., Misgeld, C., Porst, J., Wolfarth, B., Brandmaier, A.M., Kühn, S., &  
978 Lindenberger, U. (2022). Vamos en bici: Study protocol of an investigation of  
979 cognitive and neural changes following language training, physical exercise training,  
980 or a combination of both. *BioRxiv*. doi: 10.1101/2022.01.30.478181
- 981 Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T.,  
982 Beckmann, C., Jenkinson, M., & Smith, S. M. (2009). Bayesian analysis of  
983 neuroimaging data in FSL. *NeuroImage*, 45(1), S173–S186.  
984 <https://doi.org/10.1016/j.neuroimage.2008.10.055>
- 985 Young J, Angevaren M, Rusted J, Tabet N. (2015). Aerobic exercise to improve cognitive  
986 function in older people without known cognitive impairment. *Cochrane Database of*  
987 *Systematic Reviews*. doi: 10.1002/14651858.CD005381.pub4
- 988 Zarkali, A., McColgan, P., Leyland, L.-A., Lees, A. J., Rees, G., & Weil, R. S. (2020). Fiber-  
989 specific white matter reductions in Parkinson hallucinations and visual dysfunction.  
990 *Neurology*, 94(14), e1525–e1538. <https://doi.org/10.1212/WNL.0000000000009014>  
991





## DECLARATION OF RESEARCHER CONTRIBUTIONS

Annex Declaration pursuant to Sec. 7 (3), fourth sentence, of the Doctoral Study Regulations regarding my own share of the submitted scientific or scholarly work that has been published or is intended for publication within the scope of my publication-based work

**I.** Last name, first name: Polk, Sarah Elisabeth

Institute: Department of Education and Psychology

Doctoral study subject: Psychology

Title: M.Sc. in Social, Cognitive, and Affective Neuroscience; B.A. in Psychology

**II. Numbered listing of works submitted (title, authors, where and when published and/or submitted):**

1. Wenger, E., Polk, S. E., Kleemeyer, M. M., Weiskopf, N., Bodammer, N. C., Lindenberger, U., & Brandmaier, A. M. (2022). Reliability of quantitative multiparameter maps is high for MT and PD but attenuated for R1 and R2\* in healthy young adults. *Human Brain Mapping*. <https://doi.org/10.1002/hbm.25870>
2. Polk, S. E., Kleemeyer, M. M., Köhncke, Y., Brandmaier, A. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Wenger, E., & Düzel, S. (2022). Change in Latent Gray-Matter Structural Integrity Is Associated With Change in Cardiovascular Fitness in Older Adults Who Engage in At-Home Aerobic Exercise. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2022.852737>
3. Polk, S. E., Kleemeyer, M. M., Bodammer, N. C., Misgeld, C., Porst, J., Wolfarth, B., Kühn, S., Lindenberger, U., Düzel, S., & Wenger, E. (2022). *Aerobic exercise is associated with region-specific changes in volumetric, tensor-based, and fixel-based measures of white matter integrity in healthy older adults* [Manuscript under revision].

**III. Explanation of own share of these works:**

The amount of the work completed by myself is evaluated on the following scale:

all – the vast majority – most – part.

Regarding II. 1.: Data collection (part), data analysis and programming (the vast majority), data visualization (the vast majority), manuscript preparation (part).

Regarding II. 2.: Data collection (part), data analysis and programming (all), data visualization (all), interpretation and discussion of results (the vast majority), and manuscript preparation (the vast majority).

Regarding II. 3.: Data collection (part), (pre)processing of MRI data (most), data analysis and programming (all), data visualization (all), interpretation and discussion of results (the vast majority), and manuscript preparation (the vast majority).

**IV. Names and e-mail addresses for the relevant co-authors:**

Regarding II. 1.:

Dr. Elisabeth Wenger  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
wenger@mpib-berlin.mpg.de

Dr. Maike M. Kleemeyer  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
kleemeyer@mpib-berlin.mpg.de

Prof. Dr. Nikolaus Weiskopf  
Max Planck Institute for Human Cognitive and Brain Sciences  
Stephanstraße 1A  
04103 Leipzig  
weiskopf@cbs.mpg.de

Dr. Nils C. Bodammer  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
bodammer@mpib-berlin.mpg.de

Prof. Dr. Ulman Lindenberger  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
seklindenberger@mpib-berlin.mpg.de

Dr. Andreas Brandmaier  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
brandmaier@mpib-berlin.mpg.de

Regarding II. 2.:

Dr. Maike M. Kleemeyer (s.a.)

Dr. Ylva Köhncke  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
koehncke@mpib-berlin.mpg.de

Dr. Andreas Brandmaier (s.a.)

Dr. Nils C. Bodammer (s.a.)

Dr. med. Carola Misgeld  
Charité – University Medicine Berlin  
Philippstraße 10  
10115 Berlin  
carola.misgeld@charite.de

Johanna Porst  
Charité – University Medicine Berlin  
Philippstraße 13, House 11  
10115 Berlin  
johanna.porst@charite.de

Univ.-Prof. Dr. med. Bernd Wolfarth  
Charité – University Medicine Berlin  
Philippstraße 13 Haus 11  
10115 Berlin  
bernd.wolfarth@charite.de

Prof. Dr. Simone Kühn  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
kuehn@mpib-berlin.mpg.de

Prof. Dr. Ulman Lindenberger (s.a.)

Dr. Elisabeth Wenger (s.a.)

Dr. Sandra Düzel  
Max Planck Institute for Human Development  
Lentzeallee 94  
14195 Berlin  
duzel@mpib-berlin.mpg.de

Regarding II. 3.:

Dr. Maike M. Kleemeyer (s.a.)

Dr. Nils C. Bodammer (s.a.)

Dr. med. Carola Misgeld (s.a.)

Johanna Porst (s.a.)

Prof. Dr. med. Bernd Wolfarth (s.a.)

Prof. Dr. Simone Kühn (s.a.)

Prof. Dr. Ulman Lindenberger (s.a.)

Dr. Sandra Düzel (s.a.)

Dr. Elisabeth Wenger (s.a.)

Berlin, 2022

Sarah E. Polk