

Changes in fitness are associated with changes in hippocampal microstructure and hippocampal volume among older adults



Maike Margarethe Kleemeyer^{a,*}, Simone Kühn^{a,b}, John Prindle^a, Nils Christian Bodammer^a, Lars Brechtel^c, Alexander Garthe^d, Gerd Kempermann^{d,e}, Sabine Schaefer^{a,f}, Ulman Lindenberger^{a,g}

^a Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

^b University Clinic Hamburg-Eppendorf, Clinic and Polyclinic for Psychiatry and Psychotherapy, Martinistrasse 52, 20246 Hamburg, Germany

^c SMS Medical Institute Berlin/Berliner Akademie für Sportmedizin e.V., Olympiapark Berlin, Hanns-Braun-Straße/Adlerplatz, 14053 Berlin, Germany

^d German Center for Neurodegenerative Diseases, Fetscherstraße 105, 01307 Dresden, Germany

^e Center for Regenerative Therapies TU Dresden, Fetscherstraße 105, 01307 Dresden, Germany

^f Leipzig University, Sportwissenschaftliche Fakultät, Jahnallee 59, 04109 Leipzig, Germany

^g European University Institute, Via dei Roccettini 9, 50014 San Domenico di Fiesole (FI), Italy

ARTICLE INFO

Article history:

Accepted 10 November 2015

Available online 14 November 2015

Keywords:

Aging

Fitness

Physical exercise

Hippocampal volume

Hippocampal microstructure

Latent difference modeling

ABSTRACT

This study investigates the effects of fitness changes on hippocampal microstructure and hippocampal volume. Fifty-two healthy participants aged 59–74 years with a sedentary lifestyle were randomly assigned to either of two levels of exercise intensity. Training lasted for six months. Physical fitness, hippocampal volumes, and hippocampal microstructure were measured before and after training. Hippocampal microstructure was assessed by mean diffusivity, which inversely reflects tissue density; hence, mean diffusivity is lower for more densely packed tissue. Mean changes in fitness did not differ reliably across intensity levels of training, so data were collapsed across groups. Multivariate modeling of pretest–posttest differences using structural equation modeling (SEM) revealed that individual differences in latent change were reliable for all three constructs. More positive changes in fitness were associated with more positive changes in tissue density (i.e., more negative changes in mean diffusivity), and more positive changes in tissue density were associated with more positive changes in volume. We conclude that fitness-related changes in hippocampal volume may be brought about by changes in tissue density. The relative contributions of angiogenesis, gliogenesis, and/or neurogenesis to changes in tissue density remain to be identified.

© 2015 Elsevier Inc. All rights reserved.

Introduction

During the last decade, evidence has been accumulated documenting beneficial effects of exercise in preserving cognitive abilities as well as brain structure and function in older adulthood (for reviews, see, e.g., Bherer et al., 2013; Erickson et al., 2015; Hötting and Röder, 2013). The hippocampus has attracted much attention because some of the most compelling evidence for positive effects of fitness on cognition has been found for this brain region, in both animal and human studies. The hippocampus is a small brain region located in the

medial temporal lobes, which plays a major role in spatial and contextual memory formation (Squire et al., 2004). It is one of the first brain regions affected by age-related atrophy and has been associated with age-related disorders such as Alzheimer's disease (Barnes et al., 2009). On the other hand, the hippocampus is one of two brain regions for which the potential for neurogenesis is preserved into late adulthood (Erickson et al., 1998; Spalding et al., 2013). Specifically, the hippocampus is responsive to exercise in animal models (Kronenberg et al., 2003; van Praag et al., 1999).

Multiple studies point to stable or even increased hippocampal volumes after completing 3–12 months of physical exercise interventions (Erickson et al., 2011; Maass et al., 2015; Niemann et al., 2014), using either manual segmentation or automated subcortical segmentation procedures for the MR images. However, the cellular mechanisms underlying these volumetric changes remain largely unknown. Some insight can be gained from animal studies showing that animals with access to running wheels as opposed to those held in standard cages show increased capillary density, and produce new neurons (for reviews, see Kempermann, 2012; Thomas et al., 2012; Voss et al., 2013).

* Corresponding author.

E-mail addresses: kleemeyer@mpib-berlin.mpg.de (M.M. Kleemeyer), kuehn@mpib-berlin.mpg.de (S. Kühn), prindle@mpib-berlin.mpg.de (J. Prindle), bodammer@mpib-berlin.mpg.de (N.C. Bodammer), lars.brechtel@medical-institute-berlin.de (L. Brechtel), alexander.garthe@dzne.de (A. Garthe), gerd.kempermann@dzne.de (G. Kempermann), sabine.schaefer@uni-leipzig.de (S. Schaefer), seklindenberger@mpib-berlin.mpg.de (U. Lindenberger).

Consistent with the animal literature, [Pereira et al. \(2007\)](#) demonstrated that after 3 months of exercise in middle-aged human adults, changes in fitness correlated with increases in dentate gyrus cerebral blood volume (CBV), and by extrapolating from animal data, claimed that these changes are related to neurogenesis. A recent study replicated this finding in older adults, reporting associations among changes in fitness, hippocampal perfusion, and volume of the hippocampal head in the context of a 3-month fitness intervention program ([Maass et al., 2015](#)). At the cellular level, these observations may reflect either dilation of existing blood vessels or angiogenesis.

Parameters derived from diffusion tensor imaging (DTI) may shed additional light on underlying changes in brain microstructure. DTI measures the diffusion of water molecules within tissue and is commonly used to determine white matter integrity ([Bammer, 2003](#); [Beaulieu, 2002](#)). DTI can also be applied to gray matter, where it likewise provides information on how freely water molecules can diffuse through tissue. When many barriers (mostly cell membranes) are present, that is, when the tissue is densely packed, mean diffusivity (MD) would be low, and vice versa. In this way, DTI can complement volumetric measures with information on tissue (or barrier) density (cf. [Lövdén et al., 2013](#)). Thus, in the case of exercise-induced volumetric increases that are accompanied by increases in the bulk of membranes, including potentially angiogenesis, gliogenesis, or neurogenesis (hyperplasia), we would expect to observe an increase in barrier density, that is, a decrease in mean diffusivity as opposed to unchanged diffusivity, which would be more consistent with a mere hypertrophy of cells that were already present before the intervention. [Tian et al. \(2014\)](#) provided initial evidence for such an association between higher fitness and lower hippocampal MD, showing that the amount of self-reported exercise activities predicts lower MD in medial temporal lobe (and cingulate cortex) about 10 years later in very old adults.

The present study investigates whether hippocampal tissue changes as captured by MD potentially mediate the association between fitness changes and changes in hippocampal volumes. Latent difference modeling ([McArdle and Nesselroade, 1994](#); cf. [Raz et al., 2005](#)), a variant of structural equation modeling (SEM), was used to represent individual differences in changes before and after the intervention. This method greatly reduces the problem of unreliability in the measurement of longitudinal change by using identical sets of more than one observed variable to define equivalent latent constructs at each measurement occasion, and then computing difference scores on the basis of these latent constructs. In doing so, the method effectively separates variance at the construct level from measurement-specific variance and error, and generates difference scores that are more reliable than difference scores based on observed variables (cf. [McArdle and Nesselroade, 1994](#)). To this end, we examined the covariance of change in fitness, hippocampal MD, and hippocampal volume in the context of a fitness intervention study with healthy sedentary older adults. Elderly participants were randomly assigned to either of two levels of fitness-training intensity. Before and after the 6-month training phase, participants performed a maximal graded exercise test to assess their training-related fitness improvements, and underwent T1-weighted and diffusion-weighted MR imaging. We hypothesized: (a) a negative association between changes in fitness and changes in MD, in that an increase in fitness would lead to lower diffusivity in the hippocampus, thus reflecting higher tissue density, and (b) a negative association between changes in MD and changes in volume, in that greater decrements in diffusivity, again reflecting more positive changes in tissue density, would lead to more positive changes in volume.

Materials & methods

Participants

Seven hundred and twenty-three community-dwelling older adults were contacted via local newspaper advertisements; from this initial

pool, fifty-seven individuals were enrolled in the study. Participants met the following inclusion criteria: (1) age between 59 and 75 years; (2) physical inactivity prior to study enrollment (metabolic rate < 40 based on the Freiburg Questionnaire of physical activity in German, [Frey et al., 1999](#)); (3) MMSE score ≥ 26 ; (4) free of neurological, psychiatric, or cardiovascular diseases; (5) right handedness; (6) no contraindication for heart-rate controlled exercise training (e.g., beta blockers); and (7) suitability for an MR environment (e.g., no magnetic implants, claustrophobia). During training, three participants dropped out due to health issues unrelated to the study, and two dropped out due to motivational issues, resulting in an effective final sample of fifty-two participants who completed the intervention (mean age = 66.0 years, SD = 4.36, age range = 59–74 years; 20 men). Participants gave informed written consent to the study procedure, which was approved by the Ethics Committee of the German Psychological Society (DGPs), and were paid for study participation. Study adherence was incentivized with a bonus system.

Design

Before (pretest) and after (posttest) the 6-month aerobic fitness intervention, as well as another six months after completing the intervention (maintenance), participants underwent a comprehensive assessment distributed over six testing sessions including a battery of questionnaires, cognitive tests, and motor tests as well as cardiovascular fitness assessment and an MR session. A mock scanner session was also included prior to the very first MR assessment to familiarize participants with the MR procedure.

Training

After completing the pretest assessment, participants were randomly assigned to either of two training regimens, high intensity (HI) or low intensity (LI). Groups were counterbalanced for age, sex, years of education, digit-symbol, and MMSE scores.

Participants in both groups came to the lab to exercise on stationary bikes. During the first three weeks, participants exercised twice a week, with each session lasting 25 min in the first week, 40 min in the second week and 55 min in the third week. After the third week, participants trained three times a week for 55 min each session. For the HI group, training intensity was calibrated to result in a heart rate at 80 % of the individual's ventilatory anaerobic threshold ([Wasserman et al., 1990](#)), as determined from a maximal graded exercise test at pretest. The LI group exercised at a constant resistance of 10 W irrespective of heart rates. During the last 21 sessions, 5 intervals of 2 min each were integrated after 20 min of training in order to further increase variance in fitness gains. Whereas the LI group only increased the cadence in these time windows from 60–70 to 80–90 cycles/min, the HI group also increased the intensity to a resistance corresponding to 110 % of the individual's ventilatory anaerobic threshold. To adapt intensity levels, heart rate was centrally and automatically monitored using the training software *custo cardio concept* (*custo med GmbH, Ottobrunn, Germany*). A staff member additionally controlled compliance for each participant and each training session. Participants exercised in groups of up to six persons at a time. Groups were not separated by intensity level, and participants were informed about differences between the two training regimens only after the termination of the study.

Cardiovascular fitness assessment

Aerobic fitness was assessed using a maximal graded exercise test on a cycle ergometer. The test started at 10 W, increased to 25 W after 2 min followed by 25 W increments every 2 min until total exhaustion or signs of cardiac or respiratory distress. A sports physician continuously monitored the cardiogram, oxygen uptake, heart rate, and blood pressure. As outcomes of the fitness assessment, four parameters

were assessed: the maximum oxygen consumption at exhaustion ($VO_2\text{max}$), the oxygen consumption at the ventilatory anaerobic threshold ($VO_2\text{AT}$), the maximum performance capacity ($W\text{max}$), and the performance capacity at the ventilatory anaerobic threshold ($W\text{AT}$). $VO_2\text{max}$ for each subject and each time point was only used if at least one of the following criteria was met: (1) leveling-off of the oxygen uptake, (2) respiratory ratio ≥ 1.11 , (3) maximum heart rate $\geq 200 - \text{age}$, and (4) maximum oxygen equivalent ≥ 30 . Note that values of performance capacity are always in relation to body weight, as heavier people can more easily achieve higher loads.

MRI data acquisition and processing

MRI protocol

Brain images were acquired on a Siemens TIM Trio 3 T MRI scanner (Siemens, Erlangen, Germany) using a 32-channel head coil. T1-weighted images were acquired using a standard magnetization-prepared rapid gradient-echo (MPRAGE) sequence (TR = 2500 ms, TE = 4.76 ms, TI = 1100 ms, flip angle = 7° , acquisition matrix = $256 \times 256 \times 176$, 1 mm isotropic voxels).

Diffusion-weighted images (DWI) were acquired using a single-shot EPI sequence with the following parameters: TR = 8000 ms, TE = 93 ms, 62 slices, 1 average, b-value 1000 s/mm^2 , 60 diffusion encoding directions, voxel-size $2 \times 2 \times 2 \text{ mm}^3$. In addition, 7 images without diffusion weighting were acquired, which were equally distributed between the diffusion-weighted images.

Image processing

T1-weighted images. In order to remove the image inhomogeneities from the 32-channel head coil before pre-processing, the T1-weighted images were bias corrected using the N4ITK bias field correction as implemented in Slicer4.1 (<http://www.slicer.org>, Fedorov et al., 2012). We made the following changes to the default settings: an increase in the number of iterations to 500, 400, 300 as well as a decreased shrink factor of 2. A binary brain mask was used to improve correction quality.

Cortical reconstruction and volumetric segmentation were performed using the FreeSurfer image analysis suite version 5.3 (<http://surfer.nmr.mgh.harvard.edu/>). To reduce variability of results across time points and thus increase the robustness and sensitivity of the overall longitudinal analysis, images were automatically processed with the longitudinal processing stream implemented in FreeSurfer (Reuter et al., 2012). Specifically an unbiased within-subject template was created from the two time points using robust, inverse consistent registration (Reuter et al., 2010). Several processing steps, such as skull stripping, Talairach transforms, atlas registration as well as spherical surface maps and parcellations were then initialized with common information from the within-subject template, significantly increasing reliability and statistical power (Reuter et al., 2012). All reconstructed data were visually checked for segmentation accuracy at each time point. No manual interventions with the MRI data were needed. Volumes of bilateral hippocampi in mm^3 were read out for each subject and each time point using FreeSurfer's "asegstats2table," and then included as observed variables in the LDMS.

To cross-validate the FreeSurfer analyses, we reanalyzed the T1-weighted images using FMRIB's model-based segmentation/registration tool FIRST (Patenaude et al., 2011) with its default settings, including affine registration to an MNI space template, segmentation of subcortical structures based on shape/appearance models constructed from manually segmented images, and boundary correction. Segmentations from each participant and at each time point were visually checked for segmentation accuracy. Volumes of bilateral hippocampi in mm^3 were read out for each subject and each time point to include them into the LDMS.

Diffusion-weighted images. Diffusion-weighted images (DWI) were first quality controlled using DTIprep (Oguz et al., 2014). This detects and corrects artifacts caused by eddy-currents, head motion, vibration and pulsation, venetian blind artifacts, as well as slice-wise and gradient-wise inconsistencies of signal intensity. Out of our 104 datasets, an average of 3 gradients per dataset was excluded. DWIs were then analyzed using the FMRIB Software Library (FSL, Jenkinson et al., 2012) as well as FMRIB's Diffusion Toolbox (FDT). Since motion and eddy current correction was already performed by DTIprep, we only generated a binary brain mask for the DWIs from the first non-diffusion-weighted image using the brain extraction tool (Smith, 2002) before fitting the tensor. Dtifit (part of the FDT tool in FSL) was used to fit a diffusion tensor model to every voxel included in the brain mask, producing the final diffusion maps. Participants' diffusion data were aligned into MNI space using nonlinear registration as implemented in TBSS (Smith et al., 2006). All diffusion maps were visually checked for registration accuracy at each time point. Mean diffusivity values in mm^2/s were extracted for bilateral hippocampi using anatomical regions of interest from the Harvard-Oxford Atlas.

Cognitive assessment

A comprehensive battery of cognitive tests was assessed, covering perceptual speed, executive control, episodic memory, reasoning, and vocabulary. We failed to establish a latent difference model for a pertinent cognitive ability such as episodic memory, working memory, or spatial navigation that would show reliable individual differences in change. Hence, we refrain from addressing associations between fitness-related changes in hippocampal integrity and cognitive changes in this article because the presence of reliable variance in change is a prerequisite for testing hypotheses on the covariance between changes (Hertzog et al., 2006, 2008).

Statistical analyses

Statistical analyses were performed using SPSS (IBM Corp., IBM SPSS Statistics, V22, Armonk, NY, USA) and R (R Foundation for Statistical Computing, V 3.1.2, Vienna, Austria). To test for effects on the observed variables, we used repeated measures ANOVAs with time point as within-subject factor and training group as between-subject factor as well as age and gender as covariates.

Structural equation modeling

Given that difference scores are particularly vulnerable to the effects of measurement error, we used latent difference models to analyze data on the latent level (cf. McArdle and Nesselroade, 1994; McArdle and Prindle, 2008; Raz et al., 2005; Schmiedek et al., 2010). This method allows separation of variance at the construct level from measurement error and indicator-specific variance by defining latent factors that represent the variance common to multiple indicators (i.e., observed variables). We established a latent factor of fitness based on the observed variables $VO_2\text{AT}$ and $W\text{AT}$, a latent factor of MD based on MD in the left and the right hippocampus, and a latent factor of hippocampal volume based on left and right hippocampal volumes as measured by FIRST and FreeSurfer. Using RAM notation (cf. Boker et al., 2002; McArdle and McDonald, 1984), Fig. 1 displays how latent constructs are established in the context of the latent difference score model, using the fitness factor as an example. We used the procedures described in Raz et al. (2005), with the notable exception that residual variances, which were constrained to be equal in Raz et al. (2005), were not constrained to be equal across time in the present paper in order to achieve acceptable model fit. Note, however, that the loadings of the indicators on the latent factors were constrained to be equal across time, just as in Raz et al. (2005). The resulting models retain the assumption of strong, but not strict (as in Raz et al., 2005) invariance of measurement across time.

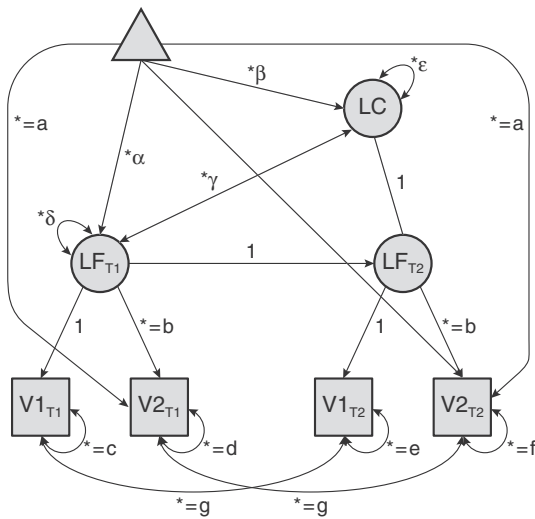


Fig. 1. Latent difference model with strong measurement invariance, adapted from Raz et al. (2005). Squares represent observed variables, circles represent latent factors, and the triangle serves to represent information regarding means and intercepts. V1–V2 = observed indicator variables (e.g., VO₂AT and WAT for fitness); LF = latent factor of construct (e.g., fitness); LC = latent change, (e.g. change in fitness from pretest to posttest); T1 = pretest (i.e., measures taken before fitness training); T2 = posttest (i.e., measures taken immediately after fitness training); α = latent mean of construct (e.g., fitness) at pretest; β = mean change in latent construct (e.g., fitness) from pretest to posttest; γ = latent covariance between individual differences at pretest and individual differences in changes; δ = latent variance at pretest; ϵ = latent variance in change; a = intercept of observed variable; b = loading of observed variable on latent factor; c, d, e, f = residual variances of observed variables; g = correlation between residuals; * = free parameter. Parameters with identical labels are constrained to be equal to each other to ensure identification and interpretability of the models.

To examine changes from pretest to posttest, we first computed univariate LDMs for the three constructs (fitness, MD, and volume) separately to examine whether (a) mean latent changes and (b) variances in change (i.e., individual differences in change) differed reliably from zero. Note that these two parameters vary freely, that is, individual differences in change can be reliably different from zero in the absence of reliable mean changes, and vice versa. Note also that interpreting correlations between changes would be dubious if one or both of the corresponding variances in change did not differ reliably from zero (cf. Ghisletta and Lindenberger, 2004). Age and gender were included as covariates when setting up the measurement model. All three constructs showed reliable variance in change. Hence we established a multivariate measurement model in which changes in all three constructs were allowed to correlate freely with one another (the corresponding R code can be found in Inline Supplementary R Code 1). We then imposed constraints on this measurement model, based on the guiding hypotheses of the present study. The resulting latent model specified unidirectional paths from changes in fitness to changes in hippocampal mean diffusivity, and from changes in hippocampal diffusivity to changes in hippocampal volume (see Inline Supplementary R Code 2). We used a chi-square test to examine whether this latent model, which is nested within the measurement model, was associated with a reliable decrement in fit. Given that we had clear a-priori hypotheses about the sign of the change–change associations, the statistical significance of the two path coefficients in the latent model was tested using one-tailed t-tests, with an alpha level of $p = 0.05$.

Inline Supplementary R Code 1 and 2 can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2015.11.026>.

To explore changes from posttest to maintenance, we used the same procedure, with one notable exception. As we had no strong hypotheses about the directionality of change–change associations during this phase of the study, we established a multivariate measurement model,

in which changes in all three constructs were allowed to correlate freely with one another, but refrained from imposing a path model.

Results

Changes from pretest to posttest

Participants of the two groups did not differ at baseline with respect to age, years of education, fitness, and MMSE (more detailed sample characteristics are presented in Table 1). Also, training adherence did not differ reliably between the two groups. Table 2 provides an overview of training effects on the variables of interest for the two groups.

Univariate latent difference modeling

For fitness, model fit for strong measurement invariance (see Fig. 1) was satisfactory ($\chi^2 = 10.334$, $df = 9$, CFI = 0.993, RMSEA = 0.053, SRMR = 0.065). For MD, a model with strong measurement invariance provided a satisfactory fit to the data ($\chi^2 = 9.384$, $df = 9$, CFI = 0.999, RMSEA = 0.029, SRMR = 0.050). For volume, model fit for strong measurement invariance was acceptable ($\chi^2 = 78.657$, $df = 38$, CFI = 0.956, RMSEA = 0.143, SRMR = 0.056).

Inspecting changes in latent means from pretest to posttest

The intervention was effective at increasing fitness levels as confirmed by a significant mean increase in fitness ($\beta = 0.413$, $SE = 0.183$, $p = 0.024$). Contrary to expectations, the HI and LI groups did not differ in mean level of fitness change ($\beta = 0.402$, $SE = 0.298$, $p = 0.178$). Hence, the two groups were collapsed into one for all analyses reported here.

We observed a mean increase in MD at the latent level ($\beta = 0.457$, $SE = 0.156$, $p = 0.003$), as well as a mean decrease in HC volume at the latent level ($\beta = -0.468$, $SE = 0.231$, $p = 0.043$). Fig. 2 depicts the mean changes for the three constructs.

Multivariate latent difference measurement model

For the multivariate measurement model with the factors fitness, volume, and MD, strong measurement invariance provided an acceptable model fit ($\chi^2 = 195.530$, $df = 131$, CFI = 0.958, RMSEA = 0.097, SRMR = 0.125). As depicted in Fig. 3, changes in fitness were related to changes in mean diffusivity ($\rho = -0.342$, $SE = 0.152$, $p = 0.024$), and changes in MD were related to changes in volume ($\rho = -0.488$, $SE = 0.215$, $p = 0.023$). In contrast, the association between changes in fitness and changes in volume was not statistically reliable ($\rho = 0.108$, $SE = 0.241$, $p = 0.655$).

Table 1

Sample characteristics for the two training groups. Values are means (s.d.). The metabolic rate was assessed via the Freiburg Questionnaire of physical activity (Frey et al., 1999). Abbreviations: MMSE: Mini-Mental State examination, BMI: Body mass index.

Characteristic	Group		p-Value
	High intensity	Low intensity	
N	25	27	
Male	10 (40.0%)	10 (37.0%)	
Age	66.10 (4.16)	65.93 (4.63)	$t_{(50)} = -0.142$, $p = 0.887$
Education (years)	11.0 (1.57)	11.13 (1.71)	$t_{(50)} = 0.056$, $p = 0.956$
MMSE	29.24 (1.33)	28.85 (1.41)	$t_{(50)} = -1.020$, $p = 0.313$
Training adherence (max 75)	70.72 (3.82)	69.11 (7.66)	$t_{(50)} = -0.946$, $p = 0.348$
Hormone replacement therapy (women)	5 (33.3%)	6 (35.3%)	
Smoking	1 (4.0%)	1 (3.7%)	
Treated hypertension	4 (16.0%)	3 (11.1%)	
BMI	26.60 (4.33)	25.35 (3.94)	$t_{(50)} = -1.089$, $p = 0.281$
Digit symbol	38.13 (8.89)	39.81 (6.31)	$t_{(48)} = 0.777$, $p = 0.441$
Metabolic rate (self-reported)	16.92 (10.50)	20.15 (11.14)	$t_{(50)} = 1.073$, $p = 0.289$

Table 2

Pre- and post-intervention values (mean ± s.d.) for fitness measures, hippocampal volume, and mean diffusivity. Abbreviations: FreeSurfer HC: hippocampal volumes based on FreeSurfer, FIRST HC: hippocampal volumes based on FIRST, MD: hippocampal mean diffusivity.

	High intensity		Low intensity	
	Pre	Post	Pre	Post
Vo ₂ max (ml/min/kg)	21.26 (± 4.76)	23.32 (± 5.06)	21.69 (± 5.76)	22.34 (± 5.29)
Vo ₂ AT (ml/min/kg)	15.60 (± 3.18)	16.54 (± 3.49)	15.20 (± 4.25)	16.05 (± 3.57)
Wmax (W/kg) ^{*,a}	1.66 (± 0.36)	1.87 (± 0.41)	1.73 (± 0.45)	1.81 (± 0.50)
WAT (W/kg)	1.01 (± 0.28)	1.14 (± 0.35)	1.10 (± 0.37)	1.11 (± 0.34)
FreeSurfer left HC (mm ³)	3692.15 (± 421.12)	3678.70 (± 423.04)	3824.37 (± 476.45)	3827.36 (± 497.01)
FreeSurfer right HC (mm ³)	3778.10 (± 399.20)	3746.16 (± 404.86)	3895.97 (± 468.86)	3889.15 (± 465.40)
FIRST left HC (mm ³)	3658.92 (± 203.53)	3612.32 (± 480.09)	3674.44 (± 477.30)	3648.56 (± 527.43)
FIRST right HC (mm ³)	3619.24 (± 479.92)	3572.00 (± 457.73)	3769.44 (± 447.25)	3773.63 (± 455.69)
MD left HC (mm ² /s)	0.0012 (± .0001)	.0013 (± .0001)	0.0012 (± .0001)	.0012 (± .0001)
MD right HC (mm ² /s)	0.0012 (± .0002)	.0013 (± .0002)	0.0012 (± .0001)	.0012 (± .0002)

* Indicates a significant time × group interaction from the repeated measures ANOVA (p < 0.05).

^a F(1,48) = 5.207; p = .027; η² = 0.098, follow-up ANOVAs separately for groups did not show significant changes (all p > .11).

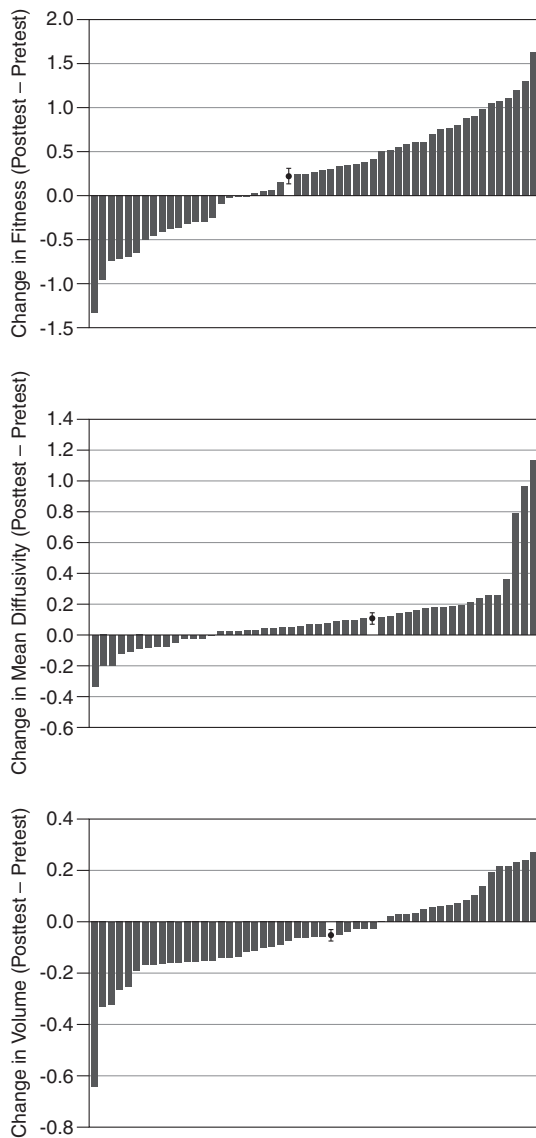


Fig. 2. Individual differences in change for fitness (top), hippocampal MD (middle), and hippocampal volume (bottom). Data for the two time points were pooled before z-transformation to preserve mean differences between time points, and then averaged across indicators. Each bar represents the posttest minus pretest difference for one participant. Difference scores are sorted in ascending order. Circles represent mean changes, with error bars corresponding to standard errors of the mean.

Multivariate latent difference path model

Next, we fit the data to a path model corresponding to the guiding hypotheses of the present study that an increase in fitness leads to lower diffusivity in the hippocampus, and that greater decrements (or lower increments) in diffusivity, reflecting more positive changes tissue density, lead to more positive (or less negative) changes in volume. This path model was nested within the multivariate latent difference measurement model. A direct comparison between the two models revealed that the path model was not associated with a reliable decrease in fit in comparison to the measurement model (−2 log likelihood ratio = 1.265; df = 1; p = 0.261). The hypothesized path model is shown in Fig. 4. The fit of the model to the data was satisfactory (χ² = 196.795, df = 132, CFI = 0.958, RMSEA = 0.097, SRMR = 0.125). According to this model, changes in fitness account for changes in MD (β = −0.330 SE = 0.173, p = 0.028) and changes in MD lead to changes in volume (β = −0.488, SE = 0.212, p = 0.011).

Changes from posttest to maintenance

Univariate latent difference modeling

For fitness, model fit for strict measurement invariance was satisfactory (χ² = 14.153, df = 12, CFI = 0.990, RMSEA = 0.059, SRMR = 0.054). For MD, a model with strong measurement invariance provided a satisfactory fit to the data (χ² = 14.764, df = 9, CFI = 0.986, RMSEA = 0.111, SRMR = 0.054). For volume, model fit for strong measurement

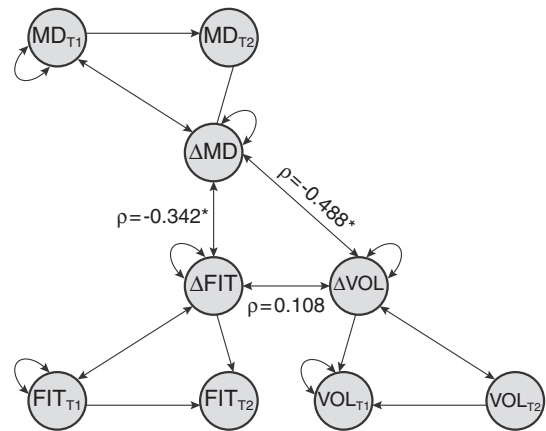


Fig. 3. Multivariate latent difference model. Changes in fitness were reliably associated with changes in MD, and changes in MD were reliably associated with changes in volume. Only latent constructs are shown. Abbreviations: FIT: latent factor fitness, VOL: latent factor volume, MD: latent factor mean diffusivity, T1: before training, T2: after training, Δ: latent change. * indicates p < 0.05.



Fig. 4. Path model. Only latent constructs are shown. Abbreviations: FIT: fitness, VOL: volume, MD: mean diffusivity, Δ : latent change. * indicates $p < 0.05$.

invariance was acceptable ($\chi^2 = 81.693$, $df = 38$, $CFI = 0.951$, $RMSEA = 0.149$, $SRMR = 0.064$).

Inspecting changes in latent means from posttest to maintenance

Six months after cessation of the controlled intervention mean fitness levels decreased significantly ($\beta = -0.480$, $SE = 0.172$, $p = 0.005$), and the two intensity groups again did not differ in mean level of fitness change ($\beta = -0.211$, $SE = 0.287$, $p = 0.461$). Mean latent changes in MD did not differ reliably from zero ($\beta = -0.012$, $SE = 0.173$, $p = 0.946$), nor did latent changes in volume ($\beta = 0.159$, $SE = 0.208$, $p = 0.445$).

Multivariate latent difference measurement model

We established a multivariate measurement model with the factors fitness, volume, and MD, assuming strict measurement invariance for fitness, and strong measurement invariance for MD and volume over time. This model yielded an acceptable model fit ($\chi^2 = 241.285$, $df = 151$, $CFI = 0.943$, $RMSEA = 0.107$, $SRMR = 0.145$). Changes from posttest to maintenance were not reliably associated across the three constructs (changes in fitness and changes in mean diffusivity: $\rho = -0.048$, $SE = 0.163$, $p = 0.768$, changes in MD and changes in volume: $\rho = -0.339$, $SE = 0.208$, $p = 0.103$, changes in fitness and changes in volume: $\rho = -0.000$, $SE = 0.217$, $p = 0.999$).

Discussion

In this study we investigated whether exercise-induced fitness improvements are associated with changes in mean diffusivity and volumes of the hippocampus. Using latent difference modeling (McArdle and Nesselroade, 1994), we were able to show that exercise-induced changes in fitness are associated with changes in hippocampal MD, and that changes in hippocampal MD are associated with changes in hippocampal volumes. The path model shown in Fig. 4 is consistent with the assumption of a temporal sequence in which fitness changes lead to changes in MD, which in turn lead to changes in volume. However, we note that a stringent empirical test of this assumption would require more than two time points.

Exercise-related changes in hippocampal volume have repeatedly been found (Erickson et al., 2011; Maass et al., 2015) but the mechanisms underlying these changes are not well understood. Given that exercise targets the cardiovascular system, improvements in brain perfusion are likely to be involved. In middle-aged adults, Pereira et al. (2007) provided initial evidence for a fitness-related increase in hippocampal perfusion. These results were recently replicated and extended in a sample of older adults by Maass et al. (2015), who reported an association between hippocampal perfusion and hippocampal volume. The association between hippocampal perfusion and hippocampal volume may reflect dilation of existing blood vessels, angiogenesis, or both. To the extent that changes in MD represent changes in the barrier density of gray matter, our findings suggest that the effects of fitness changes on volume may not be confined to dilation of preexisting vasculature but may also trigger angiogenesis. This is in line with cross-sectional findings by Bullitt et al. (2009) showing that highly active older adults exhibit a larger number of small-diameter vessels as assessed by MR angiography, and it is also consistent with relevant animal models (Bloor, 2005; Van der Borgh et al., 2009). Angiogenesis may then trigger neurogenesis, synaptogenesis, or changes in the neuropil, leading to an increase in tissue volume.

In contrast to previous work (Erickson et al., 2011; Maass et al., 2015), the volumetric changes observed in this study were not directly linked to changes in fitness. The present training program included an interval phase, during which participants in the HI group were pushed above their anaerobic threshold, whereas participants in LI groups only increased the cadence. Possibly, this may have been a stressor for some participants in the HI group. Given that stress appears to suppress adult neurogenesis (Kim and Diamond, 2002; Lucassen et al., 2010), the interval phase may have counteracted some of the positive effects of fitness training. Some animal studies reported a negative association between hippocampal neurogenesis and exercise duration (Naylor et al., 2005) and intensity (Lou et al., 2008), with the latter being quite comparable to our setup. In line with this interpretation, visual inspection of the volumetric changes suggested increased volumes in the LI group as compared to stable volumes in the HI group, though the difference was not statistically reliable.

Unfortunately, the present study lacks measures of hippocampal perfusion and/or angiography and therefore cannot disentangle the contribution of these different factors to the observed volumetric changes. More research using quantifiable and mechanistically interpretable MR parameters are needed to elucidate the ways in which fitness training changes the physiology of the aging human brain, and helps to promote healthy cognitive aging.

Unlike most published reports on fitness interventions, we used LDM to analyze associations among changes. LDM establishes latent constructs of interest by using multiple indicators and represents variance in change at the latent level. It effectively separates construct variance from residual variance, and is likely to increase statistical power as well as replicability of results across studies. We encourage more frequent use of the LDM approach in intervention research. However, we note that the method is based on strong (but testable) assumptions about measurement invariance, and requires moderate to large sample sizes.

Contrary to expectations, we found no group differences in fitness changes between the HI and the LI groups. Perhaps the two training regimes were too similar for sedentary adults, in the sense that the LI training may have been more effective in boosting fitness than expected. Furthermore, genetic factors seem to modulate exercise-induced gains in fitness (Peter et al., 2014), thereby potentially blurring effects at the group level.

The sample size of the present study is still relatively small. In addition, the present results may, to an unknown degree, reflect a bias in participation selection. Participants in the present study followed a highly sedentary lifestyle but, at the same time, were sufficiently healthy to be admitted to the study, and felt sufficiently motivated to engage in an aerobic exercise intervention program. This combination of features may have led to the selection of a particularly resilient sedentary group. It remains to be seen whether the present results replicate in larger and more representative samples.

From posttest to maintenance, fitness, MD, and hippocampal volume showed reliable variance in change, but, in contrast to the intervention period, these changes were no longer associated across constructs. This may suggest that the different constructs vary in the rate of adaptation, that participants differ in the degree to which they continued exercising, or both. In the present study, the level of exercise between posttest and maintenance was only assessed via monthly self-report questionnaires; more detailed, objective measures are lacking. Hence the interpretation of these differences in change is unclear.

Finally, it should be kept in mind that the hypothesized sequence of changes postulated in this article is empirically under-identified. According to this sequence, which corresponds to the path model shown in Fig. 4, changes in fitness lead to changes in hippocampal tissue density, which in turn lead to changes in volume. Clearly, more than two time points are needed to observe these changes separately in time in order to test the veridicality of this sequence more directly. Future research should rely to a greater extent on research designs with multiple

scanning sessions distributed before, during, and after the intervention to identify the trajectories and dynamics of experience-dependent plasticity (cf. Lövdén et al., 2013).

Conclusions

In this study we found that exercise-induced fitness changes are associated with microstructural brain changes in the hippocampus as indexed by MD, and that changes in hippocampal MD are associated with changes in hippocampal volume. These results suggest that previously reported increases in hippocampal volume may in part be driven by an increase in cell membranes operating as diffusion barriers.

Acknowledgments

This study was conducted within the “Sensorimotor-Cognitive Couplings” project at the Center for Lifespan Psychology of the Max Planck Institute for Human Development, in partial fulfilment of the doctoral dissertation of MMK. MMK participated in the International Max Planck Research School on the Life Course (LIFE; www.imprs-life.mpg.de). We cordially thank our participants for their long-term commitment to this study. We are also very grateful to Gabriele Faust for her assistance in conducting the study and to our student assistants for their great support in data collection.

References

- Bammer, R., 2003. Basic principles of diffusion-weighted imaging. *Eur. J. Radiol.* 45, 169–184.
- Barnes, J., Bartlett, J.W., van de Pol, L.A., Loy, C.T., Scahill, R.I., Frost, C., Thompson, P., Fox, N.C., 2009. A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. *Neurobiol. Aging* 30, 1711–1723.
- Beaulieu, C., 2002. The basis of anisotropic water diffusion in the nervous system—a technical review. *NMR Biomed.* 15, 435–455.
- Bherer, L., Erickson, K.I., Liu-Ambrose, T., 2013. A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. *J. Aging Res.* 2013, 657508.
- Bloor, C.M., 2005. Angiogenesis during exercise and training. *Angiogenesis* 8, 263–271.
- Boker, S.M., McArdle, J.J., Neale, M., 2002. An algorithm for the hierarchical organization of path diagrams and calculation of components of expected covariance. *Struct. Equ. Model.* 9, 174–194.
- Bullitt, E., Rahman, F.N., Smith, J.K., Kim, E., Zeng, D., Katz, L.M., Marks, B.L., 2009. The effect of exercise on the cerebral vasculature of healthy aged subjects as visualized by MR angiography. *AJNR Am. J. Neuroradiol.* 30, 1857–1863.
- 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. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3017–3022.
- Erickson, K.I., Hillman, C.H., Kramer, A.F., 2015. Physical activity, brain, and cognition. *Curr. Opin. Behav. Sci.* 4, 27–32.
- Eriksson, P.S., Perflieva, E., Björk-Eriksson, T., Alborn, A.M., Nordborg, C., Peterson, D.A., Gage, F.H., 1998. Neurogenesis in the adult human hippocampus. *Nat. Med.* 4, 1313–1317.
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J.V., Pieper, S., Kikinis, R., 2012. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* 30, 1323–1341.
- Frey, I., Berg, A., Grathwohl, D., Keul, J., 1999. Freiburger Fragebogen zur körperlichen Aktivität—Entwicklung, Prüfung und Anwendung. *Soz. Präventivmed.* 44, 55–64.
- Ghisletta, P., Lindenberger, U., 2004. Static and dynamic longitudinal structural analyses of cognitive changes in old age. *Gerontology* 50, 12–16.
- Hertzog, C., Lindenberger, U., Ghisletta, P., von Oertzen, T., 2006. On the power of multivariate latent growth curve models to detect correlated change. *Psychol. Methods* 11, 244–252.
- Hertzog, C., von Oertzen, T., Ghisletta, P., Lindenberger, U., 2008. Evaluating the power of latent growth curve models to detect individual differences in change. *Struct. Equ. Model.* 15, 541–563.
- Hötting, K., Röder, B., 2013. Beneficial effects of physical exercise on neuroplasticity and cognition. *Neurosci. Biobehav. Rev.* 37, 2243–2257.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. *Fsl. NeuroImage* 62, 782–790.
- Kempermann, G., 2012. New neurons for ‘survival of the fittest’. *Nat. Rev. Neurosci.* 13, 727–736.
- Kim, J.J., Diamond, D.M., 2002. The stressed hippocampus, synaptic plasticity and lost memories. *Nat. Rev. Neurosci.* 3, 453–462.
- Kronenberg, G., Reuter, K., Steiner, B., Brandt, M.D., Jessberger, S., Yamaguchi, M., Kempermann, G., 2003. Subpopulations of proliferating cells of the adult hippocampus respond differently to physiologic neurogenic stimuli. *J. Comp. Neurol.* 467, 455–463.
- Lou, S.J., Liu, J.Y., Chang, H., Chen, P.J., 2008. Hippocampal neurogenesis and gene expression depend on exercise intensity in juvenile rats. *Brain Res.* 1210, 48–55.
- Lövdén, M., Wenger, E., Mårtensson, J., Lindenberger, U., Bäckman, L., 2013. Structural brain plasticity in adult learning and development. *Neurosci. Biobehav. Rev.* 37, 2296–2310.
- Lucassen, P.J., Meerlo, P., Naylor, A.S., van Dam, A.M., Dayer, A.G., Fuchs, E., Oomen, C.A., Czéh, B., 2010. Regulation of adult neurogenesis by stress, sleep disruption, exercise and inflammation: implications for depression and antidepressant action. *Eur. Neuropsychopharmacol.* 20, 1–17.
- 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. *Mol. Psychiatry* 20, 585–593.
- McArdle, J.J., McDonald, R.P., 1984. Some algebraic properties of the reticular action model for moment structures. *Br. J. Math. Stat. Psychol.* 37, 234–251.
- McArdle, J.J., Nesselroade, J.R., 1994. Using multivariate data to structure developmental change. In: Cohen, S.H., Reese, H.W. (Eds.), *Life-Span Developmental Psychology: Methodological Contributions*. Lawrence Erlbaum Associates, Inc, Hillsdale, NJ, England, pp. 223–267.
- McArdle, J.J., Prindle, J.J., 2008. A latent change score analysis of a randomized clinical trial in reasoning training. *Psychol. Aging* 23, 702–719.
- Naylor, A.S., Persson, A.L., Eriksson, P.S., Jonsdottir, I.H., Thorlin, T., 2005. Extended voluntary running inhibits exercise-induced adult hippocampal progenitor proliferation in the spontaneously hypertensive rat. *J. Neurophysiol.* 93, 2406–2414.
- Niemann, C., Godde, B., Voelcker-Rehage, C., 2014. Not only cardiovascular, but also coordinative exercise increases hippocampal volume in older adults. *Front. Aging Neurosci.* 6, 170.
- Oguz, I., Farzinfar, M., Matsui, J., Budin, F., Liu, Z., Gerig, G., Johnson, H.J., Styner, M., 2014. DTIPrep: quality control of diffusion-weighted images. *Front. Neuroinform.* 8, 4.
- Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* 56, 907–922.
- Pereira, A.C., Huddleston, D.E., Brickman, A.M., Sosunov, A.A., Hen, R., McKhann, G.M., Sloan, R., Gage, F.H., Brown, T.R., Small, S.A., 2007. An in vivo correlate of exercise-induced neurogenesis in the adult dentate gyrus. *Proc. Natl. Acad. Sci. U. S. A.* 104, 5638–5643.
- Peter, I., Papandonatos, G.D., Belalcazar, L.M., Yang, Y., Erar, B., Jakicic, J.M., Unick, J.L., Balasubramanyam, A., Lipkin, E.W., Delahanty, L.M., Wagenknecht, L.E., Wing, R.R., McCaffery, J.M., Huggins, G.S., Look, A.R.G., 2014. Genetic modifiers of cardiorespiratory fitness response to lifestyle intervention. *Med. Sci. Sports Exerc.* 46, 302–311.
- Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., Acker, J.D., 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cereb. Cortex* 15, 1676–1689.
- Reuter, M., Rosas, H.D., Fischl, B., 2010. Highly accurate inverse consistent registration: a robust approach. *NeuroImage* 53, 1181–1196.
- Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 61, 1402–1418.
- Schmiedek, F., Lövdén, M., Lindenberger, U., 2010. Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Front. Aging Neurosci.* 2, 27.
- Smith, S.M., 2002. Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.
- 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., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *NeuroImage* 31, 1487–1505.
- Spalding, K.L., Bergmann, O., Alkass, K., Bernard, S., Salehpour, M., Huttner, H.B., Boström, E., Westerlund, I., Vial, C., Buchholz, B.A., Possnert, G., Mash, D.C., Druid, H., Frisén, J., 2013. Dynamics of hippocampal neurogenesis in adult humans. *Cell* 153, 1219–1227.
- Squire, L.R., Stark, C.E., Clark, R.E., 2004. The medial temporal lobe. *Annu. Rev. Neurosci.* 27, 279–306.
- Thomas, A.G., Dennis, A., Bandettini, P.A., Johansen-Berg, H., 2012. The effects of aerobic activity on brain structure. *Front. Psychol.* 3, 86.
- 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. *J. Gerontol. A Biol. Sci. Med. Sci.* 69, 1284–1290.
- Van der Borgh, K., Kóbor-Nyakas, D.E., Klauke, K., Eggen, B.J., Nyakas, C., Van der Zee, E.A., Meerlo, P., 2009. Physical exercise leads to rapid adaptations in hippocampal vasculature: temporal dynamics and relationship to cell proliferation and neurogenesis. *Hippocampus* 19, 928–936.
- van Praag, H., Kempermann, G., Gage, F.H., 1999. Running increases cell proliferation and neurogenesis in the adult mouse dentate gyrus. *Nat. Neurosci.* 2, 266–270.
- Voss, M.W., Vivar, C., Kramer, A.F., van Praag, H., 2013. Bridging animal and human models of exercise-induced brain plasticity. *Trends Cogn. Sci.* 17, 525–544.
- Wasserman, K., Beaver, W.L., Whipp, B.J., 1990. Gas exchange theory and the lactic acidosis (anaerobic) threshold. *Circulation* 81, II14–II30.