LONGITUDINAL AGING RESEARCH: COGNITION, BEHAVIOR AND NEUROSCIENCE

EDITED BY: Lutz Jäncke, Mike Martin, Christina Röcke and Susan Mérillat

PUBLISHED IN: Frontiers in Human Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-830-1
DOI 10.3389/978-2-88976-830-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

LONGITUDINAL AGING RESEARCH: COGNITION, BEHAVIOR AND NEUROSCIENCE

Topic Editors:
Lutz Jäncke, University of Zurich, Switzerland
Mike Martin, University of Zurich, Switzerland
Christina Röcke, University of Zurich, Switzerland
Susan Mérillat, University of Zurich, Switzerland

Citation: Jäncke, L., Martin, M., Röcke, C., Mérillat, S., eds. (2022). Longitudinal Aging Research: Cognition, Behavior and Neuroscience. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-830-1

Table of Contents

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 14 February 2020
doi: 10.3389/fnhum.2020.00043

The Relation Between Steroid Secretion Patterns and the Androgen Receptor Gene Polymorphism on Physical Health and Psychological Well-Being—Longitudinal Findings From the Men's Health 40+ Study

Tim Jonas Lacker[1,2], *Andreas Walther*[1,2,3], *Serena Fiacco*[1,2] *and Ulrike Ehlert*[1,2]*

[1]Clinical Psychology and Psychotherapy, University of Zurich, Zurich, Switzerland, [2]University Research Priority Program (URPP) Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland, [3]Biopsychology, TU Dresden, Dresden, Germany

OPEN ACCESS

Edited by:
Claudia Buss,
Charité Medical University of Berlin, Germany

Reviewed by:
Oliver C. Schultheiss,
University of Erlangen Nuremberg, Germany
Robert-Paul Juster,
Université de Montréal, Canada

***Correspondence:**
Ulrike Ehlert
u.ehlert@psychologie.uzh.ch

Specialty section:
This article was submitted to Health, a section of the journal Frontiers in Human Neuroscience

Received: 25 September 2019
Accepted: 29 January 2020
Published: 14 February 2020

Citation:
Lacker TJ, Walther A, Fiacco S and Ehlert U (2020) The Relation Between Steroid Secretion Patterns and the Androgen Receptor Gene Polymorphism on Physical Health and Psychological Well-Being—Longitudinal Findings From the Men's Health 40+ Study. Front. Hum. Neurosci. 14:43. doi: 10.3389/fnhum.2020.00043

Research is increasingly focusing on promoting healthy aging and the related extension of the health span by targeting crucial biological processes responsible for age-related conditions. While age-related gradual changes in steroid hormones such as testosterone, estradiol, or cortisol are well described in men, their interactions among each other or with genetic markers have not been sufficiently investigated with regard to physical health or psychological well-being. More specifically, the examination of age-related alterations in hormone interactions and the androgen receptor polymorphism, which modulates androgen action on target cells, in relation to physical health and psychological well-being represents a promising avenue for research on healthy aging in men. A total of 97 healthy aging men provided complete data on psychometric health measures as well as hormonal and genetic parameters at baseline and a 4-year follow-up assessment. Fasting saliva samples were taken at 8:00 am under standardized laboratory conditions, while the androgen receptor gene polymorphism was analyzed from dried blood spots. Longitudinal analyses revealed that psychological well-being and physical health remained stable over time. Analyses indicated that E2 moderated the course of psychological well-being, while the androgen receptor gene polymorphism moderated the course of physical health. Further, T was a strong predictor of physical health. These results suggest that the hypothalamic-pituitary-gonadal (HPG) axis might be important for the maintenance of psychological well-being in men, while physical health depends more on interindividual differences in the androgen receptor gene and T.

Keywords: healthy aging, men's health, androgen receptor, steroid hormones, gerontopsychology, physical health, mental health, biopsychosocial

INTRODUCTION

Research is increasingly focusing on promoting healthy aging and the associated extension of the health span by targeting crucial biological processes responsible for age-related health conditions (Barzilai et al., 2018). Healthy aging has become one of the leading health care goals of our time, with the World Health Organization (WHO) recently declaring a "Decade of Action on Healthy Aging" (World Health Organization, 2017). Generally speaking, healthy aging can be defined as the "process of developing and maintaining the functional ability that enables well-being in older age" (World Health Organization, 2015b, p. 28). Overall, life expectancy has increased dramatically over the last 150 years, rising even more sharply in the last few decades (World Health Organization, 2015a). At the same time, due to a slower increase in birth rates, the world population is facing a growing proportion of older individuals (Beard et al., 2016). Living the gained years of life in good health is a prioritized aim not only for each individual, but also for society in general, in order to reduce the large economic burden of chronic diseases in the elderly (Beard and Bloom, 2015; World Health Organization, 2017). To achieve this, it is crucial to investigate biological influences (e.g., hormone secretion and genetic markers) on physical health and well-being, comprising both physical and psychological aspects.

Psychological well-being changes throughout the lifespan. It is mostly measured based on the individual's subjective perception and comprises three essential components: life satisfaction, presence of positive affect and absence of negative affect (Diener and Lucas, 1999). Subjective well-being has been found to be relatively stable until very old age, whereupon it declines due to health constraints. This stability despite age-related physical health decline has been called the well-being paradox (Kunzmann et al., 2000; Hansen and Slagsvold, 2012). The individual psychological well-being can be influenced by various biological factors such as hormone secretion, with the mood-altering properties of testosterone (T) being especially crucial for aging males (Walther and Ehlert, 2015). Furthermore, genetics plays an important role, not only in physiology but also in subjective well-being (Weiss et al., 2008). More specifically, the androgen receptor gene polymorphism has been linked to personality factors that are closely associated with well-being (Weiss et al., 2008; Westberg et al., 2009). While another study with three distinct samples did not replicate the finding that the androgen receptor gene polymorphism functions as a predictor of well-being in the general population sample, it was a predictor in the two patient samples (Schneider et al., 2010). However, the aforementioned study assessed general well-being rather than specifically psychological well-being.

In most individuals, physical health deteriorates with higher age and often results in frailty and multimorbidity (Huisman et al., 2011; Lowsky et al., 2013; Guthrie and Boyd, 2018; Rockwood, 2019). However, a general population study found that even in a group of 8,059 individuals aged 85 years and older, 56% of participants reported that they did not experience any health-related limitations (Lowsky et al., 2013). Thus, one important marker of healthy aging is the subjective physical health status (Han et al., 2015). Previous research has described multiple potential biological influences, such as age-related changes in hormone secretion patterns or genetic markers (Feldman et al., 2002; Tirabassi et al., 2015; Walther et al., 2016).

Steroid hormones show distinct trajectories of stability, decrease or increase throughout the aging process. With regard to the hypothalamic-pituitary-gonadal (HPG) axis, the primary male sex steroid T significantly decreases with age, whereas estradiol (E2) shows stability or a slight decline (Vermeulen et al., 2002; Fiacco et al., 2018). For example, the effect size of the T decrease has been shown to be $r = 0.46$ in an influential population-based study with 1156 individuals (Feldman et al., 2002). These changes can mainly be attributed to the decreased function or attrition of the Leydig cells in the testes. In addition, increased concentrations of circulating sex hormone-binding globulin (SHBG) and a higher conversion rate of T to E2 by the enzyme aromatase within the body's fat cells further lead to markedly reduced testosterone levels (Vermeulen, 2000; Samaras et al., 2012). The hypothalamic-pituitary-adrenal (HPA) axis also undergoes distinct changes throughout the aging process, leading to increased cortisol secretion (Karlamangla et al., 2013). The increase in C is a consequence of age-related changes in glucocorticoid and mineralocorticoid receptor quantities, increased corticotropin-releasing hormone secretion from the hypothalamus and decreased sensitivity of the feedback mechanisms of the HPA axis (Zhou and Swaab, 1999; Gupta and Morley, 2014). Following the principle of homeostasis (Cannon, 1929; Joseph and Whirledge, 2017), mutual adaptations between the HPA and HPG axes aiming for an endocrine balance, and also within the HPG axis itself, have been suggested as important factors determining an individual's health (Sollberger and Ehlert, 2016). Thus, it is important to respect the interaction between the different hormones of the two main endocrine axes.

Investigating the interaction between different steroid hormones can be a promising and insightful strategy to investigate the interplay of different hormonal axes and the interplay of hormonal end products within a single endocrine axis (Sollberger and Ehlert, 2016). A prospective cohort study on over 2,500 men, investigating the hormone ratio and coronary heart disease, found positive associations of the T and C ratio with the risk of cardiovascular events (Smith et al., 2005). The same study further linked the T and C ratio to physical health factors such as BMI or blood pressure. Moreover, emotion regulation, which is closely related to psychological well-being, was positively associated with the T and C ratio in 19 participants of an fMRI study in healthy men (Denson et al., 2013). For another prominent hormone ratio, the T and E2 ratio (TE2r), results indicated a positive association with disease risk (Schneider et al., 2010; Gong et al., 2013). To the best of our knowledge, there are not yet any studies investigating the TE2r with respect to psychological well-being. Since hormone ratios can be criticized, obscuring the influence of each hormone on the effect, it is important to additionally respect literature on hormone interactions without the calculation of ratios (Sollberger and Ehlert, 2016). In a recent review, T has been described as influencing physical health factors such as frailty

in elderly men (Gordon and Hubbard, 2018). T has also been found to have a positive influence on psychological well-being, while a rather low C secretion can have the same effect. Studies investigating E2 need to be done, yet (Rector and Friedman, 2018). In general, since the field of positive psychology is very young, studies specifically investigating the association between hormones and psychological well-being or their interaction are very scarce (Rector and Friedman, 2018).

Genetic differences in the androgen receptor gene influence androgen action on target cells and modulate the association between androgens and physical health and well-being. The androgen receptor gene is located on the X chromosome at the location q11-q12 and comprises eight exons. The first exon of the androgen receptor gene contains a distinct cytosine-adenine-guanine (CAG) repeat sequence, which varies in repeat length between 10 and 36 for healthy individuals (Tirabassi et al., 2015). This variation has been suggested to modulate the androgen receptor binding affinity, affecting the androgen receptor action in response to receptor binding (Chamberlain et al., 1994; Rajender et al., 2008). This could lead to altered availability and function of androgens and, since the endocrine system is dynamic in nature, the HPA axis is influenced in its' function by this as well (Terburg et al., 2009). Previous research on the CAG repeat length and endocrine secretion has yielded mixed results, with one population-based longitudinal study reporting negative associations with T levels (Krithivas et al., 1999) and a more recent study failing to confirm this association (Eendebak et al., 2016). Another study suggested that higher CAG repeat length is not associated with altered T secretion, but is associated with increased E2 action (Huhtaniemi et al., 2009). Furthermore, associations with psychological traits have been found. More specifically, the CAG repeat length was negatively associated with aggression ($r = -0.365$) in a study of 645 convicted criminals, showing a moderate effect size (Rajender et al., 2008). Furthermore, individuals in the shorter CAG repeats group showed higher values on extraversion and neuroticism in a study of 141 individuals (Westberg et al., 2009). All of the above-mentioned traits can influence psychological well-being (Strickhouser et al., 2017). Finally, the CAG repeat length has previously been associated with health conditions. A low CAG repeat was associated with a greater risk of prostate cancer, with odds ratios of up to 3.7, in an influential review article (Nelson and Witte, 2002), while a higher CAG repeat was associated with infertility in a study comparing 37 infertile with 50 fertile individuals (Wallerand et al., 2001).

The biopsychosocial model (Engel, 1980; Wade and Halligan, 2017) postulates interactions between psychosocial and biological markers. However, to date, no study has investigated the interplay between the influence of hormonal axes and the androgen receptor gene on changes in psychological well-being and subjective physical health specifically in healthy aging men. We hypothesize that physical health decreases, psychological well-being remains stable and both, the TCr and TE2r decrease. Furthermore, we hypothesize that the interactions between the assessed steroid hormones, through their rather short-term secretory nature, influence the change in psychological well-being, while the CAG repeat length moderates the change in physical function.

MATERIALS AND METHODS

Participants

For the current study, we longitudinally investigated 97 healthy aging men from the Men's Health 40+ study at two timepoints (Walther et al., 2016). All men were between 40 and 75 years of age at the time of the baseline measurement. To ensure the investigation of healthy aging men at both time points, participants were screened before inclusion in the study. First, all participants were asked "How would you describe your current health?," which is the first item of the Short Form 36 Health Survey (Bullinger and Kirchberger, 1998). This item is rated on a scale encompassing very bad, bad, good, very good and excellent. To be eligible for study inclusion, participants had to be fluent in the German language and had to describe their health as at least good at both time points, suggesting that participants perceived a continuously high level of subjective health during the time of study. Participants who reported acute or chronic medical conditions, psychiatric disorders, psychopharmacological treatment (including T supplementation) or psychotherapy were excluded. The study was approved by the local ethics committee and all participants accepted the terms of the study and gave their written informed consent.

Procedure

The study comprised two timepoints spanning 4 years. At each time point, the study was divided into two parts. The first part consisted of three consecutive questionnaire batteries, each lasting for 1 h, including questionnaires measuring subjective health and physical health. For the second part of the study, participants underwent a laboratory assessment at the laboratory facilities of the psychological institute. During this assessment, participants provided saliva samples and blood samples and underwent various other physiological measures such as bioelectrical impedance analysis to investigate the body composition. Detailed information on the study procedure is reported elsewhere (Walther et al., 2016).

Questionnaire

For the investigation of psychological well-being and physical health, the German version of the Short Form 36 Health Survey (SF-36) questionnaire was used (Bullinger and Kirchberger, 1998). The SF-36 is a widely used, well-validated self-report questionnaire consisting of eight subscales, of which the psychological well-being and physical health subscales were used for the current study. The internal consistency for the scales used was in line with the originally reported values, being mostly over $\alpha = 0.70$. For the physical health scale, the internal consistency was $\alpha = 0.66$ at baseline and $\alpha = 0.70$ at follow-up. For the psychological well-being scale, the internal consistency was $\alpha = 0.80$ at baseline and $\alpha = 0.73$ at follow-up.

Hormone Analyses

All hormone samples were taken between 8:00 and 8:15 a.m. during the laboratory assessment. Participants provided fasting

saliva samples (SaliCaps, IBL International GmbH, Hamburg, Germany) under the supervision of trained study personnel. After collection, samples were stored at −20°C until analysis. T and E2 were analyzed using a luminescence immunoassay (Goncharov et al., 2006). The intra- and inter-assay variation for T and E2 was below 10%, with sensitivities of 1.8 pg/ml for T and 0.3 pg/ml for E2. C was measured using an enzyme-linked immunosorbent assay, with an intra- and inter-assay variation below 10% and a sensitivity of 30 pg/ml (Chiappin et al., 2007). To standardize the units, C was transformed from nmol/L to pg/ml. Afterward, hormone ratios were calculated. After calculation, both ratios were log-transformed to smooth the distribution and therefore were then normally distributed (Sollberger and Ehlert, 2016).

Genetic Analyses

Blood samples were assessed using the dried blood spot method (Fischer et al., 2019b), in which participants provided four drops of blood on a filter paper consisting of pure cellulose (Protein Saver Snap Apart, Forest Farm Industrial Estate, Cardiff, UK). Analysis of DNA with dried blood spots has been shown to be a reliable and valid procedure to analyze polymorphisms (Demirev, 2013). The sample collection took place under standardized laboratory conditions immediately after saliva samples were taken. After drying the samples for at least 4 h, they were frozen and stored at −20°C until analysis. During analysis, genomic DNA was extracted from three filter paper punches of 3 mm diameter each, using the QIAGEN QIAamp DNA Investigator Kit (Qiagen, Hombrechtikon, Switzerland). The genetic analyses themselves were conducted applying the capillary electrophoresis method, using the Applied Biosystems 3730XL Sequencer (Thermo Fisher Scientific, Waltham, MA, USA) with protocols and primers according to Westberg et al. (2001). For the determination of the fragment lengths, the GeneMapper Software v3.7 (Thermo Fisher Scientific, Waltham, MA, USA) was used.

Statistical Analysis

To investigate the association of steroid hormone and their interactions and the androgen receptor gene polymorphism with the change in psychological well-being and physical health over time, two different sets of analyses were conducted. All analyses were performed in R (v 3.4.3), using the "lme4" package (Bates et al., 2015). The normality of the residuals was tested visually by inspecting the QQ-plots. Hormone variables were log-transformed to facilitate normality of the residuals. No crude violations of the normality of the residuals were detected. The first step of the analyses investigated the longitudinal linear mixed model change in psychological well-being, self-reported physical health and the hormone ratios. In a subsequent step, two separate moderation analyses were conducted to investigate whether the hormones and their interactions or the androgen receptor gene polymorphism influence the change in psychological well-being or physical health over time. It is important to note that for all analyses with dynamic parameters (except for the androgen receptor), we used longitudinal data instead of only baseline values, in order to incorporate the dynamic nature into the analyses. Moderation analyses with both hormone ratios were conducted in individual models for each ratio, with subsequent individual interaction analyses of the single hormones. Models for the hormonal interactions included both interaction terms for T and E2, or T and C, respectively, in the same model. The two-step analysis was conducted to enable a broader analysis of steroid hormone interactions, followed by a detailed, statistical interaction term analysis to determine which of the hormones exhibits a stronger impact on the effect. Taken together there were seven separate models. Two individual models for the hormone ratios*time, three separate models for all single hormones as predictors and one model for the T*time and C*time and one for the T*time and E2*time interaction terms. All analyses were controlled for physical activity, educational level, income, medication intake, and fat mass. To control for differing trajectories between different age groups, age was also used as a covariate. Furthermore, for all analyses with the hormones and their interactions, the androgen receptor gene polymorphism was added as a covariate. For the androgen receptor gene polymorphism moderations (i.e., CAG repeat length), (longitudinal) *T* values were added as a covariate to account for the potential interdependence of the androgen receptor and T.

RESULTS

Sample Characteristics

At baseline, 271 participants were included in the study. However, only 130 participants completed the screening for the follow-up and 97 participants completed the whole study at the second time point. The participants were invited at the T1 for a cross-sectional study and were then re-invited for a follow-up approximately 4 years after initial testing. Due to the initial cross-sectional design of the study, most participants who refused participation for the second wave did so because they did not expect a follow-up. After completion of the screening, participants who dropped out reported that they did not have sufficient time for the study or they were no longer eligible (i.e., no longer reported at least good self-rated health). Detailed information on the dropout rate is described elsewhere (Lacker et al., 2019).

The sample characteristics are described in **Table 1**. Participants were on average 61.26 (*SD* = 10.02) years old at baseline and 65.18 (*SD* = 9.98) years old at follow-up. The BMI was approximately 25 at both time points, with only minor changes. The majority of participants were in a relationship or married at both time points. Over 40% of participants held a university degree, while 24% of participants reported having completed secondary school as their highest educational attainment. Most participants earned between 50,000 and 1,50,000 CHF (50,000–1,50,000 USD) annually. At both time points, over 80% of participants reported being non-smokers. Around 60% of participants at baseline and 53% at follow-up did not take any regular medication. Of the participants taking medication, almost half took antihypertensive medication

TABLE 1 | Sample characteristics.

	Baseline			Follow-Up		
	N	**M/Freq.**	***SD*/%**	***N***	**M/Freq.**	***SD*/%**
Age	97	61.26	10.02	97	65.18	9.98
BMI	95	25.16	2.57	97	25.51	4.31
Fat mass in percent (%)	95	22.42	5.43	97	20.58	7.03
CAG repeats	–	-	-	94	17.8	3.07
C (pg/ml)	95	1,424.3	1,396.5	97	2,169.9	1,597.5
T (pg/ml)	93	42.53	18.44	96	37.88	26.28
E2 (pg/ml)	94	1.46	1.06	97	1.63	1.69
Physical health	97	65.47	3.93	96	65.36	4.09
Psychological well-being	97	25.58	2.87	96	25.67	2.87
Marital status (%)	97			96		
Single		8	8.25		5	5.21
In relationship		8	8.25		10	10.42
Married		70	72.16		70	72.92
Divorced		10	10.31		10	10.42
Widowed		1	1.03		1	1.04
Education (%)	97			96		
Middle school		8	8.25		9	9.38
Secondary school		23	23.7		23	23.96
Grammar school		7	7.22		7	7.29
University degree		43	44.33		46	47.92
other		16	16.49		11	11.46
Income (in CHF, %)	97			96		
No income		1	1.03		1	1.04
Up to 30,000		3	3.09		3	3.13
30,001–50,000		7	7.22		4	4.17
50,001–1,00,000		35	36.08		38	39.58
1,00,001–1,50,000		35	36.08		38	39.58
1,50,001–2,00,000		12	12.37		7	7.29
More than 2,00,000		4	4.12		5	5.21
Smoking (%)	97			96		
Non-smoking		82	84.54		83	86.46
Smoking		15	15.46		13	13.54
Medication (%)	97			96		
No		59	60.08		51	53.13
Yes (details see below)		38	39.18		45	46.88
Type of medication[a] (%)	38			45		
Antihypertensives		23	45.09		28	45.16
Vitamins and minerals		1	1.96		5	8.06
Other		27	52.94		29	46.77
Physical activity (%)	97			96		
Less than 1 h		1	1.03		1	1.04
1–3 h		32	32.99		20	20.83
4–6 h		40	41.24		51	53.13
7 or more		24	24.74		24	25.00

[a]*Participants could indicate more than one medication. Note. N, Number of participants; M, Mean; Freq., Frequency; SD, Standard deviation; %, Percent; C, Cortisol; T, Testosterone; E2, Estradiol; CHF, Swiss Francs.*

or vitamins and minerals, and a further 50% took "other" medication, which included mostly cholesterol and bowel-regulating medications. With regard to physical activity, baseline results indicated that 66% of participants were physically active for 4–6 h or more per week, while this proportion increased by 12%, to a total of 78%, at follow-up.

Mixed Model Analyses and Descriptive Data

The mean CAG repeat length was 17.8 (*SD* = 3.07) and approximately normally distributed (data not shown). Participants had a raw mean C level of 1,424.3 (*SD* = 1,396.5) pg/ml at baseline, which increased to 2169.9 (*SD* = 1,597.5) pg/ml at follow-up. For raw T, the results indicated a mean T level of 42.53 (*SD* = 18.44) pg/ml at baseline and a decreased T level of 37.88 (*SD* = 26.28) at follow-up. Raw E2, however, remained rather stable, with a slight increase from 1.46 (*SD* = 1.06) pg/ml at baseline to 1.63 (*SD* = 0.17) at follow-up. For the hormone ratios, the raw Testosterone/Cortisol ratio decreased from 0.06 (*SD* = 0.05) pg/ml to 0.03 (*SD* = 0.03) pg/ml, while the raw Testosterone/Estradiol ratio decreased from 43.35 (*SD* = 30.02) pg/ml to 38.52 (*SD* = 41.95) pg/ml. Mixed model analyses for the change in single hormones can be derived from Lacker et al. (2019). The self-reported physical health remained stable, with values of 65.47 (*SD* = 3.93) at baseline and 65.36 (*SD* = 4.09) at follow-up. Psychological well-being remained similarly stable, with a mean of 25.58 (*SD* = 2.87) at baseline and 25.67 (*SD* = 2.87) at follow-up. The mixed model analyses with random intercepts

TABLE 2 | Results of the linear mixed model analyses.

	Linear mixed model results over time						
	AIC_0	AIC	Estimate	*SE*	*p*	95% CI_{lower}	95% CI_{upper}
Physical health	1,042	1,059	0.143	0.389	0.367	−0.627	0.911
Psychological well-being	898	912	−0.058	0.262	0.825	−0.578	0.459
Linear mixed model results moderation analyses for physical health							
TCr*time	1,042	993	0.556	0.496	0.265	−0.436	1.534
TE2r*time	1,042	996	−0.346	0.564	0.540	−1.470	0.765
T	1,042	992	1.343	0.400	<0.001	0.544	2.150
C	1,042	1,000	0.443	0.320	0.168	−0.194	1.093
T*time	1,042	986	−0.188	0.865	0.828	−1.960	1.530
C*time	1,042	986	0.001	0.591	0.999	−1.169	1.196
E2	1,042	1,011	0.034	0.356	0.923	−0.673	0.744
T*time	1,042	993	−1.049	0.961	0.277	−3.027	0.865
E2*time	1,042	993	0.741	0.595	0.215	−0.432	1.944
CAG repeat*time	1,042	1,010	−0.271	0.117	0.023	−0.503	−0.039
Linear mixed model results moderation analyses for psychological well-being							
TCr*time	898	863	−0.289	0.343	0.401	−0.974	0.387
TE2r*time	898	864	1.076	0.375	0.005	0.336	1.820
T	898	870	0.222	0.284	0.436	−0.365	0.826
C	898	870	−0.121	0.218	0.579	−0.554	0.320
T*time	898	866	−0.027	0.600	0.964	−1.238	1.157
C*time	898	866	0.533	0.411	0.198	−0.285	1.379
E2	898	876	0.414	0.246	0.094	−0.080	0.912
T*time	898	867	1.165	0.643	0.073	−0.124	2.434
E2*time	898	867	−1.006	0.395	0.013	−1.785	−0.217
CAG repeat*time	898	881	−0.081	0.079	0.312	−0.238	0.077

N, 97; T, Testosterone; C, Cortisol; E2, Estradiol; AIC0, Akaike criteron of the null model; AIC, Akaike criterion; SE, standard error; p, p-value; CI, Confidence interval.

(detailed information shown in **Table 2**) yielded no significant results for the change in physical health and psychological well-being, while both hormone ratios decreased significantly (TCr: $b = -0.756$, $SE = 0.134$, $p < 0.001$; TE2r: $b = -0.347$, $SE = 0.098$, $p < 0.001$).

The results of the moderation analyses are described in **Table 2**. Moderation analyses for the longitudinal change in physical health did not yield significant results for the interaction of T and E2, T and C and the hormone ratios. However, T was a significant predictor of physical health, meaning that it significantly influenced the intercept for physical health ($b = 1.34$, $SE = 0.40$, $p < 0.001$). Additionally, when using the CAG repeat length ($b = -0.27$, $SE = 0.12$, $p = 0.023$) as moderator (CAG repeat*time) a significant moderation effect on the change in physical health was found, showing an increase for lower CAG repeats. For the longitudinal change in psychological well-being, on the other hand, the results indicated a significant moderation effect of E2 ($b = -1.01$, $SE = 0.40$, $p = 0.013$), but only a trend for T ($b = 1.17$, $SE = 0.64$, $p = 0.073$), on the changes over time, whereas the CAG repeat and the interaction of T and C were not significant moderators of the change in psychological well-being. Furthermore, there was a significant moderation effect of the TE2r on psychological well-being ($b = 1.076$, $SE = 0.374$, $p = 0.005$).

Figure 1 graphically depicts the three significant moderation effects; however, it is important to note that the grouping was conducted for reasons of simplicity, and all moderation analyses used continuous variables as moderators.

Additionally, the supplementary material shows a correlation matrix for all variables that were used in the study. **Supplementary Tables S1, S2** show bootstrapped zero-order two-tailed correlations and two-tailed partial correlations, controlling for age, education, income, fat mass, and physical activity.

DISCUSSION

In this longitudinal study on healthy aging men, we found that both hormone ratios decreased over time, while self-reported physical health and psychological well-being remained stable. Moreover, E2 moderated the course of psychological well-being, insofar as the change in psychological well-being differs between individuals in relation to the (longitudinal) E2 values, if *T* values are taken into account. None of the other assessed hormones moderated psychological well-being or physical health. T was, however, found to be a significant predictor of physical health. Additionally, the androgen receptor gene polymorphism had an impact on the course of physical health: the trajectory of physical health changes with the number of CAG repeats.

In general, the literature on age-related hormone alterations in men reports a decrease in sex hormones, with the exception of E2, and an increase in glucocorticoid levels is also well documented (Fiacco et al., 2018). Our findings support this, showing a decrease in T, an increase in C, and stable E2 levels (Lacker et al., 2019).

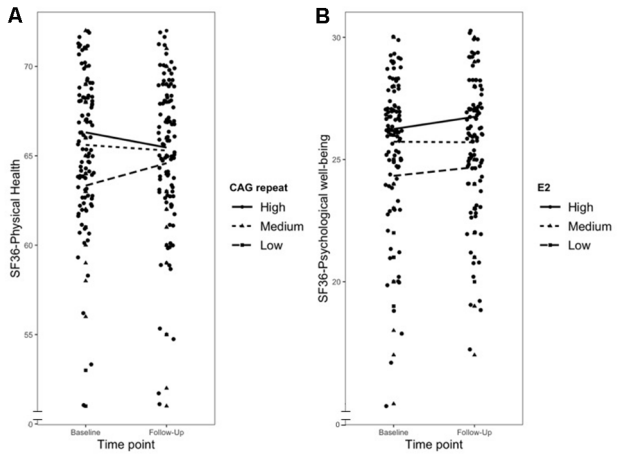


FIGURE 1 | Graph showing moderation effects for all significant interactions. Panel **(A)** showing the influence of the androgen receptor polymorphism cytosine-adenine-guanine (CAG repeat) on the course of physical health, Panel **(B)** showing the influence of E2 on the course of psychological well-being. High is 1SD above the sample mean, Medium the sample mean and Low is 1SD below the sample mean.

The finding that physical health and psychological well-being remained stable in our sample is highly interesting considering that in the general population, a continuous age-related decline in both health dimensions has been reported (Murphy et al., 2017). A multi-national study conducted over a similar time span in a large general population sample revealed a stronger decrease in physical health compared to our sample, but an equally low decrease in psychological well-being (Bourassa et al., 2017). The findings from another study suggest that psychological well-being, and especially dimensions such as life satisfaction or negative affect, might decline in advanced age only (Hansen and Slagsvold, 2012). However, with the present sample of middle-aged and older men, we cannot draw definitive conclusions in this regard.

E2 seems to positively influence the course of psychological well-being in specifically healthy aging men, as shown in the present analyses. Interestingly, while a recent meta-analysis showed increased E2 levels for depressed men (Fischer et al., 2019a), previous research suggested that E2 alone did not have an impact on markers of well-being such as mood (Beer et al., 2006). The latter finding was supported in our result, where the interaction of E2 was only significant after adding the interaction with T. Also in line with the latter finding, the TE2r was found to be associated with depressive symptoms in a very specific population of obese men (Monteagudo et al., 2016). The high-fat mass in these men presumably led to high aromatase activity and consequently to a parallel increase in E2 and a decrease in T (Vermeulen, 2000; Samaras et al., 2012). Concluding from these findings in obese men, we argue that it is generally not higher E2 alone that is responsible for reduced psychological well-being in aging men, but rather the interaction with reduced T levels. This ultimately results in an imbalance in the endpoints of the HPG axis, with effects on psychological well-being, where E2 is the main reason for this effect. This aspect also enables the integration with the recent meta-analysis by Fischer et al. (2019a).

The balance between the hormonal endpoints of the HPG and HPA axes may have an influence on psychological well-being. Previous research suggested functional connectivity between these two axes, through the existence of sex and adrenal steroid receptors in brain regions such as the hippocampus, amygdala or nuclei of the hypothalamus (Viau, 2002). In accordance with this, meta-analyses found T and C to be associated with mental health, which is closely related to psychological well-being (Adam et al., 2017; Walther et al., 2019). This implies that the balance between the two can also play a role in mental health maintenance. At first glance, it is thus surprising that we were unable to replicate this effect in the current sample. However, the present finding is in line with a previous report from our group in the same sample, in which we showed that psychosocial markers and biological markers change independently of each other (Lacker et al., 2019). Therefore, the lack of a relationship between the interaction of T and C and psychological well-being might be a phenomenon of healthy aging.

While the interplay between the endpoints of the HPG and HPA axes does not seem to explain changes in psychological well-being, T alone seems to be a predictor of physical health. This finding is in line with the literature. For example, a previous study with 407 participants revealed a low TCr to be an indicator of poor physical health, where T, but not C, differed significantly between a healthy and a non-healthy group (Wang et al., 2013). A recent review also supports the impact of T on physical health (Fiacco et al., 2018). In general, T shows anabolic effects, while C exhibits catabolic effects, with the former improving aspects of physical health (e.g., bone density) and the latter decreasing physical health (Dennison et al., 1999; Snyder et al., 2017). This might explain why T, but not C, was a significant predictor of physical health in our healthy aging sample. Similarly to psychological well-being, the lack of influence of the T and C interaction on physical health might reflect a healthy aging phenomenon, due to increased physical health in the investigated sample and thus adequate T and C values.

The interaction between the endpoints of the HPG axis does not seem to influence physical health in healthy aging men, as the T and E2 interactions did not have an impact on physical health in our sample. A previous general population study, however, concluded that the TE2r can be important for longevity (Menke et al., 2010). More specifically, low E2, as well as low T secretion, were associated with a higher cardiovascular risk. The same was found in another study for all-cause mortality (Hsu et al., 2016). In our study, T was a predictor of physical health only when the interaction with E2 was not included in the model. However, participants with higher T exhibited higher physical health, which is in line with the previously mentioned studies.

The androgen receptor gene polymorphism was suggested to potentially alter the course of physical health. According to our results, the androgen receptor gene polymorphism does appear to influence the course of subjectively experienced physical health even after controlling for longitudinal T secretion. Surprisingly, however, previous findings emphasized the role of the androgen receptor gene polymorphism in physical health

in the patient samples, but not in a general population sample (Schneider et al., 2010). Furthermore, it has been shown that the androgen receptor gene polymorphism can modulate androgen action in the target cells (Eendebak et al., 2016), thus suggesting a potential mechanism to explain its influence on physical health. This is relevant, for example, in the patient samples with potentially altered endocrine secretion patterns. In the present study, we were able to replicate the association with subjective physical health, thus further strengthening the link between the androgen receptor gene polymorphism and physical health.

The mean length for the androgen receptor gene polymorphism in the present study was 17.8, and thus below the generally reported mean of approximately 21. This can be explained by the applied capillary electrophoresis method. Other studies using the same method also reported lower repeats of approximately three base-pairs for the androgen receptor gene polymorphism (Mansfield et al., 1996; Westberg et al., 2001; Boorman et al., 2002). Capillary electrophoresis can potentially include a systematic error. Due to an extensive pairing in the CAG region, the probability that this region will melt completely during the analyses is decreased (Mansfield et al., 1996; Boorman et al., 2002). This can lead to an underestimation of the true repeat length when comparing fragment peaks with standard curves (Boorman et al., 2002).

Some limitations of the present study need to be mentioned. The first refers to the high dropout rate, which was mostly due to the *post hoc* planning of the follow-up. Second, potential conclusions that are drawn from this study only refer to healthy aging men. The additional investigation of a non-healthy group could be beneficial. Third, the methodological issues of the genetic analyses hinder comparisons of the androgen receptor gene polymorphism length with other study findings, although the sample itself was very homogeneous. Fourth, the internal consistency for the physical health questionnaire scale was rather low in the current sample. This may point to the need to validate the questionnaire in a healthy aging population. Finally, there are also some statistical constraints within the study: we used a two-time-point design, which could only incorporate random intercepts, whereas more time points would enable detailed random slope analyses (Curran et al., 2010).

Besides this, our study has several distinct strengths. The longitudinal design enabled the detailed investigation of intraindividual changes over time. Moreover, we used validated questionnaires and protocols for data collection. Applying strict inclusion criteria, we specifically investigated healthy aging men. Furthermore, we applied sophisticated statistical methods, respecting interindividual differences while focusing on intraindividual changes.

To sum up, we found that E2 influenced the course of psychological well-being, while T was a predictor of physical health. Furthermore, the androgen receptor gene polymorphism was associated with the course of self-reported physical health but not the course of psychological well-being. Since self-reported perceived health and actual health are closely related (Dainese et al., 2011), these results provide new insights into the role of distinct endocrine and genetic parameters for the maintenance of health throughout the aging process in men. However, the interactions of biological and psychological aspects in healthy aging men remain poorly understood and require further study.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Canton of Zurich. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TL and UE designed the concept of the study. TL and AW organized and conducted the study and collected the data. TL wrote the first draft of the manuscript. UE, SF, and AW contributed with important intellectual content and edited subsequent versions of the manuscript.

FUNDING

The study was funded by the University Research Priority Program (URPP) "Dynamics of Healthy Aging" of the University of Zurich. The funders played no role in the design, execution or analysis of the current study.

ACKNOWLEDGMENTS

First, we would like to thank the Men's Health 40+ study team, namely Nina Chiastra, Rahel Furrer and Anja Vandersmissen, without whom the study would not have been possible. We also wish to thank Dr. Firouzeh Farahmand and Elena Gardini for their extraordinary work in performing all hormone and genetic analyses. We are further grateful to all of the participants for their time and motivation to join the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00043/full#supplementary-material.

REFERENCES

Adam, E. K., Quinn, M. E., Tavernier, R., McQuillan, M. T., Dahlke, K. A., and Gilbert, K. E. (2017). Diurnal cortisol slopes and mental and physical health outcomes: a systematic review and meta-analysis. *Psychoneuroendocrinology* 83, 25–41. doi: 10.1016/j.psyneuen.2017.05.018

Barzilai, N., Cuervo, A. N., and Austad, S. (2018). Aging as a biological target for prevention and therapy. *JAMA* 320, 1321–1322. doi: 10.1001/jama.2018.9562

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:2015. doi: 10.18637/jss.v067.i01

Beard, J. R., and Bloom, D. E. (2015). Towards a comprehensive public health response to population ageing. *Lancet* 385, 658–661. doi: 10.1016/s0140-6736(14)61461-6

Beard, J. R., Officer, A., De Carvalho, I. A., Sadana, R., Pot, A. M., Michel, J. P., et al. (2016). The world report on ageing and health: a policy framework for healthy ageing. *Lancet* 387, 2145–2154. doi: 10.1016/S0140-6736(15)00516-4

Beer, T. M., Bland, L. B., Bussiere, J. R., Neiss, M. B., Wersinger, E. M., Garzotto, M., et al. (2006). Testosterone loss and estradiol administration modify memory in men. *J. Urol.* 175, 130–135. doi: 10.1016/S0022-5347(05)00049-2

Boorman, D. W., Guo, Y., Visvanathan, K., Helzlsouer, K., and O'Brien, T. G. (2002). Automated fragment analysis method for determining androgen receptor CAG repeat length. *Biotechniques* 33, 140–143. doi: 10.2144/02331md01

Bourassa, K. J., Memel, M., Woolverton, C., and Sbarra, D. A. (2017). Social participation predicts cognitive functioning in aging adults over time: comparisons with physical health, depression and physical activity. *Aging Ment. Health* 21, 133–146. doi: 10.1080/13607863.2015.1081152

Bullinger, M., and Kirchberger, I. (1998). *SF-36: Fragebogen Zum Gesundheitszustand; Handanweisung.* Göttingen: Hogrefe, Verlag für Psychologie.

Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiol. Rev.* 9, 399–431. doi: 10.1152/physrev.1929.9.3.399

Chamberlain, N. L., Driver, E. D., and Miesfeld, R. L. (1994). The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res.* 22, 3181–3186. doi: 10.1093/nar/22.15.3181

Chiappin, S., Antonelli, G., Gatti, R., and Elio, F. (2007). Saliva specimen: a new laboratory tool for diagnostic and basic investigation. *Clin. Chim. Acta* 383, 30–40. doi: 10.1016/j.cca.2007.04.011

Curran, P. J., Obeidat, K., and Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* 11, 121–136. doi: 10.1080/15248371003699969

Dainese, S. M., Allemand, M., Ribeiro, N., Bayram, S., Martin, M., and Ehlert, U. (2011). Protective factors in midlife. *GeroPsych* 24, 19–29. doi: 10.1024/1662-9647/a000032

Demirev, P. A. (2013). Dried blood spots: analysis and applications. *Anal. Chem.* 85, 779–789. doi: 10.1021/ac303205m

Dennison, E., Hindmarsh, P., Fall, C., Kellingray, S., Barker, D., Phillips, D., et al. (1999). Profiles of endogenous circulating cortisol and bone mineral density in healthy elderly men. *J. Clin. Endocrinol. Metab.* 84, 3058–3063. doi: 10.1210/jcem.84.9.5964

Denson, T. F., Ronay, R., von Hippel, W., and Schira, M. M. (2013). Endogenous testosterone and cortisol modulate neural responses during induced anger control. *Soc. Neurosci.* 8, 165–177. doi: 10.1080/17470919.2012.655425

Diener, E., and Lucas, R. E. (1999). "Personality and subjective well-being," in *Well-Being: The Foundations of Hedonic Psychology*, eds E. Diener, D. Kahneman and N. Schwarz (New York, NY: Russell Sage Foundation), 213–229.

Eendebak, R. J. A. H., Huhtaniemi, I. T., Pye, S. R., Ahern, T., W O'Neill, T., Bartfai, G., et al. (2016). The androgen receptor gene CAG repeat in relation to 4-year changes in androgen-sensitive endpoints in community-dwelling older european men. *Eur. J. Endocrinol.* 175, 583–593. doi: 10.1530/eje-16-0447

Engel, G. L. (1980). The clinical application of the biopsychosocial model. *Am. J. Psychiatry* 137, 535–544. doi: 10.1176/ajp.137.5.535

Feldman, H. A., Longcope, C., Derby, C. A., Johannes, C. B., Araujo, A. B., Coviello, A. D., et al. (2002). Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the massachusetts male aging study. *J. Clin. Endocrinol. Metab.* 87, 589–598. doi: 10.1210/jcem.87.2.8201

Fiacco, S., Walther, A., and Ehlert, U. (2018). Steroid secretion in healthy aging. *Psychoneuroendocrinology* 105, 64–78. doi: 10.1016/j.psyneuen.2018.09.035

Fischer, S., Ehlert, U., and Amiel Castro, R. (2019a). Hormones of the hypothalamic-pituitary-gonadal (HPG) axis in male depressive disorders - a systematic review and meta-analysis. *Front. Neuroendocrinol.* 55:100792. doi: 10.1016/j.yfrne.2019.100792

Fischer, S., Obrist, R., and Ehlert, U. (2019b). How and when to use dried blood spots in psychoneuroendocrinological research. *Psychoneuroendocrinology* 108:190–196. doi: 10.1016/j.psyneuen.2019.06.011

Goncharov, N., Katsya, G., Dobracheva, A., Nizhnik, A., Kolesnikova, G., Herbst, V., et al. (2006). Diagnostic significance of free salivary testosterone measurement using a direct luminescence immunoassay in healthy men and in patients with disorders of androgenic status. *Aging Male* 9, 111–122. doi: 10.1080/13685530600713060

Gong, Y., Xiao, H., Li, C., Bai, J., Cheng, X., Jin, M., et al. (2013). Elevated T/E 2 ratio is associated with an increased risk of cerebrovascular disease in elderly men. *PLoS One* 8:e61598. doi: 10.1371/journal.pone.0061598

Gordon, E. H., and Hubbard, R. E. (2018). Physiological basis for sex differences in frailty. *Curr. Opin. Physiol.* 6, 10–15. doi: 10.1016/j.cophys.2018.02.013

Gupta, D., and Morley, J. E. (2014). Hypothalamic-pituitary-adrenal (HPA) axis and aging. *Compr. Physiol.* 4, 1495–1510. doi: 10.1002/cphy.c130049

Guthrie, B., and Boyd, C. M. (2018). Clinical guidelines in the context of aging and multimorbidity. *Public Policy Aging Rep.* 28, 143–149. doi: 10.1093/ppar/pry038

Han, K., Lee, Y., Gu, J., Oh, H., Han, J., and Kim, K. (2015). Psychosocial factors for influencing healthy aging in adults in Korea. *Health Qual. Life Outcomes* 13:31. doi: 10.1186/s12955-015-0225-5

Hansen, T., and Slagsvold, B. (2012). The age and subjective well-being paradox revisited: a multidimensional perspective. *Norsk Epidemiol.* 22, 187–195. doi: 10.5324/nje.v22i2.1565

Hsu, B., Cumming, R. G., Naganathan, V., Blyth, F. M., Le Couteur, D. G., Hirani, V., et al. (2016). Temporal changes in androgens and estrogens are associated with all-cause and cause-specific mortality in older men. *J. Clin. Endocrinol. Metab.* 101, 2201–2210. doi: 10.1210/jc.2016-1025

Huhtaniemi, I. T., Pye, S. R., Limer, K. L., Thomson, W., O'Neill, T. W., Platt, H., et al. (2009). Increased estrogen rather than decreased androgen action is associated with longer androgen receptor CAG repeats. *J. Clin. Endocrinol. Metab.* 94, 277–284. doi: 10.1210/jc.2008-0848

Huisman, M., Poppelaars, J., van der Horst, M., Beekman, A. T. F., Brug, J., van Tilburg, T. G., et al. (2011). Cohort profile: the longitudinal aging study amsterdam. *Int. J. Epidemiol.* 40, 868–876. doi: 10.1093/ije/dyQ239

Joseph, D. N., and Whirledge, S. (2017). Stress and the HPA axis: balancing homeostasis and fertility. *Int. J. Mol. Sci.* 18:E2224. doi: 10.3390/ijms18102224

Karlamangla, A. S., Friedman, E. M., Seeman, T. E., Stawksi, R. S., and Almeida, D. M. (2013). Daytime trajectories of cortisol: demographic and socioeconomic differences-findings from the national study of daily experiences. *Psychoneuroendocrinology* 38, 2585–2597. doi: 10.1016/j.psyneuen.2013.06.010

Krithivas, K., Yurgalevitch, S. M., Mohr, B. A., Wilcox, C. J., Batter, S. J., Brown, M., et al. (1999). Evidence that the CAG repeat in the androgen receptor gene is associated with the age-related decline in serum androgen levels in men. *J. Endocrinol.* 162, 137–142. doi: 10.1677/joe.0.1620137

Kunzmann, U., Little, T. D., and Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudinal evidence from the Berlin aging study. *Psychol. Aging* 15, 511–526. doi: 10.1037/0882-7974.15.3.511

Lacker, T. J., Walther, A., and Ehlert, U. (2019). Age-related alterations in endocrine markers do not match changes in psychosocial measures - findings from the men's health 40+ longitudinal study. *Psychoneuroendocrinology* 107, 17–18. doi: 10.1016/j.psyneuen.2019.07.048

Lowsky, D. J., Olshansky, S. J., Bhattacharya, J., and Goldman, D. P. (2013). Heterogeneity in healthy aging. *J. Gerontol. A Biol. Sci. Med. Sci.* 69, 640–649. doi: 10.1093/gerona/glt162

Mansfield, E. S., Vainer, M., Enad, S., Barker, D. L., Harris, D., Rappaport, E., et al. (1996). Sensitivity, reproducibility, and accuracy in short tandem repeat genotyping using capillary array electrophoresis. *Genome Res.* 6, 893–903. doi: 10.1101/gr.6.9.893

Menke, A., Guallar, E., Rohrmann, S., Nelson, W. G., Rifai, N., Kanarek, N., et al. (2010). Sex steroid hormone concentrations and risk of death in US men. *Am. J. Epidemiol.* 171, 583–592. doi: 10.1093/aje/kwp415

Monteagudo, P. T., Falcão, A. A., Verreschi, I. T. N., and Zanella, M. T. (2016). The imbalance of sex-hormones related to depressive symptoms in obese men. *Aging Male* 19, 20–26. doi: 10.3109/13685538.2015.1084500

Murphy, S. L., Xu, J., Kochanek, K. D., Curtin, S. C., and Arias, E. (2017). Deaths: final data for 2015. *Natl. Vital Stat. Rep.* 66, 1-75. Available online at: https://stacks.cdc.gov/view/cdc/50011.

Nelson, K. A., and Witte, J. S. (2002). Androgen receptor CAG repeats and prostate cancer. *Am. J. Epidemiol.* 155, 883–890. doi: 10.1093/aje/155.10.883

Rajender, S., Pandu, G., Sharma, J. D., Gandhi, K. P. C., Singh, L., and Thangaraj, K. (2008). Reduced CAG repeats length in androgen receptor gene is associated with violent criminal behavior. *Int. J. Legal Med.* 122, 367–372. doi: 10.1007/s00414-008-0225-7

Rector, B. J. L., and Friedman, E. M. (2018). "Hormones and well-being," in *Handbook of Well-Being*, eds E. Diener, S. Oishi and L. Tay (Salt Lake City, UT: DEF Publishers), 1–16.

Rockwood, K. (2019). Frailty and aging medicine. *Aging Med.* 2, 4–6. doi: 10.1002/agm2.12060

Samaras, N., Frangos, E., Forster, A., Lang, P. O., and Samaras, D. (2012). Andropause: a review of the definition and treatment. *Eur. Geriatr Med.* 3, 368–373. doi: 10.1016/j.eurger.2012.08.007

Schneider, G., Nienhaus, K., Gromoll, J., Heuft, G., Nieschlag, E., and Zitzmann, M. (2010). Aging males' symptoms in relation to the genetically determined androgen receptor CAG polymorphism, sex hormone levels and sample membership. *Psychoneuroendocrinology* 35, 578–587. doi: 10.1016/j.psyneuen.2009.09.008

Smith, G. D., Ben-Shlomo, Y., Beswick, A., Yarnell, J., Lightman, S., and Elwood, P. (2005). Cortisol, testosterone and coronary heart disease: prospective evidence from the caerphilly study. *Circulation* 112, 332–340. doi: 10.1161/CIRCULATIONAHA.104.489088

Snyder, P. J., Kopperdahl, D. L., Stephens-Shields, A. J., Ellenberg, S. S., Cauley, J. A., Ensrud, K. E., et al. (2017). Effect of testosterone treatment on volumetric bone density and strength in older men with low testosterone: a controlled clinical trial. *JAMA Intern. Med.* 177, 471–479. doi: 10.1001/jamainternmed.2016.9539

Sollberger, S., and Ehlert, U. (2016). How to use and interpret hormone ratios. *Psychoneuroendocrinology* 63, 385-397. doi: 10.1016/j.psyneuen.2015.09.031

Strickhouser, J. E., Zell, E., and Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychol.* 36, 797-810. doi: 10.1037/hea0000475

Terburg, D., Morgan, B., and van Honk, J. (2009). The testosterone-cortisol ratio: a hormonal marker for proneness to social aggression. *Int. J. Law Psychiatry* 32, 216–223. doi: 10.1016/j.ijlp.2009.04.008

Tirabassi, G., Cignarelli, A., Perrini, S., delli Muti, N., Furlani, G., Gallo, M., et al. (2015). Influence of CAG repeat polymorphism on the targets of testosterone action. *Int. J. Endocrinol.* 2015:298107. doi: 10.1155/2015/298107

Vermeulen, A. (2000). Andropause. *Maturitas* 34, 5–15. doi: 10.1016/s0378-5122(99)00075-4

Vermeulen, A., Kaufman, J. M., Goemaere, S., and Van Pottelberg, I. (2002). Estradiol in elderly men. *Aging Male* 5, 98-102. doi: 10.1080/tam.5.2.98.102

Viau, V. (2002). Functional cross-talk between the hypothalamic-pituitary-gonadal and -Adrenal axes. *J. Neuroendocrinol.* 14, 506–513. doi: 10.1046/j.1365-2826.2002.00798.x

Wade, D. T., and Halligan, P. W. (2017). The biopsychosocial model of illness: a model whose time has come. *Clin. Rehabil.* 31, 995–1004. doi: 10.1177/0269215517709890

Wallerand, H., Rémy-Martin, A., Chabannes, E., Bermont, L., Adessi, G. L., and Bittard, H. (2001). Relationship between expansion of the CAG repeat in exon 1 of the androgen receptor gene and idiopathic male infertility. *Fertil. Steril.* 76, 769–774. doi: 10.1016/s0015-0282(01)01987-2

Walther, A., and Ehlert, U. (2015). Steroid secretion and psychological well-being in men 40+. *Neurobiology of Men's Mental Health*. Available online at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS. . .PAGE=reference. . .D=psyc12 . . .NEWS=N. . .AN=2015-46819-024. Accessed September 17, 2019.

Walther, A., Breidenstein, J., and Miller, R. (2019). Association of testosterone treatment with alleviation of depressive symptoms in men: a systematic review and meta-analysis. *JAMA Psychiatry* 76, 31–40. doi: 10.1001/jamapsychiatry.2018.2734

Walther, A., Philipp, M., Lozza, N., and Ehlert, U. (2016). The rate of change in declining steroid hormones: a new parameter of healthy aging in men? *Oncotarget* 7, 60844-60857. doi: 10.18632/oncotarget.11752

Wang, L., Zhao, X., Chen, J., Guo, X., Liang, X., Yi, D., et al. (2013). Biological indicators of sub-optimal health status. *J. Tradit. Chin. Med.* 33, 647–650. doi: 10.1016/s0254-6272(14)60036-4

Weiss, A., Bates, T. C., and Luciano, M. (2008). Happiness is a Personal(ity) thing: the genetics of personality and well-being in a representative sample. *Psychol. Sci.* 19, 205–210. doi: 10.1111/j.1467-9280.2008.02068.x

Westberg, L., Baghaei, F., Rosmond, R., Hellstrand, M., Landén, M., Jansson, M., et al. (2001). Polymorphisms of the androgen receptor gene and the estrogen receptor β gene are associated with androgen levels in women. *J. Clin. Endocrinol. Metab.* 86, 2562–2568. doi: 10.1210/jcem.86.6.7614

Westberg, L., Henningsson, S., Landén, M., Annerbrink, K., Melke, J., Nilson, S., et al. (2009). Influence of androgen receptor repeat polymorphisms on personality traits in men. *J. Psychiatry Neurosci.* 34, 205–213.

World Health Organization. (2015a). *World Health Statistics 2015*. Luxemburg: World Health Organization.

World Health Organization. (2015b). *World Report on Ageing and Health*. Luxemburg: World Health Organization.

World Health Organization. (2017). 10 priorities towards a decade of healthy ageing, 1–20. Available online at: https://www.who.int/ageing/WHO-ALC-10-priorities.pdf. Accessed September 17, 2019.

Zhou, J. N., and Swaab, D. F. (1999). Activation and degeneration during aging: a morphometric study of the human hypothalamus. *Microsc. Res. Tech.* 44, 36–48. doi: 10.1002/(SICI)1097-0029(19990101)44:1<36::AID-JEMT5>3.0.CO;2-F

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 26 June 2020
doi: 10.3389/fnhum.2020.00238

Eavesdropping on Autobiographical Memory: A Naturalistic Observation Study of Older Adults' Memory Sharing in Daily Conversations

Aubrey A. Wank[1], Matthias R. Mehl[2], Jessica R. Andrews-Hanna[3,4,5], Angelina J. Polsinelli[6], Suzanne Moseley[7], Elizabeth L. Glisky[8] and Matthew D. Grilli[1,4,9]**

[1]Human Memory Laboratory, Department of Psychology, University of Arizona, Tucson, AZ, United States, [2]Naturalistic Observation of Social Interaction Laboratory, Department of Psychology, University of Arizona, Tucson, AZ, United States, [3]Neuroscience of Emotion and Thought Laboratory, Department of Psychology, University of Arizona, Tucson, AZ, United States, [4]Evelyn F. McKnight Brain Institute, University of Arizona, Tucson, AZ, United States, [5]Cognitive Science Program, Department of Philosophy, University of Arizona, Tucson, AZ, United States, [6]Department of Neurology, Indiana University School of Medicine, Indianapolis, IN, United States, [7]Minnesota Epilepsy Group, St. Paul, MN, United States, [8]Aging and Cognition Laboratory, Department of Psychology, University of Arizona, Tucson, AZ, United States, [9]Department of Neurology, University of Arizona, Tucson, AZ, United States

OPEN ACCESS

Edited by:
Mike Martin,
University of Zurich, Switzerland

Reviewed by:
Carlos J. Gomez-Ariza,
University of Jaén, Spain
Julia Mayas,
National University of Distance Education (UNED), Spain

***Correspondence:**
Aubrey A. Wank
aalw@email.arizona.edu
Matthew D. Grilli
mdgrilli@email.arizona.edu

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

Received: 03 March 2020
Accepted: 29 May 2020
Published: 26 June 2020

Citation:
Wank AA, Mehl MR, Andrews-Hanna JR, Polsinelli AJ, Moseley S, Glisky EL and Grilli MD (2020) Eavesdropping on Autobiographical Memory: A Naturalistic Observation Study of Older Adults' Memory Sharing in Daily Conversations. Front. Hum. Neurosci. 14:238. doi: 10.3389/fnhum.2020.00238

The retrieval of autobiographical memories is an integral part of everyday social interactions. Prior laboratory research has revealed that older age is associated with a reduction in the retrieval of autobiographical episodic memories, and the ability to elaborate these memories with episodic details. However, how age-related reductions in episodic specificity unfold in everyday social contexts remains largely unknown. Also, constraints of the laboratory-based approach have limited our understanding of how autobiographical semantic memory is linked to older age. To address these gaps in knowledge, we used a smartphone application known as the Electronically Activated Recorder, or "EAR," to unobtrusively capture real-world conversations over 4 days. In a sample of 102 cognitively normal older adults, we extracted instances where memories and future thoughts were shared by the participants, and we scored the shared episodic memories and future thoughts for their make-up of episodic and semantic detail. We found that older age was associated with a reduction in real-world sharing of autobiographical episodic and semantic memories. We also found that older age was linked to less episodically and semantically detailed descriptions of autobiographical episodic memories. Frequency and level of detail of shared future thoughts yielded weaker relationships with age, which may be related to the low frequency of future thoughts in general. Similar to laboratory research, there was no correlation between autobiographical episodic detail sharing and a standard episodic memory test. However, in contrast to laboratory studies, episodic detail production while sharing autobiographical episodic memories was weakly related to episodic detail

production while describing future events, unrelated to working memory, and not different between men and women. Overall, our findings provide novel evidence of how older age relates to episodic specificity when autobiographical memories are assessed unobtrusively and objectively "in the wild."

Keywords: episodic specificity, autobiographical memory, episodic memory, semantic memory, cognitive aging, naturalistic observation

INTRODUCTION

Remembering experiences from our personal history, or autobiographical memories, is thought to aid in the development and maintenance of the self, facilitate social communication, and guide behavior (Pillemer, 1992; Bluck and Alea, 2002; Pasupathi et al., 2002; Bluck, 2003; Bluck et al., 2005). Importantly, the act of recalling and sharing autobiographical memories allows us to communicate a wealth of information with others, from general knowledge to details of a one-time event (Conway and Pleydell-Pearce, 2000), with each form serving distinct purposes (Waters et al., 2014). Such disclosures of personal information in conversation are thought to increase intimacy and empathy within relationships, provide opportunities to teach others, and aid decision making (Alea and Bluck, 2003, 2007). In this sense, autobiographical memory sharing helps us navigate our social interactions.

In prior research, much attention has been given to understanding the relationship between older age and the way autobiographical memories are shared. One reliable finding is that normal cognitive aging is linked to a reduction in the *episodic specificity*, or the frequency of episodic retrieval and/or the vivid elaboration, of autobiographical memories. For instance, prior research has found that, when asked to reflect on the past, older adults retrieve fewer unique life events in comparison to young adults, and in counterbalance recall more events that are extended over time or are repeated events (Piolino et al., 2006; Ros et al., 2009, 2017; Ford et al., 2014). When asked to elaborate these episodic memories, older adults also tend to incorporate fewer episodic details relative to young adults, replacing them with semantic details, or factual knowledge about one's life story and the broader context of the event (Levine et al., 2002; Addis et al., 2008; Gaesser et al., 2011; Devitt et al., 2017). Critically, some work has further shown that reductions in episodic specificity are evident when comparing young-old to old-old adults (e.g., ages 65–75 vs. +75 or ages 60–69 vs. 70–79), indicating that this effect continues in advanced normal cognitive aging (Piolino et al., 2002; De Beni et al., 2013). The age-related reduction in episodic specificity, broadly reflecting a move away from episodic retrieval, has been interpreted in a few ways. One explanation is that older age is associated with a compensatory shift towards conceptual retrieval that is due to reduced executive resources (Piolino et al., 2010) and stronger coupling of two large-scale brain networks (Turner and Spreng, 2015; Spreng et al., 2018), resulting in a tendency to view things in a semantic light. Alternatively, a natural alteration in narrative style, perhaps related to new perspectives on a lifetime of experiences or changes in language use, could contribute to reduced episodic specificity with age (James et al., 1998; Trunk and Abrams, 2009; Gaesser et al., 2011). Both of these explanations can account for why prior work has commonly found that reduced episodic retrieval is accompanied by increased semantic retrieval. Regardless of the reason(s), these findings and theories suggest that in everyday life, increased age may be linked to less sharing of one's episodic past.

Despite the vast evidence for age-related reductions in episodic specificity and its potential importance to how memories are shared in everyday social life, this quality of memory retrieval has been largely studied in an analog of the real-world, namely through laboratory interviews (Williams and Broadbent, 1986; Levine et al., 2002; St. Jacques and Levine, 2007; Addis et al., 2008; Barnabe et al., 2012). In one approach, participants generate memories in response to positive and negative cue words, and they are subsequently scored as "episodic" or as more general memories of the past (Williams and Broadbent, 1986). Other laboratory interviews instruct individuals to retrieve episodic memories from a particular time or period and to focus on describing episodic details (Levine et al., 2002). Often, participants are provided cues, such as neutral words (e.g., "tree"), and once a natural ending point is reached, the experimenter may probe for additional information. These reports are then parsed into individual details, with each one scored according to whether it is episodic or "internal" to an event or more generic or "external," providing a fine-grained, objective picture of recollection (Levine et al., 2002).

Although, there are practical and psychometric benefits to this structured approach, the laboratory interview departs from the contexts encountered naturally in everyday life in several ways, including how autobiographical memories are likely cued and with whom they are commonly shared. Diary studies and other naturalistic thought sampling methods, which require individuals to record events or report on them, have revealed commonalities between laboratory-derived and real-world memories (Levine et al., 2002; Berntsen and Hall, 2004; Schlagman and Kvavilashvili, 2008; D'Argembeau et al., 2011). Yet, by their nature, these methods do not capture in-the-moment, naturalistic outward sharing of autobiographical memories with others, and these studies have not focused on assessing episodic specificity at the level of detail commonly done in the laboratory. Therefore, it remains unclear how age-related alterations in episodic specificity unfold in real-world social contexts.

One way to address this gap in knowledge is to utilize ambulatory assessment technologies that allow for the unobtrusive recording of everyday conversation as it happens (Mehl, 2017). The Electronically Activated Recorder or EAR

(Mehl et al., 2001; Mehl, 2017) is one method for such an approach (see Robbins, 2017). The EAR is a mobile smartphone application that periodically samples blocks of ambient sounds, including conversation, from one's moment-to-moment environment, providing what amounts to an acoustic log of one's day. Recently, the EAR has been applied to assess autobiographical sharing in natural, everyday social life, revealing outcomes that replicate and go beyond findings from the laboratory. For instance, prior work using the EAR has shown that in everyday social interactions, older adults tend to be past-oriented (Demiray et al., 2018) and share fewer future thoughts than young adults (Brianza and Demiray, 2019). Older adults also appear to share autobiographical memories in their natural social relationships for specific motives (Demiray et al., 2019). The EAR, therefore, presents a unique opportunity to ask whether, in everyday conversation, older age is associated with a reduction in sharing episodic memories and/or episodic detail.

New technologies that unobtrusively capture many memories over time can also be used to further investigate the influence of older age on autobiographical semantic memory. The relationship between older age and autobiographical semantic memory has been difficult to fully appreciate based on laboratory studies for a few reasons, chiefly that participants are typically required to recall a certain number of memories, and the amount of time given to memory sharing is controlled. As such, it remains unclear if older adults, under less constrained conditions, would retrieve autobiographical semantic memories or details at the rates commonly reported in laboratory-based studies. Concerning the real-world, according to both executive coupling (Turner and Spreng, 2015; Spreng et al., 2018) and narrative style accounts (James et al., 1998; Trunk and Abrams, 2009; Gaesser et al., 2011), older age may not alter the tendency to share autobiographical memories but rather reduced episodic retrieval may be counterbalanced by spared or increased sharing of semantic memories and semantic details (Devitt et al., 2017). However, to counter this position, prior research has revealed that broad structural and functional decline in the default network of the brain seems to emerge in older age and continue to unfold with advancing decades of life (Andrews-Hanna et al., 2007, 2019; Fjell et al., 2009), including in regions implicated in semantic memories and social concepts (Lambon Ralph et al., 2012; Andrews-Hanna et al., 2014). According to the integrity hypothesis of default network functioning (Andrews-Hanna et al., 2014, 2019), these findings suggest that while older adults may turn to autobiographical semantic memories and details if retrieval is evoked externally, older age may be connected to a global reduction in the retrieval of autobiographical memories—both episodic and semantic.

To address these questions about the relationship between older age and autobiographical memory sharing in natural social contexts, we conducted a secondary data analysis from a sample of 102 cognitively healthy older adults who, as part of a study on everyday cognition, completed laboratory-based cognitive testing, and for 4 days, wore the EAR (Polsinelli et al., 2020). In the present sample, The EAR recorded for 30 s every 6–18 min except during a 6-h overnight period starting 30 min after participants' bedtime (M = 10:54 pm; range = 9:00 pm–12:30 am). From the EAR recordings, we extracted autobiographical memories and future thoughts being shared by the participants in social situations (Demiray et al., 2018) to understand the sharing frequency of different forms of personal knowledge (i.e., episodic vs. semantic, past and future). We then scored for detailed composition within each instance of autobiographical episodic sharing using the well-established Autobiographical Interview (AI) protocol (Levine et al., 2002).

Based on research previously conducted on autobiographical memory retrieval in the laboratory, we hypothesized that older age would be associated with reduced episodic retrieval in everyday conversation in two ways. First, we expected older age to be linked to a reduction in autobiographical episodic memory sharing. Second, we predicted that older age would be associated with lower generation of episodic details while describing autobiographical episodic memories. Regarding autobiographical semantic memories, according to the executive coupling (Turner and Spreng, 2015; Spreng et al., 2018) and narrative style accounts (James et al., 1998; Trunk and Abrams, 2009; Gaesser et al., 2011) of autobiographical memory and normal cognitive aging, increased age among older adults may be positively associated with the generation of semantic memories and semantic details while describing episodic memories. However, according to the integrity hypothesis of default network functioning (Andrews-Hanna et al., 2014, 2019) increased age among older adults may be associated with a general reduction in autobiographical memory retrieval in social contexts, including semantic memories, which may be revealed through both frequency and semantic elaborateness (i.e., detail generation). Finally, in light of evidence that age-related cognitive differences in autobiographical memory are also reflected in future thinking (Addis et al., 2008; Madore et al., 2014; De Brigard et al., 2016), we expected that findings in the memory domain (e.g., older age linked to reduced episodic memory retrieval) would also be reflected in future thought sharing (e.g., older age associated with less sharing of future episodic thoughts).

To complement these age-related analyses and provide a clearer picture of the degree to which real-world memory sharing mirrors that of laboratory-based findings in older adults, we also examined a few additional features. First, we investigated whether, as shown in laboratory research, episodic specificity of one's memories is related to how detailed individuals are when they describe future episodic thoughts (Addis et al., 2008; Hill and Emery, 2013). Second, we examined whether two laboratory-based cognitive testing findings are found while sharing memories in the real-world, namely whether greater episodic specificity is associated with better working memory (Addis et al., 2008; Ros et al., 2009, 2017; Piolino et al., 2010), but weakly related to impersonal, laboratory-based measures of episodic learning and memory (Palombo et al., 2015; Grilli et al., 2018a). Third, similar to laboratory findings, we assessed whether episodic specificity is higher in women relative to men when memory sharing is assessed in real-world social contexts (Davis, 1999; MacDonald et al., 2000; Niedźwieńska, 2003; Pillemer et al.,

2003; Andreano and Cahill, 2009; Wang et al., 2011; Fuentes and Desrocher, 2013; Grysman and Hudson, 2013; Wang, 2013).

MATERIALS AND METHODS

Participants

The present study utilized data from a previously conducted study on cognitive aging using the EAR technology (Polsinelli et al., 2020). Participants (46 males/56 females) were between the ages of 65 and 90 (M = 76.12, SD = 6.00) and had, on average, 16.60 years of education (SD = 2.32, range = 12–22)[1]. All participants scored ≥25 on the Mini-Mental Status Examination, which indicated that their cognition was within normal limits[2]. Participants provided written informed consent following the Institutional Review Board of the University of Arizona.

Materials and Procedure

As mentioned, the present study was a secondary data analysis of the study conducted by Polsinelli et al. (2020). Polsinelli and colleagues adhered to the standard EAR protocol (Mehl, 2017), under which the participants are trained on how to use the EAR device before the study period (e.g., how to charge the EAR at night, how to wear the EAR to maximize recording quality). Participants wear the EAR while going about their days, unaware when exactly the device is recording. Through its sampling, the EAR protects privacy (i.e., takes snippets out of their larger conversational context) and enables at-scale empirical studies. Wearing the EAR is minimally bothersome and it has been successfully used, with good acceptance and compliance, in age groups ranging from childhood to old age (3 years to 93 years) and with both healthy and clinical populations. In the present study, compliance with the EAR procedures was high, with only 10 participants experiencing any notable technical difficulties at any point in the study period (i.e., failing to recharge overnight properly). When participants returned their EAR device, they completed a standard EAR evaluation measure (Mehl and Holleran, 2007). On average, participants' reported low obtrusiveness for themselves (e.g., "To what extent did the EAR impede your daily activities?"; M = 1.87, SD = 0.64; 1 = "not at all" through 5 = "a great deal") and bystanders (e.g., "To what extent did the EAR influence the behavior of people around you?," M = 1.94, SD = 0.83). Participants wore the EAR for approximately 4 days (a weekend and two weekdays) and gathered, on average, 310 30-s sound file samples (SD = 62, range = 91–405). The total number of sound files included only those during which participants were awake and deemed (i.e., coded as) wearing the EAR. The recorded sound files were transcribed as part of the original data analysis plan. For our secondary data analysis, we first determined whether each sound file contained speech by the participant and then whether the files captured any form of autobiographical memories or future thoughts being verbally shared with another person. We then separated this set into four categories, namely autobiographical episodic memories, semantic memories, future episodic thoughts, and future semantic thoughts. Consistent with prior descriptions, we scored autobiographical memories and future thoughts as episodic if they pertained to a specific event occurring at a particular time and place. Memories and future thoughts that lacked such specificity were scored as semantic.

We used the AI scoring protocol (Levine et al., 2002) to analyze the content of a subset of these autobiographical sound files. We selected the AI protocol because of its well-validated status as a measure of episodic specificity (Levine et al., 2002; St. Jacques and Levine, 2007; Devitt et al., 2017; Grilli et al., 2018a), its adaptability for scoring future event episodic specificity (Addis et al., 2008; Madore et al., 2014), and neural evidence linking episodic and semantic details to distinct regions/pathways of the default network (Hodgetts et al., 2017; Palombo et al., 2018; Memel et al., 2020). Consistent with the standard AI scoring, all sound files we tagged as containing episodic memories and future episodic thoughts were segmented into individual details and each detail was scored as episodic or semantic. Details were scored as episodic if they described the event content or sequencing, timing, location, perceptual quality, or thoughts or emotions of an event. Details that described facts about the self or world knowledge and details that were personal, non-specific events (e.g., extended or repeated events) were scored as semantic. Each detail was further scored as past or future-oriented based on the temporal nature of the event to which the detail was attached. The original AI protocol includes three additional detail types, namely repetitions, meta-comments, and descriptions of events external to the main event being described in the autobiographical interview. In our application of the AI protocol to EAR sound files, we did not attempt to score these categories. For repetitions and meta-comments, in the absence of insight into the social context, it was often difficult to reliably determine whether someone was absentmindedly thinking aloud, repeating oneself, or sharing information with someone who may not have heard or been present for the initial sharing of that detail. Notably, we did not repeatedly score the same information (e.g., we did not "double score" the re-sharing of the same episodic detail), but we also did not count these details as repetitions. We did not use the external detail category, because we did not assume that a single sound file must capture no more than one episodic memory. Relatedly, although future episodic thoughts might be scored as external details in the laboratory-based application of the AI, we simply scored them as episodic details if attached to episodic memory. To be consistent, we also included future-oriented semantic details attached to episodic memories as semantic details. We want to emphasize that while we slightly departed from the original AI scoring protocol, studies of healthy populations, both young and older, commonly find that semantic details capture the vast majority of "external" contents (Levine et al., 2002; Murphy et al., 2008; Bastin et al., 2013). **Figure 1** shows examples of scored sound files.

Two standard cognitive tests that were administered to a subgroup of participants in the original studies were included

[1]Education was not significantly related to age, gender, or any of the variables reported in the "Results" section.

[2]Mini-Mental Status Examination scores for two participants were not available.

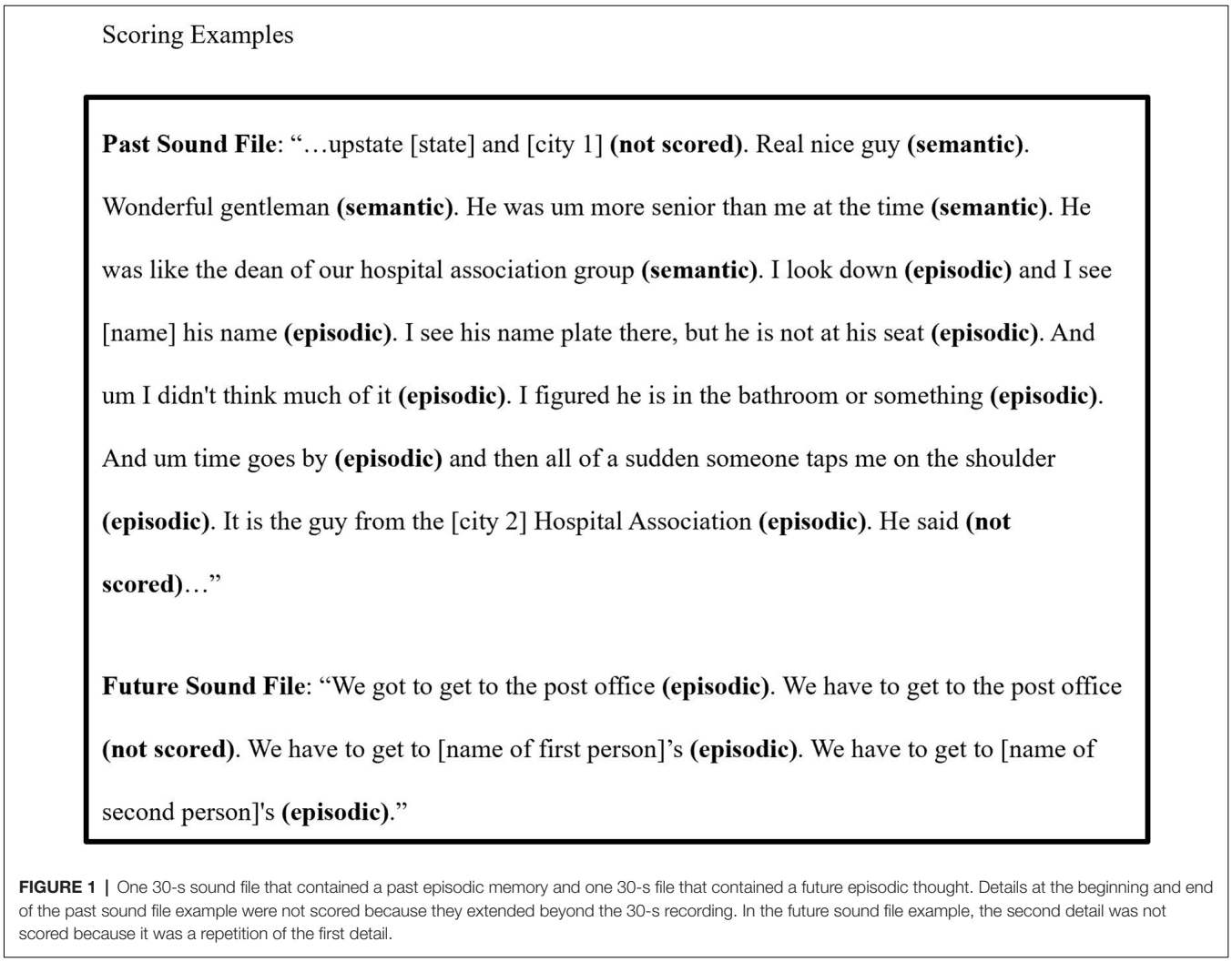


FIGURE 1 | One 30-s sound file that contained a past episodic memory and one 30-s file that contained a future episodic thought. Details at the beginning and end of the past sound file example were not scored because they extended beyond the 30-s recording. In the future sound file example, the second detail was not scored because it was a repetition of the first detail.

in this secondary data analysis. One was a test of working memory, measured using the Digit Span Backward subtest raw score from the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III DSB; Wechsler, 1997), and the other was a test of episodic memory, measured with the California Verbal Learning Test, First or Second Edition long delay cued recall raw score (CVLT LDC; Delis et al., 1987, 2000). We selected these two cognitive tests because they have been used in laboratory-based studies that have found relationships between episodic detail generation and working memory (Addis et al., 2008) or have suggested no relationship between episodic detail generation and verbal episodic memory (Grilli et al., 2018a). Only a subset of participants received these tests because, in the original study (Polsinelli et al., 2020), the cognitive battery evolved across two dissertation projects (Polsinelli, 2017; Moseley, 2018), and therefore the entire study sample did not receive the same cognitive tests. Approximately two-thirds of the sample received WAIS-III DSB (n = 72, M = 7.47, SD = 2.26). Twenty-three participants received the first edition of the CVLT LDC, recalling 12.78 words on average (SD = 2.56), and 46 received the second edition (M = 10.28, SD = 2.99). In total, 69 participants had CVLT LDC data from either the first or second edition (M = 11.12, SD = 3.07)[3].

Statistical Procedures

Consistent with established procedures to determine interrater reliability, one primary rater scored all participants and a secondary rater independently scored approximately 20%, or 21 participants (Verfaellie et al., 2014; Grilli et al., 2018b). Cronbach's alpha was calculated to determine how reliably we could identify the sound files that contained autobiographical memories or future thoughts and the episodic and semantic scoring of the details within the episodic memories and future episodic thoughts.

Two sets of analyses examined the relationship between age and autobiographical memory and future thought

[3]Because 23 participants only had long delay cued recall scores from the first edition of the CVLT, we ran the correlation with EAR episodic specificity separately for those with first edition CVLT scores and those with second edition CVLT scores. In both cases, we obtained the same results (i.e., no correlation between episodic memory and past episodic specificity).

sharing. First, to investigate whether age was related to a reduction in the sharing of episodic or semantic memories, we conducted non-parametric (Spearman) partial correlations to examine the relationship between age and sound files that contained autobiographical memories—overall, and both episodic and semantic memories separately (Kim, 2015). Non-parametric analyses were selected given that some data were not normally distributed. The number of sound files that contained any speech from the participant was used as the covariate in these analyses to control for the wide variation of speech production across participants. These analyses were repeated to examine the correlation between age and future thought sharing. Second, to investigate whether age was associated with lower detail generation while describing episodic memories, we conducted Spearman correlations between age and the average number of total details, as well as episodic and semantic details separately, captured per autobiographical episodic memory. A test of correlations from dependent samples with overlapping variables was used to compare the magnitudes of statistically significant partial correlations when appropriate (Steiger, 1980; Diedenhofen and Musch, 2015). This second set of analyses also were conducted on details generated during future episodic thought sharing.

To investigate additional features regarding the quality of social sharing of autobiographical memory, Spearman correlations were conducted to determine if the average frequency of episodic details per future episodic thought description correlated with autobiographical episodic memory description. We also used Spearman correlations to examine the association of the average number of episodic details produced in sound files that contained autobiographical episodic memories with working memory (i.e., WAIS-III DSB) and with a standardized cognitive test of verbal episodic memory (i.e., CVLT LDC). To contextualize our autobiographical memory findings more broadly, we also examined the relationship between age and our laboratory-based measures of working memory and verbal episodic memory. Finally, we conducted an independent samples *t*-test, to investigate gender effects on average episodic detail production (Fox and Weisberg, 2019). All analyses were two-tailed. We did not conduct an *a priori* power analysis and none of the analyses were pre-registered. All analyses and graphs were conducted in or created with RStudio (Wickham, 2007, 2016; Kim, 2015; Auguie, 2017; Fox and Weisberg, 2019; R Core Team, 2019).

RESULTS

Interrater Reliability

Good to excellent reliability was achieved for the total number of sound files that contained either an autobiographical memory or future thought, and for episodic and semantic subtypes separately (Cronbach's alpha range 0.86–0.99). Excellent reliability was also achieved for the total number of details produced across sound files that contained an autobiographical episodic memory or future thought, whether analyzed together or separately (Cronbach's alpha range 0.91–0.995).

Sample Included in Autobiographical Memory Analyses

To ensure that individual scores were reliable, we identified four participants who provided fewer than five sound files with an autobiographical memory or autobiographical future thought and an average of 7.50 (SD = 3.11) total details. In comparison, the remaining 98 participants produced an average of 27.16 (SD = 18.43) sound files with autobiographical memories or autobiographical future thoughts and approximately 84.98 (SD = 70.36) total details. Thus, these four participants were excluded based on having too few memories/thoughts and details to likely be reliable. Notably, they were removed before any further analyses of the data. In the remaining sample, the average age was 76.07 (SD = 5.94), with 44 men (age: M = 75.98, SD = 6.49) and 54 women (age: M = 76.15, SD = 5.51). There was an average of 81.84 sound files (SD = 44.04) that captured speech and 23.65 sound files (SD = 15.77) that included past autobiographical sharing. Of these sound files, an average of 11.17 (SD = 8.95) contained an autobiographical episodic memory, and participants generated an average of 44.24 (SD = 42.85) total details for these memories. Also, on average, 2.93 (SD = 2.99) focused on future episodic thoughts, with participants generating an average of 6.07 (SD = 6.88) total details for these thoughts.

Relationships Between Age and Autobiographical Sharing

Age and the Frequency of Autobiographical Memory and Future Thought Sharing

When controlling for the amount of conversation captured by the EAR, age was negatively correlated with overall autobiographical memory sharing, $r_{s(95)} = -0.32$, $p = 0.001$ (**Figure 2**). Age was significantly and negatively related to sharing episodic memories, $r_{s(95)} = -0.24$, $p = 0.02$, and to sharing semantic memories, $r_{s(95)} = -0.29$, $p = 0.004$ (both including amount of conversation as a covariate). A comparison of the magnitude of these two effects revealed that they were not significantly different, $z = 0.45$, $p = 0.65$.

Alternately, age and autobiographical future thought sharing, covaried for amount of conversation, were not significantly correlated, $r_{s(95)} = -0.09$, $p = 0.38$ (**Figure 3**). However, whereas the correlation between production of future episodic thoughts and age was not significant, $r_{s(95)} = -0.03$, $p = 0.80$, the correlation between future semantic thoughts and age was significant and negative, $r_{s(95)} = -0.21$, $p = 0.04$ (both including amount of conversation as a covariate).

Age and Autobiographical Memory and Future Thought Detail Sharing

In regard to past memory detail generation, age was negatively related to the average number of total details produced while describing autobiographical episodic memories, $r_{s(94)} = -0.30$, $p = 0.003$ (**Figure 4**). Age was significantly and negatively related to average episodic, $r_{s(94)} = -0.23$, $p = 0.02$, and semantic details, $r_{s(94)} = -0.30$, $p = 0.003$, separately. The magnitude of these two relationships did not significantly differ, $z = 0.57$, $p = 0.57$.

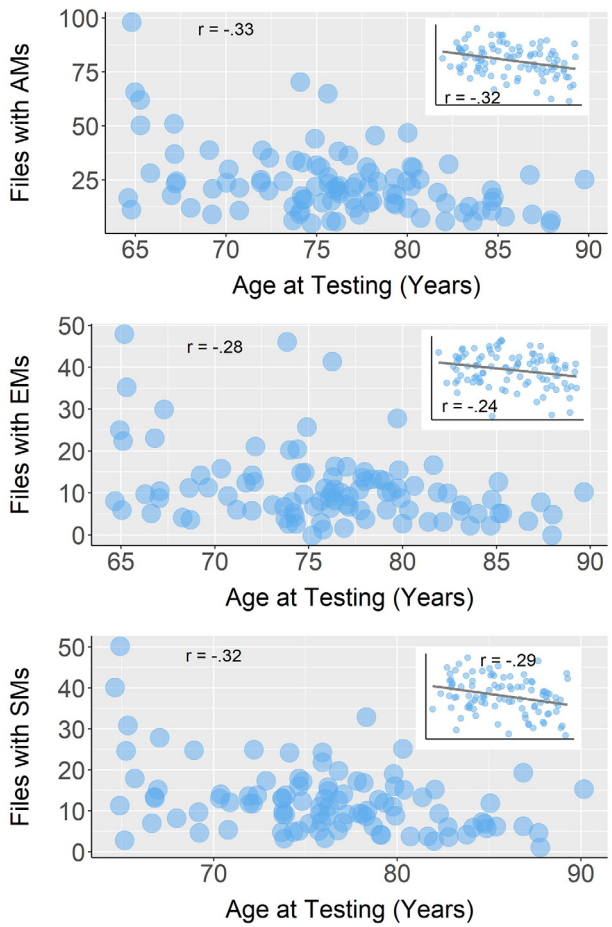


FIGURE 2 | Correlations between age and **(A)** sound files with autobiographical memories (AMs), **(B)** sound files with episodic memories (EMs), and **(C)** sound files with semantic memories (SMs). Raw data and correlation coefficients of the bivariate Spearman correlations are represented in the larger plots. The partial Spearman correlations are shown in the small plots with regression lines. For all three partial correlations, we used a covariate of the number of sound files where participants were engaged in conversation.

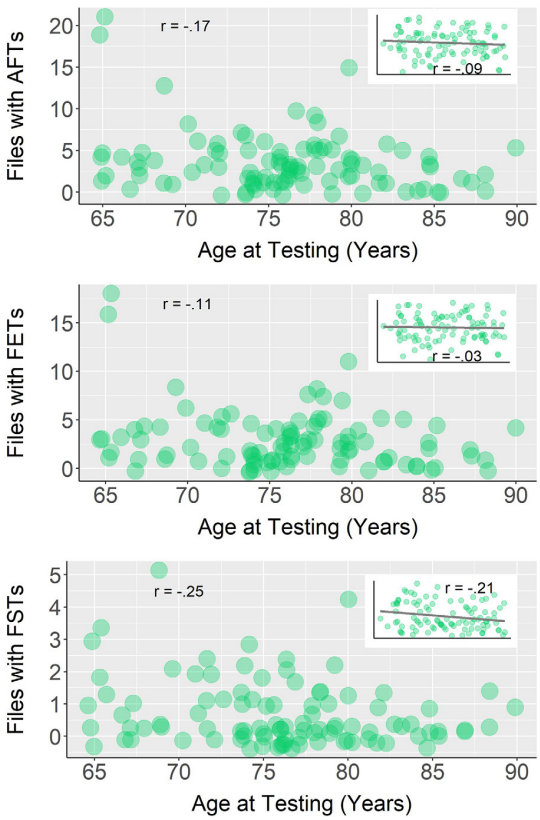


FIGURE 3 | Correlations between age and **(A)** sound files with autobiographical future thoughts (AFTs), **(B)** sound files with future episodic thoughts (FETs), and **(C)** sound files with future semantic thoughts (FSTs). Raw data and correlation coefficients of the bivariate Spearman correlations are represented in the larger plots. The partial Spearman correlations are shown in the small plots with regression lines. For all three partial correlations, we used a covariate of the number of sound files where participants were engaged in conversation.

For future episodic thought sharing, age and the average total details produced was not significant, $r_{s(80)} = 0.01$, $p = 0.90$. Similarly, there was no significant correlation between age and average episodic details, $r_{s(80)} = -0.02$, $p = 0.88$, or age and average semantic details, $r_{s(80)} = 0.04$, $p = 0.72$. **Figure 5** depicts these three relationships.

Episodic Detail Sharing in Daily Conversation

Episodic Detail Generation During Past Episodic Event and Future Episodic Thought Sharing

The correlation between the average number of episodic details produced during autobiographical episodic memory sharing and future episodic thought sharing, while positive, was not significant, $r_{s(80)} = 0.18$, $p = 0.10$ (**Figure 6**).

Laboratory-Based Cognitive Tests of Working Memory and Episodic Memory

Contrary to laboratory-based findings, our measure of working memory, DSB, was not significantly correlated with the average number of episodic details produced when participants shared autobiographical episodic memories in daily conversation, $r_{s(68)} = -0.01$, $p = 0.96$. However, consistent with laboratory-based findings, the association between a laboratory measure of episodic memory (i.e., CVLT LDC) and episodic detail provided while sharing autobiographical episodic memories was also not significant, $r_{s(65)} = 0.05$, $p = 0.66$. See **Figure 7** for both correlations. Interestingly, age was not significantly correlated with our laboratory-based cognitive measure of working memory, $r_s(68) = -0.08$, $p = 0.49$, or verbal episodic memory, $r_s(65) = -0.12$, $p = 0.32$.

Gender Comparison of Episodic Detail Generation

Contrary to some laboratory-based findings, men ($M = 2.95$, $SD = 1.00$) and women ($M = 2.81$, $SD = 0.85$) produced

FIGURE 4 | Bivariate Spearman correlations between average (avg) **(A)** total, **(B)** episodic, and **(C)** semantic details produced in sound files containing autobiographical episodic memories (EM) and age.

FIGURE 5 | Bivariate Spearman correlations between average (avg) **(A)** total, **(B)** episodic, and **(C)** semantic details produced in sound files containing future episodic thoughts (FET) and age.

comparable averages of the number of episodic details when describing episodic memories, $t_{(94)} = 0.73$, $p = 0.47$, 95% CI (−0.24, 0.51; **Figure 8**). Co-varying for age did not alter these outcomes, $p = 0.50$.

DISCUSSION

Individuals share autobiographical memories to juggle a wide range of everyday social situations, including navigating social bonds, teaching others, and solving problems (Pillemer, 1992; Bluck and Alea, 2002; Pasupathi et al., 2002; Bluck, 2003; Bluck et al., 2005). Prior laboratory-based research has extensively studied how the sharing of autobiographical memories appears to change with older age, with one key finding being that there is a robust age-related reduction in episodic memories that is evident in both the frequency at which such memories are retrieved (Piolino et al., 2006; Ros et al., 2009, 2017; Ford et al., 2014) and how much episodic detail is shared to describe unique life events (Levine et al., 2002; Addis et al., 2008; Gaesser et al., 2011; Devitt et al., 2017). The present study utilized the EAR, a method that has recently been applied to capture memories being shared in everyday conversations as they happen (Demiray et al., 2018, 2019; Brianza and Demiray, 2019), to extend the assessment of episodic specificity out of the laboratory and shed light on how older age and episodic specificity are related in the real-world.

By tracking the frequency with which autobiographical memories were generated in social conversations over 4 days, we found that, in an older adult sample, older participants shared fewer autobiographical episodic memories than younger participants. Interestingly, older age also was associated with reduced autobiographical semantic memory sharing in daily conversation. Notably, the magnitude of the effects of age on episodic memory and semantic memory sharing were similar, and not significantly different when directly compared. Together, these findings suggest that older age was linked to a global reduction in the social sharing of autobiographical memories, independent of how often individuals conversed. These results extend our knowledge of the relationship between age and episodic memory by showing that a reduction in the retrieval of episodic memories is not only evident

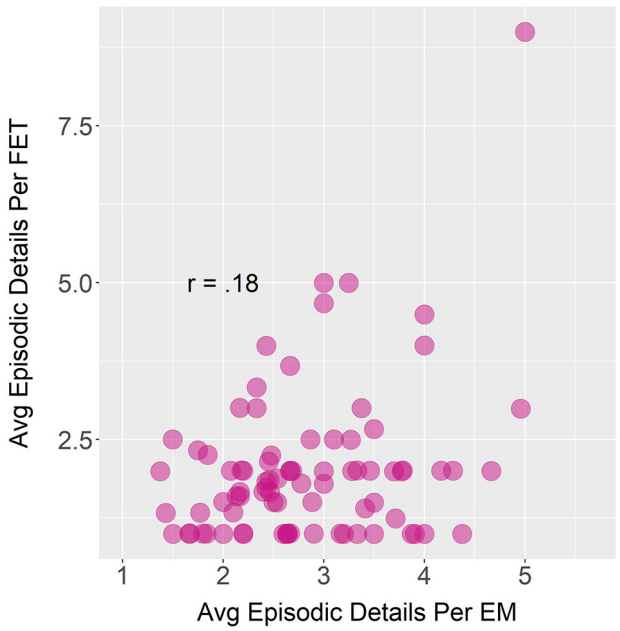


FIGURE 6 | Bivariate Spearman correlation between average (avg) past episodic details and future episodic details generated in sound files with episodic memories (EM) and future episodic thoughts (FET), respectively.

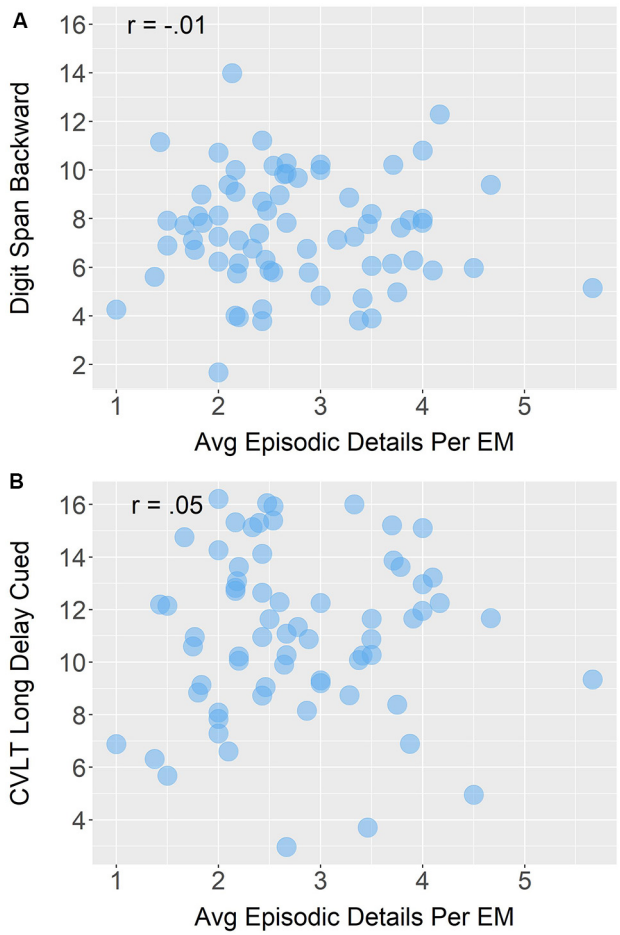


FIGURE 7 | Bivariate Spearman correlations between average (avg) episodic detail generation in sound files that contained episodic memories (EM) and **(A)** measures of working memory (i.e., Digit Span Backward raw score) and **(B)** episodic memory (i.e., California Verbal Learning Test [CVLT] long delay cued recall raw score). The jitter of data points was used to ensure minimal overlap, which pushed some CVLT scores of 16 (maximum score) above that value.

in the laboratory (Piolino et al., 2006; Ros et al., 2009, 2017; Ford et al., 2014), but also during memory sharing in daily conversations. Our findings further suggest that age-related reductions in memory retrieval may not be specific to episodic reflections, at least when autobiographical memory is assessed in daily conversations in the context of increased age among older adults. We interpret these findings as consistent with the integrity hypothesis of default network functioning (Andrews-Hanna et al., 2014, 2019), which predicts that age-related decline in the functionality of the default network should result in a broad reduction in the frequency at which one's thoughts drift towards the personal past.

Regarding details attached to autobiographical episodic memories, we found a significant negative relationship between age and overall detail generation. Interestingly, older age was not exclusively linked to less episodic detail production, as there was a significant negative relationship between age and semantic detail as well, consistent with the integrity hypothesis of default network functioning (Andrews-Hanna et al., 2014, 2019). The relationship between age and episodic detail did not differ statistically from the magnitude of the age and semantic detail relationship. Thus, similar to our analyses of overall memory sharing, our findings likely suggest that older age has a global, negative effect on memory detail generation while sharing episodic memories. While prior work indicates that reductions in episodic detail may be relevant to social cognitive behavior, including social problem-solving (Vandermorris et al., 2013; Madore and Schacter, 2014) and empathy or intentions to help others (Gaesser, 2013; Gaesser and Schacter, 2014), future research can investigate the broader implications of less semantic detail retrieval in social recollection.

Relative to the autobiographical memory findings, our naturalistic observation of autobiographical future thought-sharing produced weaker relationships with older age. These outcomes are surprising, considering the theoretical and empirical link between remembering and imagining (Schacter and Addis, 2007; Addis et al., 2008; Madore et al., 2014; Addis, 2018). Interestingly, whereas the negative correlation between older age and episodic future thought sharing was not significant, there was a significant, *negative* association between older age and future semantic thought sharing. Therefore, for both remembering and imagining in daily conversation, older age was linked to less outward retrieval of semantic information. On the one hand, these findings are difficult to reconcile with an executive coupling framework (Piolino et al., 2010; Turner and Spreng, 2015; Spreng et al., 2018) or narrative style differences

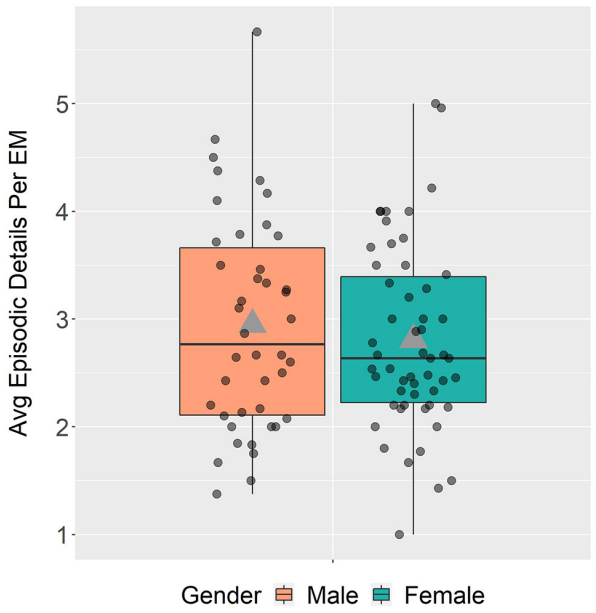


FIGURE 8 | Average (avg) episodic details per episodic memory (EM) in males and females.

(James et al., 1998; Trunk and Abrams, 2009; Gaesser et al., 2011), and seem more consistent with the integrity hypothesis of default network functioning (Andrews-Hanna et al., 2014, 2019). On the other hand, the low number of future thoughts captured for many participants raises some concerns about the stability of these findings. Similarly, the low frequency of future thought sampling may have compromised the reliability of our estimates of individual differences in details attached to future thoughts, which may explain why we did not find a significant association between age and episodic or semantic detail generation while describing future episodic thoughts. The low frequency of future thought sampling also may be relevant to understanding why we did not find that individual differences in episodic detail generation while remembering were reflected in future event sharing in daily conversation. Therefore, future studies will need to examine whether, with a greater sampling of shared future thoughts in daily conversation, stronger associations to older age are detectable.

In addition to investigating the frequency and detail make-up of autobiographical memory and future thought sharing, we also examined the degree to which a few laboratory-based outcomes on episodic detail generation were evident in the sharing of episodes in daily conversation. Similar to laboratory-based research (Palombo et al., 2015; Grilli et al., 2018a), we did not find a correlation between a list learning task and episodic detail generation while sharing autobiographical memories. Therefore, findings from both laboratory and naturalistic methods suggest that traditional verbal learning and memory tests may not provide much insight into the nature of real-world episodic specificity. More broadly, these results highlight the clinical importance of assessing episodic specificity of autobiographical memory sharing.

There were also two notable differences between laboratory-based work and our study of naturalistic autobiographical memory sharing. In our study, working memory was not associated with episodic specificity, a correlation that laboratory research has found in older adults. It is possible that our naturalistic sampling of autobiographical episodic memory sharing was not related to working memory because the memories captured were often shorter than what is reported in laboratory studies. We noticed that in their natural conversations, individuals often completed their description of a memory in less than the 30-s window of the EAR recording. Thus, it may be that working memory, perhaps in particular feature binding (i.e., episodic buffer; Baddeley, 2000; Piolino et al., 2010), is less important for the telling of shorter memories, and instead crucial for integrating and maintaining content over extended sharing. Another possibility is that our ability to capture a relationship was obscured given that only memory fragments were often captured. Future research using the EAR for this purpose could increase the recording time to better capture verbalized autobiographical memory sharing and examine different aspects of working memory more fully.

In addition to working memory, in contrast to laboratory-based work, we did not reveal gender differences in autobiographical memory sharing, as men and women were equally specific in their daily lives while conversing with others about past episodes. Although the current study did not directly address an underlying mechanism for the discrepancy between the findings of our naturalistic assessment and those of laboratory-bases studies, one possibility is that social and contextual interactions may modulate episodic specificity. In other words, men and women may express various degrees of autobiographical memory specificity depending on context (e.g., other individuals involved in the conversation; Aukett et al., 1988; Grysman and Hudson, 2013) and on the purpose that sharing serves (e.g., social bonding, teaching; Bluck, 2003). These factors may be critical to whether gender differences emerge. An alternative, and not mutually exclusive, explanation could be related to differences in available cues provided in naturalistic and laboratory environments. There is evidence that men may be more sensitive to visual/spatial cuing than women such that particular brain regions associated with greater detail, reliving, and richness showed greater activation to visuospatial compared to verbal cues in an autobiographical memory task (St. Jacques et al., 2011). These results might mean that the level of episodic specificity in men and women depends on the environmental cues of the experiment (i.e., visual/spatial cues in naturalistic studies and verbal cues in laboratory studies). Finally, many of the studies examining gender differences in autobiographical memory specificity were conducted in younger individuals (Wang et al., 2011; Fuentes and Desrocher, 2013; Wang, 2013). These findings have been extended to older individuals (Pillemer et al., 2003), but more research should be done to fully investigate possible gender differences in the specificity of socially shared personal episodes in older individuals.

The present study has a few main limitations that are worth considering. First, although episodic specificity differences have been found in laboratory-based studies of older adult cohorts

(Piolino et al., 2002; De Beni et al., 2013), more often, an older adult group is compared to a young adult group. Thus, it may be that older adults, as a cohort, tend to generate more semantic memories in everyday life and provide more semantic details than young individuals, as would be predicted by executive coupling (Piolino et al., 2010; Turner and Spreng, 2015; Spreng et al., 2018) and narrative style theoretical models (James et al., 1998; Trunk and Abrams, 2009; Gaesser et al., 2011), but there is a relative decline in these features across older adulthood. A future study could address this possibility by evaluating autobiographical memory sharing across the adult age spectrum. A young vs. older adult comparison could further evaluate the degree to which other laboratory-based findings related to aging and autobiographical memory translate to real-world remembering in daily conversations, such as a relationship to working memory. Second, as is the case for studies of real-world behavior, our naturalistic recording approach inherently provides decreased experimental control over study conditions when compared to laboratory-based studies. Future research will need to examine whether real-world contextual features, such as with whom and where one is speaking, influence the relationship between older age and autobiographical memory. Such social and environmental differences could have contributed to the divergence between laboratory and real-world autobiographical memory assessment reported in the present study. In some respects, variability in environmental and social features can be an advantage, revealing contexts that prominently affect performance among older adults. We suggest that by pairing both in- and out-of-laboratory designs, one could strike a balance between the generalizability that naturalistic observation affords and the high level of experimental control of the laboratory. A third limitation is that the EAR can only capture verbalized memories and future thoughts, leaving in the dark memories and simulations that are part of internal cognition. This may be particularly noteworthy for the naturalistic assessment of future thinking, given that we captured fewer instances of sharing future thoughts (thus replicating Demiray et al., 2018). Other everyday assessment methods, such as diary studies or other thought sampling designs of self-reported vividness and content of future thoughts, may be better suited for capturing and assessing the quality of future thoughts. Fourth, with longer EAR snippets, we would be in a better position to examine the dynamic relationship between episodic and semantic detail use (Devitt et al., 2017). Relatedly, collecting EAR data over more than 4 days would likely boost the frequency of captured future thought-sharing in daily conversation, which may improve our ability to reliably evaluate the relationships between real-world autobiographical future thinking and older age. More days of EAR data collection would also presumably boost the frequency of autobiographical memories captured and variety of environmental contexts, which would allow for further investigation of how contextual features of daily conversation (e.g., with whom or where one is engaged in a conversation) relate to autobiographical memory sharing. Finally, while the selection of the two standard cognitive tests used in the present study was based on data from prior studies (Addis et al., 2008; Grilli et al., 2018a), not all participants received them and there was only one test per domain. A future study could administer a larger battery to more participants that includes a greater range of assessments for each cognitive construct.

Despite these limitations, the present study sheds new light on how older individuals share autobiographical memories in real-world contexts, and in the process, lays the foundation for future research using new technologies to further investigate a host of clinical and functional implications of memory sharing. On the clinical side, reduced episodic specificity has been documented in a range of neurologic conditions (e.g., dementia and amnesia, Irish et al., 2011, 2012; Palombo et al., 2018) and psychiatric conditions (e.g., schizophrenia and depression; Potheegadoo et al., 2014; Söderlund et al., 2014; MacDougall et al., 2015). Both in-laboratory and unobtrusive naturalistic acquisition of autobiographical memory could be useful clinical tools for tracking changes in episodic specificity in these populations. For example, evaluating baseline episodic specificity in-clinic or in-laboratory could indicate a need for follow-up assessments of specificity using naturalistic methods such as the EAR. Furthermore, including real-world observation of autobiographical memory could provide increased coherence in clinicians' understanding of subjective patient complaints with observed deficits. Given that our standard working memory and verbal episodic memory tests were not significantly related to age, real-world memory assessment may be more sensitive to age-related cognitive changes than many laboratory-based cognitive tests, which tend to be socially decontextualized. However, we acknowledge that in future work, other standard cognitive tests will need to be compared to naturalistic assessment. On the functional side, autobiographical memory has been linked to performance on problem-solving tasks (Vandermorris et al., 2013; Madore and Schacter, 2014; McFarland et al., 2017), creative thinking (Madore et al., 2015), and emotion regulation (Jing et al., 2016). Therefore, naturalistic assessment of daily conversation using new technologies such as the EAR has the potential to be another important method for understanding the degree to which autobiographical memories are being adaptively applied to a variety of cognitive processes that are critical for wellbeing (Seifert et al., 2018).

In sum, in the present study, we used a new technology that has been recently applied to study memory sharing in natural social contexts (Demiray et al., 2018, 2019; Brianza and Demiray, 2019) and found that cross-sectionally, older individuals in the sample demonstrated lower frequencies of autobiographical memory sharing in everyday conversations, a relationship that affected both episodic and semantic memories. For episodic memories, older age was also linked to a reduction in the level of detail at which events were described. In terms of future autobiographical thought sharing, age was neither related to sharing of future episodic thoughts nor sharing of all details in future episodic thoughts. Interestingly, age was negatively correlated with future semantic thought sharing. Additional analyses revealed a commonality with findings from laboratory research (i.e., a lack of a relationship between specificity and a cognitive test of episodic memory) and differences (i.e., lack of a relationship between episodic detail in memories and future thoughts, lack of a relationship between episodic specificity and

working memory, and gender differences). Overall, our findings align with recent research showing that the EAR and similar new technologies for unobtrusively capturing cognition "in the wild" can complement laboratory-based approaches and provide new insights into the actual use of autobiographical memory in everyday life.

DATA AVAILABILITY STATEMENT

The dataset analyzed for this study and the corresponding data analysis code can be found here: https://osf.io/f3euv/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board University of Arizona. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Contributor Role Taxonomy (CRediT) taxonomy. AW: methodology, formal analysis, data curation, writing—original draft, writing—reviewing and editing, and visualization. MM: conceptualization, methodology, writing—reviewing and editing, and supervision. JA-H: conceptualization, writing—reviewing and editing. AP: investigation, data curation, project administration, and funding acquisition. SM: investigation, data curation, and project administration. EG: investigation, resources, writing—reviewing and editing, and supervision. MG: conceptualization, methodology, resources, supervision of formal analysis, writing—original draft, writing—review and editing, and supervision.

FUNDING

This research was supported by the Mind and Life Institute's 1440 Grant (AP; 2015-1440-Polsinelli). This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1746060 (AW). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. MG and JA-H were supported by project grants from the Arizona Alzheimer's Consortium/Arizona Department of Health Services.

ACKNOWLEDGMENTS

We thank research assistant, Leah Mann, for her help with reliability scoring and Dr. Mieke Verfaellie for helpful comments on an earlier version of this manuscript.

REFERENCES

Addis, D. R. (2018). Are episodic memories special? On the sameness of remembered and imagined event simulation. *J. R. Soc. N. Z.* 48, 64–88. doi: 10.1080/03036758.2018.1439071

Addis, D. R., Wong, A. T., and Schacter, D. L. (2008). Age-related changes in the episodic simulation of future events. *Psychol. Sci.* 19, 33–41. doi: 10.1111/j.1467-9280.2008.02043.x

Alea, N., and Bluck, S. (2003). Why are you telling me that? A conceptual model of the social function of autobiographical memory. *Memory* 11, 165–178. doi: 10.1080/741938207

Alea, N., and Bluck, S. (2007). I'll keep you in mind: the intimacy function of autobiographical memory. *Appl. Cogn. Psychol.* 21, 1091–1111. doi: 10.1002/acp.1316

Andreano, J. M., and Cahill, L. (2009). Sex influences on the neurobiology of learning and memory. *Learn. Mem.* 16, 248–266. doi: 10.1101/lm.918309

Andrews-Hanna, J. R., Grilli, M. D., and Irish, M. (2019). "A review and reappraisal of the default network in normal aging and dementia," in *Oxford Research Encyclopedia of Psychology*, eds R. T. Knight, S. D. Neupert, N. D. Anderson, H. W. Wahl and N. A. Pachana (Oxford: Oxford University Press).

Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N Y Acad. Sci.* 1316, 29–52. doi: 10.1111/nyas.12360

Andrews-Hanna, J. R., Snyder, A. Z., Vincent, J. L., Lustig, C., Head, D., Raichle, M. E., et al. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron* 56, 924–935. doi: 10.1016/j.neuron.2007.10.038

Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3.* Available online at: https://CRAN.R-project.org/package=gridExtra. Accessed May 13, 2020.

Aukett, R., Ritchie, J., and Mill, K. (1988). Gender differences in friendship patterns. *Sex Roles* 19, 57–66. doi: 10.1007/BF00292464

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/s1364-6613(00)01538-2

Barnabe, A., Whitehead, V., Pilon, R., Arsenault-Lapierre, G., and Chertkow, H. (2012). Autobiographical memory in mild cognitive impairment and Alzheimer's disease: a comparison between the Levine and Kopelman interview methodologies. *Hippocampus* 22, 1809–1825. doi: 10.1002/hipo.22015

Bastin, C., Feyers, D., Jedidi, H., Bahri, M. A., Degueldre, C., Lemaire, C., et al. (2013). Episodic autobiographical memory in amnestic mild cognitive impairment: what are the neural correlates? *Hum. Brain Mapp.* 34, 1811–1825. doi: 10.1002/hbm.22032

Berntsen, D., and Hall, N. M. (2004). The episodic nature of involuntary autobiographical memories. *Mem. Cognit.* 32, 789–803. doi: 10.3758/bf03195869

Bluck, S. (2003). Autobiographical memory: exploring its functions in everyday life. *Memory* 11, 113–123. doi: 10.1080/741938206

Bluck, S., and Alea, N. (2002). "Exploring the functions of autobiographical memory: why do I remember the autumn?" in *Critical Advances of Reminiscence Work*, eds J. D. Webster and B. K. Haight (New York, NY: Springer Publishing Company), 61–75.

Bluck, S., Alea, N., Habermas, T., and Rubin, D. C. (2005). A Tale of three functions: the self-reported uses of autobiographical memory. *Soc. Cogn.* 23, 91–117. doi: 10.1521/soco.23.1.91.59198

Brianza, E., and Demiray, B. (2019). Future time perspective and real-life utterances about the future in young and older adults. *GeroPsych* 32, 161–173. doi: 10.1024/1662-9647/a000216

Conway, M. A., and Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychol. Rev.* 107, 261–288. doi: 10.1037/0033-295x.107.2.261

D'Argembeau, A., Renaud, O., and Van der Linden, M. (2011). Frequency, characteristics, and functions of future-oriented thoughts in daily life. *Appl. Cogn. Psychol.* 25, 96–103. doi: 10.1002/acp.1647

Davis, P. J. (1999). Gender differences in autobiographical memory for childhood emotional experiences. *J. Pers. Soc. Psychol.* 76, 498–510. doi: 10.1037/0022-3514.76.3.498

De Beni, R., Borella, E., Carretti, B., Zavagnin, M., Lazzarini, L., and Milojevi, G. (2013). Remembering the past and imagining the future: age-related differences between young, young-old and old-old. *Aging Clin. Exp. Res.* 25, 89–97. doi: 10.1007/s40520-013-0003-3

De Brigard, F., Giovanello, K. S., Stewart, G. W., Lockrow, A. W., O'Brien, M. M., and Spreng, R. N. (2016). Characterizing the subjective experience of episodic past, future, and counterfactual thinking in healthy younger and older adults. *Q. J. Exp. Psychol.* 69, 2358–2375. doi: 10.1080/17470218.2015.1115529

Delis, D. C., Kramer, J. H., Kaplan, E., and Ober, B. A. (2000). *California Verbal Learning Test, Second Edition (CVLT-II).* San Antonio, TX: Psychological Corporation.

Delis, D. C., Kramer, J., Kaplan, E., Ober, B. A., and Fridlund, A. (1987). *The California Verbal Learning Test.* New York, NY: Psychological Corporation.

Demiray, B., Mehl, M. R., and Martin, M. (2018). Conversational time travel: evidence of a retrospective bias in real life conversations. *Front. Psychol.* 9:2160. doi: 10.3389/fpsyg.2018.02160

Demiray, B., Mischler, M., and Martin, M. (2019). Reminiscence in everyday conversations: a naturalistic observation study of older adults. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 74, 745–755. doi: 10.1093/geronb/gbx141

Devitt, A. L., Addis, D. R., and Schacter, D. L. (2017). Episodic and semantic content of memory and imagination: a multilevel analysis. *Mem. Cognit.* 45, 1078–1094. doi: 10.3758/s13421-017-0716-1

Diedenhofen, B., and Musch, J. (2015). Cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 10:e0121945. doi: 10.1371/journal.pone.0121945

Fjell, A. M., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Ras, N., et al. (2009). High consistency of regional cortical thinning in aging across multiple samples. *Cereb. Cortex* 19, 2001–2012. doi: 10.1093/cercor/bhn232

Ford, J. H., Rubin, D. C., and Giovanello, K. S. (2014). Effects of task instruction on autobiographical memory specificity in young and older adults. *Memory* 22, 722–736. doi: 10.1080/09658211.2013.820325

Fox, J., and Weisberg, S. (2019). *An R Companion to Applied Regression, Third Edition.* Thousand Oaks, CA: Sage.

Fuentes, A., and Desrocher, M. (2013). The effects of gender on the retrieval of episodic and semantic components of autobiographical memory. *Memory* 21, 619–632. doi: 10.1080/09658211.2012.744423

Gaesser, B. (2013). Constructing memory, imagination, and empathy: a cognitive neuroscience perspective. *Front. Psychol.* 3:576. doi: 10.3389/fpsyg.2012.00576

Gaesser, B., and Schacter, D. L. (2014). Episodic simulation and episodic memory can increase intentions to help others. *Proc. Natl. Acad. Sci. U S A* 111, 4415–4420. doi: 10.1073/pnas.1402461111

Gaesser, B., Sacchetti, D. C., Addis, D. R., and Schacter, D. L. (2011). Characterizing age-related changes in remembering the past and imagining the future. *Psychol. Aging* 26, 80–84. doi: 10.1037/a0021054

Grilli, M. D., Wank, A. A., Bercel, J. J., and Ryan, L. (2018a). Evidence for reduced autobiographical memory episodic specificity in cognitively normal middle-aged and older individuals at increased risk for Alzheimer's disease dementia. *J. Int. Neuropsychol. Soc.* 24, 1073–1083. doi: 10.1017/s1355617718000577

Grilli, M. D., Wank, A. A., and Verfaellie, M. (2018b). The life stories of adults with amnesia: insights into the contribution of the medial temporal lobes to the organization of autobiographical memory. *Neuropsychologia* 110, 84–91. doi: 10.1016/j.neuropsychologia.2017.03.013

Grysman, A., and Hudson, J. A. (2013). Gender differences in autobiographical memory: developmental and methodological considerations. *Dev. Rev.* 33, 239–272. doi: 10.1016/j.dr.2013.07.004

Hill, P. F., and Emery, L. J. (2013). Episodic future thought: contributions from working memory. *Conscious. Cogn.* 22, 677–683. doi: 10.1016/j.concog.2013.04.002

Hodgetts, C. J., Postans, M., Warne, N., Varnava, A., Lawrence, A. D., and Graham, K. S. (2017). Distinct contributions of the fornix and inferior longitudinal fasciculus to episodic and semantic autobiographical memory. *Cortex* 94, 1–14. doi: 10.1016/j.cortex.2017.05.010

Irish, M., Addis, D. R., Hodges, J. R., and Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: evidence from semantic dementia. *Brain* 135, 2178–2191. doi: 10.1093/brain/aws119

Irish, M., Hornberger, M., Lah, S., Miller, L., Pengas, G., Nestor, P. J., et al. (2011). Profiles of recent autobiographical memory retrieval in semantic dementia, behavioural-variant frontotemporal dementia and Alzheimer's disease. *Neuropsychologia* 49, 2694–2702. doi: 10.1016/j.neuropsychologia.2011.05.017

James, L. E., Burke, D. M., Austin, A., and Hulme, E. (1998). Production and perception of "verbosity" in younger and older adults. *Psychol. Aging* 13, 355–367. doi: 10.1037/0882-7974.13.3.355

Jing, H. G., Madore, K. P., and Schacter, D. L. (2016). Worrying about the future: an episodic specificity induction impacts problem solving, reappraisal and well-being. *J. Exp. Psychol. Gen.* 145, 402–418. doi: 10.1037/xge0000142

Kim, S. (2015). *ppcor: Partial and Semi-Partial (part) Correlation. R package version 1.1.* Available online at: https://CRAN.R-project.org/package=ppcor. Accessed May 13, 2020.

Lambon Ralph, M. A., Ehsan, S., Baker, G. A., and Rogers, T. T. (2012). Semantic memory is impaired in patients with unilateral anterior temporal lobe resection for temporal lobe epilepsy. *Brain* 135, 242–258. doi: 10.1093/brain/awr325

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., and Moscovitch, M. (2002). Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol. Aging* 17, 677–689. doi: 10.1037/0882-7974.17.4.677

MacDonald, S., Uesiliana, K., and Hayne, H. (2000). Cross-cultural and gender differences in childhood amnesia. *Memory* 8, 365–376. doi: 10.1080/09658210050156822

MacDougall, A. G., McKinnon, M. C., Herdman, K. A., King, M. J., and Kiang, M. (2015). The relationship between insight and autobiographical memory for emotional events in schizophrenia. *Psychiatry Res.* 226, 392–395. doi: 10.1016/j.psychres.2014.12.058

Madore, K. P., Addis, D. R., and Schacter, D. L. (2015). Creativity and memory: effects of an episodic-specificity induction on divergent thinking. *Psychol. Sci.* 26, 1461–1468. doi: 10.1177/0956797615591863

Madore, K. P., Gaesser, B., and Schacter, D. L. (2014). Constructive episodic simulation: dissociable effects of a specificity induction on remembering, imagining, and describing in young and older adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 40, 609–622. doi: 10.1037/a0034885

Madore, K. P., and Schacter, D. L. (2014). An episodic specificity induction enhances means-end problem solving in young and older adults. *Psychol. Aging* 29, 913–924. doi: 10.1037/a0038209

McFarland, C. P., Primosch, M., Maxon, C. M., and Stewart, B. T. (2017). Enhancing memory and imagination improves problem solving among individuals with depression. *Mem. Cognit.* 45, 932–939. doi: 10.3758/s13421-017-0706-3

Mehl, M. R. (2017). The electronically activated recorder (EAR): a method for the naturalistic observation of daily social behavior. *Curr. Dir. Psychol. Sci.* 26, 184–190. doi: 10.1177/0963721416680611

Mehl, M. R., and Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR). *Eur. J. Psychol. Assess.* 23, 248–257. doi: 10.1027/1015-5759.23.4.248

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., and Price, J. H. (2001). The electronically activated recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behav. Res. Methods* 33, 517–523. doi: 10.3758/bf03195410

Memel, M., Wank, A. A., Ryan, L., and Grilli, M. D. (2020). The relationship between episodic detail generation and anterotemporal, posteromedial, and hippocampal white matter tracts. *Cortex* 123, 124–140. doi: 10.1016/j.cortex.2019.10.010

Moseley, S. (2018). *Cognitive and Psychosocial Associations of Hearing Loss in Older Adults.* Tucson, AZ: University of Arizona. Dissertation.

Murphy, K. J., Troyer, A. K., Levine, B., and Moscovitch, M. (2008). Episodic, but not semantic, autobiographical memory is reduced in amnestic mild cognitive impairment. *Neuropsychologia* 46, 3116–3123. doi: 10.1016/j.neuropsychologia.2008.07.004

Niedźwieńska, A. (2003). Gender differences in vivid memories. *Sex Roles* 49, 321–331. doi: 10.1023/A:1025156019547

Palombo, D. J., Alain, C., Söderlund, H., Khuu, W., and Levine, B. (2015). Severely deficient autobiographical memory (SDAM) in healthy adults: a new mnemonic syndrome. *Neuropsychologia* 72, 105–118. doi: 10.1016/j.neuropsychologia.2015.04.012

Palombo, D. J., Sheldon, S., and Levine, B. (2018). Individual differences in autobiographical memory. *Trends Cogn. Sci.* 22, 583–597. doi: 10.1016/j.tics.2018.04.007

Pasupathi, M., Lucas, S., and Coombs, A. (2002). Conversational functions of autobiographical remembering: long-married couples talk about conflicts and pleasant topics. *Discourse Process.* 34, 163–192. doi: 10.1207/s15326950dp3402_3

Pillemer, D. B. (1992). "Remembering personal circumstances: a functional analysis," in *Affect and Accuracy in Recall: Studies of "Flashbulb" Memories*, eds E. Winograd and U. Neisser (New York, NY: Cambridge University Press), 236–264.

Pillemer, D. B., Wink, P., DiDonato, T. E., and Sanborn, R. L. (2003). Gender differences in autobiographical memory styles of older adults. *Memory* 11, 525–532. doi: 10.1080/09658210244000117

Piolino, P., Coste, C., Martinelli, P., Macé, L., Quinette, P., Guillery-Girard, B., et al. (2010). Reduced specificity of autobiographical memory and aging: do the executive and feature binding functions of working memory have a role? *Neuropsychologia* 48, 429–440. doi: 10.1016/j.neuropsychologia.2009.09.035

Piolino, P., Desgranges, B., Benali, K., and Eustache, F. (2002). Episodic and semantic remote autobiographical memory in ageing. *Memory* 10, 239–257. doi: 10.1080/09658210143000353

Piolino, P., Desgranges, B., Clarys, D., Guillery-Girard, B., Taconnat, L., Isingrini, M., et al. (2006). Autobiographical memory, autonoetic consciousness, and self-perspective in aging. *Psychol. Aging* 21, 510–525. doi: 10.1037/0882-7974.21.3.510

Polsinelli, A. (2017). *Cognitive and Emotional Associations of Mindfulness in Older Adults.* Tucson, AZ: University of Arizona. Dissertation.

Polsinelli, A. J., Moseley, S. A., Grilli, M. D., Glisky, E. L., and Mehl, M. R. (2020). Natural, everyday language use provides a window into the integrity of older adults' executive functioning. *J. Gerontol. B Psychol.* doi: 10.1093/geronb/gbaa055 [Epub ahead of print].

Potheegadoo, J., Cordier, A., Berna, F., and Danion, J. M. (2014). Effectiveness of a specific cueing method for improving autobiographical memory recall in patients with schizophrenia. *Schizophr. Res.* 152, 229–234. doi: 10.1016/j.schres.2013.10.046

R Core Team. (2019). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.* Vienna: Austria. Available online at: https://www.R-project.org/. Accessed May 13, 2020.

Robbins, M. L. (2017). Practical suggestions for legal and ethical concerns with social environment sampling methods. *Soc. Psychol. Pers. Sci.* 8, 573–580. doi: 10.1177/1948550617699253

Ros, L., Latorre, J. M., and Serrano, J. P. (2009). Working memory capacity and overgeneral autobiographical memory in young and older adults. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 17, 89–107. doi: 10.1080/13825580903042650

Ros, L., Latorre, J. M., Serrano, J. P., and Ricarte, J. J. (2017). Overgeneral autobiographical memory in healthy young and older adults: differential age effects on components of the Capture and Rumination, Functional Avoidance and Impaired Executive Control (CaRFAX) model. *Psychol. Aging* 32, 447–459. doi: 10.1037/pag0000175

Schacter, D. L., and Addis, D. R. (2007). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. T. R. Soc. B.* 362, 773–786. doi: 10.1098/rstb.2007.2087

Schlagman, S., and Kvavilashvili, L. (2008). Involuntary autobiographical memories in and outside the laboratory: how different are they from voluntary autobiographical memories? *Mem. Cognit.* 36, 920–932. doi: 10.3758/mc.36.5.920

Seifert, A., Hofer, M., and Allemand, M. (2018). Mobile data collection: smart, but not (yet) smart enough. *Front. Neurosci.* 12:971. doi: 10.3389/fnins.2018.00971

Söderlund, H., Moscovitch, M., Kumar, N., Daskalakis, Z. J., Flint, A., Herrmann, N., et al. (2014). Autobiographical episodic memory in major depressive disorder. *J. Abnorm. Psychol.* 123, 51–60. doi: 10.1037/a0035610

Spreng, R. N., Lockrow, A. W., DuPre, E., Setton, R., Spreng, K. A. P., and Turner, G. R. (2018). Semanticized autobiographical memory and the default-executive coupling hypothesis of aging. *Neuropsychologia* 110, 37–43. doi: 10.1016/j.neuropsychologia.2017.06.009

St. Jacques, P. L., Conway, M. A., and Cabeza, R. (2011). Gender differences in autobiographical memory for everyday events: retrieval elicited by SenseCam images versus verbal cues. *Memory* 19, 723–732. doi: 10.1080/09658211.2010.516266

St. Jacques, P. L., and Levine, B. (2007). Ageing and autobiographical memory for emotional and neutral events. *Memory* 15, 129–144. doi: 10.1080/09658210601119762

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87, 245–251. doi: 10.1037/0033-2909.87.2.245

Trunk, D. L., and Abrams, L. (2009). Young and older adults' communicative goals influence off-topic speech in autobiographical narratives? *Psychol. Aging* 24, 324–337. doi: 10.1037/a0015259

Turner, G. R., and Spreng, R. N. (2015). Prefrontal engagement and reduced default network suppression co-occur and are dynamically coupled in older adults: the default-executive coupling hypothesis of aging. *J. Cogn. Neurosci.* 27, 2462–2476. doi: 10.1162/jocn_a_00869

Vandermorris, S., Sheldon, S., Winocur, G., and Moscovitch, M. (2013). Differential contributions of executive and episodic memory functions to problem solving in younger and older adults. *J. Int. Neuropsychol. Soc.* 19, 1087–1096. doi: 10.1017/S1355617713000982

Verfaellie, M., Bousquet, K., and Keane, M. M. (2014). Medial temporal and neocortical contributions to remote memory for semantic narratives: evidence from amnesia. *Neuropsychologia* 61, 105–112. doi: 10.1016/j.neuropsychologia.2014.06.018

Wang, Q. (2013). Gender and emotion in everyday event memory. *Memory* 21, 503–511. doi: 10.1080/09658211.2012.743568

Wang, Q., Hou, Y., Tang, H., and Wiprovnick, A. (2011). Travelling backwards and forwards in time: culture and gender in the episodic specificity of past and future events. *Memory* 19, 103–109. doi: 10.1080/09658211.2010.537279

Waters, T. E. A., Bauer, P. J., and Fivush, R. (2014). Autobiographical memory functions served by multiple event types. *Appl. Cogn. Psychol.* 28, 185–195. doi: 10.1002/acp.2976

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition.* San Antonio, TX: Psychological Corporation.

Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Softw.* 21, 1–20. doi: 10.18637/jss.v021.i12

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer-Verlag.

Williams, J. M., and Broadbent, K. (1986). Autobiographical memory in suicide attempters. *J. Abnorm. Psychol.* 95, 144–149. doi: 10.1037/0021-843x.95.2.144

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.

The handling Editor declared a past co-authorship with one of the author MM.

frontiers in Human Neuroscience

METHODS
published: 03 July 2020
doi: 10.3389/fnhum.2020.00261

Functional Analysis of Continuous, High-Resolution Measures in Aging Research: A Demonstration Using Cerebral Oxygenation Data From the Irish Longitudinal Study on Aging

John D. O'Connor[1], Matthew D. L. O'Connell[1,2], Roman Romero-Ortuno[1,3], Belinda Hernández[1], Louise Newman[1], Richard B. Reilly[4], Rose Anne Kenny[1] and Silvin P. Knight[1]*

[1]The Irish Longitudinal Study on Aging, Trinity College, The University of Dublin, Dublin, Ireland, [2]Department of Population Health Sciences, King's College London, London, United Kingdom, [3]The Global Brain Health Institute, Trinity College, The University of Dublin, Dublin, Ireland, [4]Trinity Centre for Biomedical Engineering, Trinity College, The University of Dublin, Dublin, Ireland

OPEN ACCESS

Edited by:
Christina Röcke, University of Zurich, Switzerland

Reviewed by:
Veronica Guadagni, University of Calgary, Canada
Marta Zimmer, Budapest University of Technology and Economics, Hungary

***Correspondence:**
John D. O'Connor
john.oconnor@tcd.ie

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

Received: *18 February 2020*
Accepted: *12 June 2020*
Published: *03 July 2020*

Citation:
O'Connor JD, O'Connell MDL, Romero-Ortuno R, Hernández B, Newman L, Reilly RB, Kenny RA and Knight SP (2020) Functional Analysis of Continuous, High-Resolution Measures in Aging Research: A Demonstration Using Cerebral Oxygenation Data From the Irish Longitudinal Study on Aging. Front. Hum. Neurosci. 14:261. doi: 10.3389/fnhum.2020.00261

Background: A shift towards the dynamic measurement of physiologic resilience and improved technology incorporated into experimental paradigms in aging research is producing high-resolution data. Identifying the most appropriate analysis method for this type of data is a challenge. In this work, the functional principal component analysis (fPCA) was employed to demonstrate a data-driven approach to the analysis of high-resolution data in aging research.

Methods: Cerebral oxygenation during standing was measured in a large cohort [The Irish Longitudinal Study on Aging (TILDA)]. FPCA was performed on tissue saturation index (TSI) data. A regression analysis was then conducted with the functional principal component (fPC) scores as the explanatory variables and transition time as the response.

Results: The mean ± SD age of the analysis sample was 64 ± 8 years. Females made up 54% of the sample and overall, 43% had tertiary education. The first PC explained 96% of the variance in cerebral oxygenation upon standing and was related to a baseline shift. Subsequent components described the recovery to before-stand levels (fPC2), drop magnitude and initial recovery (fPC3 and fPC4) as well as a temporal shift in the location of the minimum TSI value (fPC5). Transition time was associated with components describing the magnitude and timing of the nadir.

Conclusions: Application of fPCA showed utility in reducing a large amount of data to a small number of parameters which summarize the inter-participant variation in TSI upon standing. A demonstration of principal component regression was provided to allow for continued use and development of data-driven approaches to high-resolution data analysis in aging research.

Keywords: data-driven analysis, functional principal component analysis, near infrared spectroscopy, cerebral oxygenation, ageing, orthostatic hypotension

INTRODUCTION

Measures of physiologic resilience are of increasing interest to the field of aging as they can identify an older person's vulnerability to negative outcomes when faced with a stressor (Hadley et al., 2017; Varadhan et al., 2018; Kuchel, 2018). Assessment of resilience requires the tracking of a response to a stressor such as monitoring of the recovery of the cardiovascular system to standing from a supine position (Kim et al., 2011; Romero-Ortuno et al., 2011; McCrory et al., 2016; O'Connell et al., 2018). Improved technology in aging research can record stressor response at high spatial and temporal resolution. While this provides an opportunity to investigate physiological phenomena occurring in shorter timeframes and at smaller scales, it also presents a technical challenge for researchers in identifying the most appropriate methods to analyze the data. Recent publications are showing a growing interest in applying robust analysis methods for collinear, high-resolution data (Odden and Melzer, 2019; Wallace et al., 2019). An example of this type of data is generated in the study of orthostatic hypotension (OH) through monitoring of the neurovascular reaction to standing. Upon standing, both blood pressure and cerebral oxygenation show a rapid decline within 10–20 s after initiation (Finucane et al., 2017; O'Connor et al., 2020). This is followed by rapid recovery, with both signals returning to near baseline by 40 s after standing on average (Finucane et al., 2017; O'Connor et al., 2020). This is generally followed by a slower recovery over the next few minutes (Finucane et al., 2017; O'Connor et al., 2020).

An important risk factor for falls and other health outcomes, OH is assessed by measuring blood pressure recovery after standing (active stand protocol) or passive tilting (head-up tilt test; Kenny et al., 1986; Romero-Ortuno et al., 2013). The consensus definition for OH is a drop in systolic blood pressure (SBP) of $\geq$20 mmHg and/or a drop of $\geq$10 mmHg in diastolic blood pressure (DBP) within 3 min of standing (Kaufmann, 1996; Freeman et al., 2011). Traditionally, this was measured at discrete time points using a sphygmomanometer and more recently acquired continuously at high temporal resolution *via* beat-to-beat monitoring. The continuous monitoring of BP upon standing allows for an individual's response profile to be examined in much greater detail. For example, assessment of the initial reaction to standing can reveal impairment; when SBP drops by $\geq$40 mmHg in the first 15 s this is referred to as initial orthostatic hypotension (Finucane et al., 2014, 2017; McCrory et al., 2016; O'Connell et al., 2018).

The effect of insufficiency in peripheral blood pressure may not translate to cerebral hypoperfusion due to the brain's intrinsic mechanisms for maintaining sufficient cerebral tissue perfusion (i.e., cerebral autoregulation; CAR; Xing et al., 2017). Assessment of dynamic CAR, occurring in response to rapid changes in BP, has been described using Transcranial Doppler Ultrasound (TCD) to quantify cerebral blood flow (CBF) with quick deflation of a leg cuff driving changes in BP (Aaslid et al., 1989). A sit-to-stand maneuver can also be performed to invoke changes in systemic BP which allow the tracking of CAR (van Beek et al., 2008). Whereas TCD measures the CBF in the basal arteries of the brain, near-infrared spectroscopy (NIRS) may be used to measure cerebral oxygenation levels (Colier et al., 1997; Mehagnoul-Schipper et al., 2000; Kawaguchi et al., 2001; Kim et al., 2018; Mol et al., 2019). Although sit-to-stand or supine-to-stand tests are suitable for older adults, the speed of postural transition is heterogeneous and the hemeodynamic response is influenced by standing speed (de Bruïne et al., 2017; Mol et al., 2019; O'Connor et al., 2020). This is not unexpected given that much of the physiological response to standing is driven by muscle activation in the lower limbs (van Wijnen et al., 2017). In addition to the magnitude of central and peripheral hemeodynamic changes following orthostasis, the timing of nadir is of interest. Particularly when the speed of transition has been shown to vary from 2 to 27 s in older adults (O'Connor et al., 2020). Evaluation of CAR with aging may be confounded by the time taken to complete the postural transition. Commercial NIRS devices acquire data at a high temporal resolution, methodologies for analyzing these data have, however, focused on discrete time-points and feature identification with little consideration of timing differences (Briggs et al., 2018, 2019).

Functional principal component analysis (fPCA) is an established method for time-series analysis and data dimensionality reduction which can be used to identify complex trends in data including temporal changes. This work aimed to describe an application of fPCA to assess the association between transition speed during a supine-to-stand challenge and cerebral oxygenation in a large sample of older adults.

METHODS

Cohort

The Irish Longitudinal Study on Aging (TILDA) began data collection in 2009, following a cohort (n = 8,504) of adults aged 50 and over as well as a small number of partners under the age of 50. Data is collected on health, social and financial circumstances (Donoghue et al., 2018). Trinity College Dublin granted ethical approval for the study, the design was compliant with the principles set out in the Declaration of Helsinki and participants provided informed, written consent. This work focuses on data collected from participants who completed TILDA's health assessment center during the third wave of data collection (2014–2015). Since the primary aim of the article was to describe the implementation and interpretation of FPCA, all those with sufficient data were included in the analysis with a preference for adjusting for covariates as opposed to exclusion. Sample characteristics which have previously been associated with cerebral oxygenation such as depression and blood pressure are presented in **Table 1** (Lucas et al., 2010; Briggs et al., 2019).

Active Stand Protocol

Cerebral oxygenation measurements in TILDA were acquired in a quiet room with a controlled temperature of between 21°C and 23°C. During the active stand, a NIRS device (Portalite; Artinis Medical Systems, Zetten, Netherlands) was fixed to the forehead (3 cm lateral and 3.5 cm superior to the nasion) in approximately in the FP1 position of the ten-twenty electrode system (Klem et al., 1999). The Oxysoft (V3.0.53, Artemis Medical Systems, Zetten, Netherlands) software facilitated data

TABLE 1 | Characteristics of the sample.

Age (years)	64.2 ± 7.9 [39–93]			-
Sex	Male: 45.8% (1,261)	Female: 54.3% (1,495)		-
Highest education achieved	Primary/none: 16.2% (445)	Secondary: 40.4% (1,113)	Third/higher: 43.5% (1,198)	-
Hypertensive	Yes: 35.3% (971)	No: 64.7% (1,781)		Missing: 4
Taking any anti-hypertensives (ATC C02, C03, C07, C08, C09)	Yes: 36.8% (1,015)	No: 63.2% (1,741)		-
CES-D (Centre for Epidemiological Studies Depression scale)	2.9 ± 3.5			Missing: 10
Taking at least one anti-depressant (ATC N06A)	Yes: 7.4% (203)	No: 92.6% (2,553)		-
Orthostatic hypotension	Yes: 13.1% (361)	No: 86.9% (2,395)		-
Smoking history	Current: 9.7% (266)	Past: 42.5% (1,172)	Never: 47.8% (1,318)	-
Body mass index	28.3 ± 4.7			Missing: 3

Orthostatic hypotension was defined as those who had a blood pressure deficit of ≥20 mmhg systolic or ≥10 mmhg diastolic at 40 s after stand given that a previous study shows that most people have recovered by then (Finucane et al., 2017).

collection and allowed manual signal marking at the beginning of a rest period. A digital photoplethysmography device (Finometer MIDI device, Finapres Medical Systems BV, Amsterdam, The Netherlands) was attached to each participant and used for measuring continuous blood pressure and identifying the timing of postural transitions *via* a built-in height correction unit. Data were acquired continuously while participants lay supine for 10 min before transitioning to a standing position and remaining standing for a further 3 min. The sampling frequency of the NIRS device was 50 Hz.

Signal Processing and Analysis

Data from the NIRS files were extracted and processed using MATLAB (R2018a, TheMathWorks, Inc., MA, USA). The device recorded both HbO_2 (μM) and HHb (μM), from which the summary measure of tissue saturation index (TSI) was calculated using Equation 1.

$$TSI = \frac{HbO_2 \times 100}{HbO_2 + HHb} \quad (1)$$

Participants were removed from the analysis for missing/erroneous height sensor data or rest markers, as the time of their stand could not be reliably identified. After visual inspection of the signals, exclusions were applied for data suspected to be a measurement error—TSI falling below 10%, absolute HbO_2 or HHb of less than 0.1 μM in over a quarter of data points, TSI change of over 45% overall, TSI change of less than 0.1% over first 30 s of stand and participants with a variation of more than 10% in the TSI baseline. Based on the manufacturer's manual, those with an average fit factor (a measure of agreement between sensors on the NIRS device) of less than 98% were also removed. The time when the stand occurred was identified from the height sensor data, and the NIRS response profiles were subsequently aligned such that the zero second time-point corresponded to the point where the participant had started standing. Data was retained from 1 min before stand to 3 min after meaning each participant's signal had 12,001 samples (i.e., 4 min of 50 Hz data).

Functional Principal Component Analysis

Overview and Motivation

The fPCA was performed using a freely available MATLAB toolbox (Ramsay and Silverman, 2005; Ramsay et al., 2009). The purpose of applying fPCA, in this context, was to describe each participant's response using a series of independent functions (i.e., functional principal components). These are defined by a functional loading across time for the NIRS recording which describes the influence of the signal at that time on the component. The functional principal component (fPC) score can then be calculated and is an estimation of the contribution of a component to an individual's response. These scores may be standardized to Z-scores for interpretability where, for instance, a participant may have a score of +2SD on a specific fPC meaning their response was +2SD above the mean response for that component. This can then be interpreted graphically by plotting the mean ± 2SD of each of the components and is referred to in previous work as single component reconstruction (Brandon et al., 2013). To aid interpretation of the principal components, raw data were averaged over 10 s periods and a correlation matrix with the scores was produced. This enables the reader to assess which regions of the response are correlated with the components. Along with the graphical interpretation, scores can also be leveraged for statistical inference by using them as independent variables in regression (i.e., principal component regression).

Basis Functions, Knots and Smoothing

The first step of performing fPCA is fitting functions to the data. These functions are commonly defined by a linear combination of basis functions (e.g., Fourier, B-spline or wavelet bases). The choice of basis is dependent on the data input, Fourier bases are often used for periodic variation (e.g., seasonal weather changes or biomechanical phenomena repeating over a gait cycle) whereas B-splines may be used to model non-periodic processes such as the active stand (Ramsay and Silverman, 2005; Donoghue et al., 2008). The location of knots for B-splines is user-defined, they can be placed uniformly over the measuring window or located more densely at areas of the signal expected to show more variation (e.g., shortly after/before the stand in the data here). The smoothness of the function is often a subjective choice based on visual inspection of the fit of the function to individual participant data. Automated methods for choosing a smoothing parameter, such as cross-validation, exist but are recommended only for initial guidance (Ramsay and Silverman, 2002).

The NIRS data from the active stand is non-periodic and therefore B-splines were chosen as a suitable set of basis functions to describe the data. Knots were placed at 2-s intervals from 10 s before the stand to 40 s after with additional knots

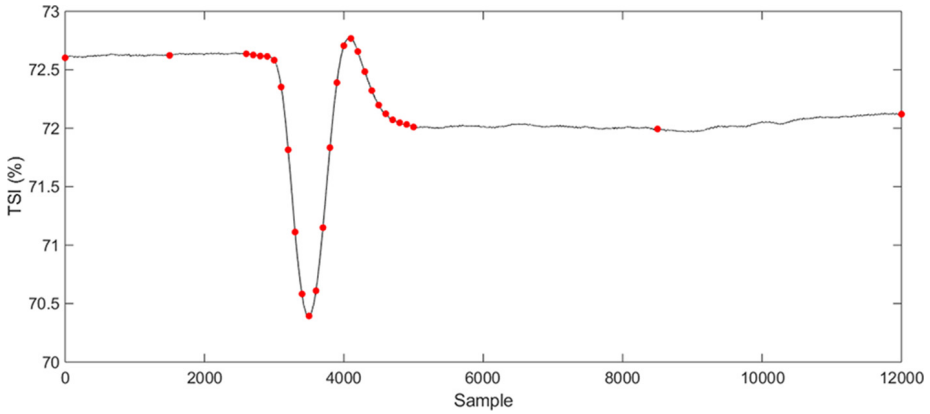


FIGURE 1 | Placement of knots located preferentially around the area of the signal expected to show variation based on the mean response and previous works.

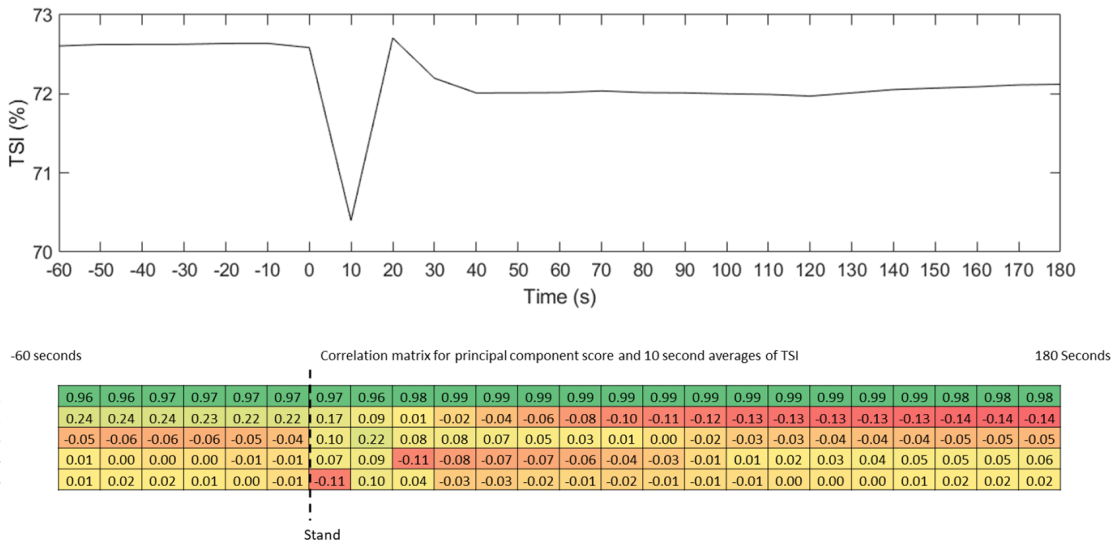


FIGURE 2 | Correlation between 10-s averages of original tissue saturation index (TSI) data and principal component scores.

located at the beginning/end, 30 s before stand and 110 s after (**Figure 1**). This was based on the shape of the mean response and previous literature on cardiovascular responses to standing (Finucane et al., 2014; O'Connell et al., 2018). A smoothing parameter of 1e-8 was chosen based on the appropriate fit to the data and generalized cross-validation results (Ramsay and Silverman, 2005).

Functional Principal Components Regression

Regression analyses may be performed using the scores calculated from fPCA. It is important to decipher how many principal components to retain for analyses of this type. Common methods for achieving this are the scree plot (a graphical representation of the cumulative variance explained by fPCs), choosing components describing a defined proportion of the variance (e.g., 90%) or generalized cross-validation. Transition time (i.e., the amount of time taken to assume a standing position from supine) was used as an outcome variable in this demonstration and linear regression was performed with five fPCs as the explanatory variables. Covariates included in the model were: age, sex, use of antihypertensive or antidepressant medication, BMI and resting mean arterial pressure (MAP). The number of fPCs to use was chosen here by identification of the "elbow" in the scree plot.

In addition to analysis of data from TILDA, a demonstration is provided in **Appendix 1** using simulated data to allow for easier implementation by other researchers (i.e., code may be used by replacing simulated data with experimental data). A MATLAB script for generating and analyzing the data is provided. The example of stand data includes a simulated trough magnitude as well as hypothetical timing of minimum, start and end of the

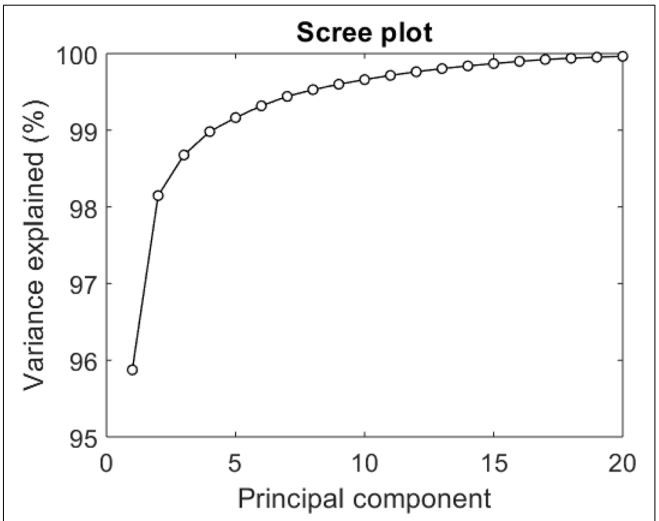


FIGURE 3 | Scree plot for assessing cumulative variance explained by additional functional principal components (fPCs).

nadir. A scree plot and single component reconstructions are presented for the simulated data (**Appendix 1**, **Figures 5** and **6**). A principal component regression (using the fPC scores in linear regression) is also performed for demonstration purposes on the simulated data with a theoretical dependent variable that is related to the simulated trough features.

RESULTS

A total of 2,756 participants' NIRS data were available for analysis (further detail provided in **Appendix 2**, **Figure 7**). Sample characteristics are detailed in **Table 1**. The mean (SD) age of the analysis sample was 64 (8) years. Females made up 54% and overall 43% had tertiary education. Levels of hypertension and antihypertensive use were similar at 35% and 37% respectively. The sample had a mean (SD) Centre for Epidemiological Studies Depression scale score of 2.9 (3.5) with 7% taking at least one antidepressant medication. Past smokers constituted 43% of the sample compared to 10% current and 48% never smokers. Body mass index was, on average, in the overweight range at 28 (5) kg/m^2.

The mean TSI response for the full sample according to the 10 s averaged data consists of a sharp drop on standing followed by an initial recovery and stabilization throughout the rest of the stand (**Figure 2**). The correlation between the original data and the fPC scores is also shown in **Figure 2** (this is analogous to loadings). Similar to previous studies, the "elbow" of the scree plot (**Figure 3**) was visually identified and 5 modes were chosen for retention (Woods and Edwards, 2007; Wilcox, 2012). The first principal component explained the majority of the variation in the TSI signal throughout the 4 min of rest and standing, accounting for 95.9% of the variation (**Figure 3**). As expected, the variance explained increased with the inclusion of more principal components with five modes explaining 99.2%.

The first component describes a baseline difference which shifts the entire trace up or down, this is evident in the component reconstruction for fPC1 (**Figure 4**) and the high correlations with all 10-s averages (**Figure 2**). The second principal component may be interpreted to represent the offset between baseline and recovery, i.e., how far above or below baseline the recovery stabilizes (**Figure 4**). This interpretation is supported by the changing sign in the correlation matrix during the transition from drop to recovery (**Figure 2**) as well as the difference between the baseline and final value on the single component reconstruction for fPC2 (**Figure 4**). Graphically, the third and fourth mode described the depth of the nadir/peak and the initial recovery (**Figures 2**, **4**). Scores for the third, fourth and fifth components for TSI were linked to the difference between the nadir depth at around 10 s and initial overshoot at around 20 s as well as the temporal shift in where the minimum TSI occurred (**Figures 2**, **4**). In particular, a low score for principal component five appeared to indicate a shifted minimum TSI value i.e., minimum TSI occurred after a longer duration following the transition to stand (**Figures 2**, **4**). Linear regression analysis of the fPCs with the response variable of transition time showed associations between all but the first component (**Table 2**). The largest regression coefficients were for fPC4 and fPC5, suggesting that transition time is linked to the timing and depth of the nadir. When covariates were added to the model, the adjusted R^2 value increased from 0.05 to 0.12 suggesting some of the variances in standing speed is driven by participant factors (**Table 3**). However, there remained significant relationships between the FPC4 and FPC5 and standing speed after adjusting for these factors (**Table 3**). Standing speed was associated with age, sex, BMI and antidepressant medication (**Table 3**).

The first three components, resulting from fPCA on the simulated data (**Figure 5** in **Appendix 1**), characterized the simulated features of baseline, trough depth and timing. The scree plot also shows that three components are the appropriate choice of fPCs to retain using the "elbow" identification method. Regression analysis showed significant relationships between the scores and simulated features.

DISCUSSION

The purpose of this work was to demonstrate a data-driven approach to the analysis of high resolution and collinear data as generated from dynamic physiological tasks in older adults. This was achieved using cerebral oxygenation data from an active stand protocol and fPCA.

The number of parameters needed to characterize an individual's TSI response profile was reduced from several thousand time-points to five fPC scores which explained a large proportion of variance in the sample. The study demonstrates how component scores can be interpreted for use in further statistical analysis. In particular, the main interests of the active stand protocol were well characterized by the fPCs. Other aspects of the response such as the depth of the initial drop or the timing of the peak

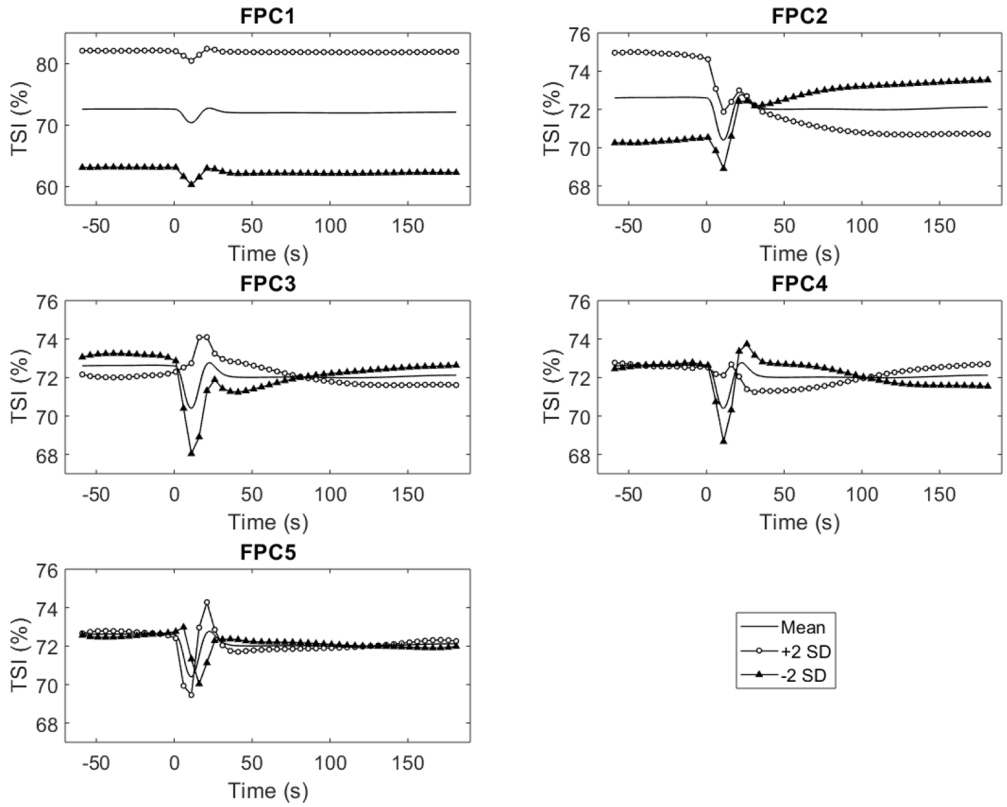


FIGURE 4 | Graphical representations of the first five principal components for TSI (±2 standard deviations).

TABLE 2 | Results from linear regression with transition time as the dependent variable and functional principal components as the explanatory variables.

Transition time(s)	Coefficient	*t*-statistic	*P*-value	95% Confidence interval	
fPC1	−0.10	−1.90	0.058	−0.206	0.003
fPC2	0.11	1.98	**0.048**	0.001	0.210
fPC3	0.11	1.99	**0.047**	0.001	0.211
fPC4	0.28	5.30	**<0.001**	0.179	0.388
fPC5	−0.54	−10.13	**<0.001**	−0.647	−0.437
Intercept	7.26	135.91	**<0.001**	7.164	7.37

Adjusted R^2 = 0.05, p-values < 0.05.

TABLE 3 | Results from linear regression with principal component scores and covariates of age, sex, body mass index (BMI), antidepressants, antihypertensives and mean arterial pressure (MAP).

Transition time (s)	Coefficient	*t*-statistic	*P*-value	95% Confidence interval	
fPC1	−0.07	−1.36	0.173	−0.173	0.031
fPC2	0.04	0.81	0.420	−0.061	0.146
fPC3	0.08	1.44	0.151	−0.029	0.186
fPC4	0.17	3.16	**0.002**	0.065	0.276
fPC5	−0.47	−8.87	**<0.001**	−0.570	−0.363
Age	0.06	8.43	**<0.001**	0.045	0.072
Sex	0.63	5.37	**<0.001**	0.403	0.866
BMI	0.12	10.17	**<0.001**	0.095	0.140
Antidepressants	0.48	2.42	**0.016**	0.092	0.877
Antihypertensives	0.20	1.71	0.088	−0.029	0.429
MAP	−0.01	−1.46	0.144	−0.015	0.002
Intercept	−0.26	−0.35	0.723	−1.680	1.166

Adjusted R^2 = 0.12, p-values < 0.05.

drop were also well described by subsequent components. Transition time demonstrated an association with several fPC scores related to the timing and magnitude of the nadir after standing. This is notable as identification of timing differences may be difficult to detect with traditional analysis methods.

The study of the orthostatic response to standing from supine is of interest in older adults given its relationship with falls, depression and mortality (McCrory et al., 2016; Finucane et al., 2017; Briggs et al., 2019). Given the range in older adult mobility, the speed at which the transition is completed is an important source of variation that is associated with both timing and magnitude of hemeodynamic parameters (de Bruïne et al., 2017; Mol et al., 2019; O'Connor et al., 2020). The musculoskeletal effort of standing causes vasodilation in the lower limbs (as evidenced by larger changes during active vs. passive stand; van Wijnen et al., 2017). A faster transition may lead to quicker dilation of the vasculature and has shown larger changes from baseline in TSI and blood pressure (O'Connor et al., 2020). A slower transition can lead to a delayed nadir in orthostatic measures and the identification of temporal differences (FPC4 and FPC5) in this work are important for the interpretation of previous studies which assume minimum values occur at the same time after stand (e.g., 10 s after stand). The time taken to transition is related to aging, sex, BMI, and antidepressant medication use; therefore, adjustments should be made for analysis of cerebrovascular function in aging using supine-to-stand tests.

The data sampling rate has been increasing in many areas of aging research including during continuous monitoring of blood pressure and cerebral oxygenation (Kawaguchi et al., 2001; Briggs et al., 2018, 2019; O'Connell et al., 2018). Commonly, data reduction is achieved *via* time point averaging (e.g., 10-s averages), however, given the relatively high sampling frequency capabilities of commercially available devices this leads to most of the collected data being discarded. As such, complex trends may be difficult, if not impossible to identify with this type of analysis; fPCA can reveal such trends in the data (e.g., temporal trends such as shifting peak/drop associated with fPC5). Principal component regression also addresses the issue of collinearity, with the ability to provide a small number of continuous independent parameters which characterize the entire response profile. The analysis pipeline demonstrated in this work may prove useful to researchers wishing to investigate links between high-resolution, collinear data collected in research into aging and outcome measures. This type of analysis, with additional development, may ultimately prove useful in a clinical setting for automated analysis of NIRS data, identifying subtle trends in the active stand response which may be associated with poor or desirable outcomes. Starting to address limitations in modeling methods for physiologic resilience tests can facilitate the development of improved prediction for important outcomes such as falls, cognition or mortality (Wallace et al., 2019).

A large proportion of the variance between individuals was explained by the first component (96%) suggesting that the differences in baseline dominated. This is similar to trends in other hemeodynamic measures during standing such as heart rate and blood pressure. However, the proportion of variance explained by the baseline shift is larger in the case of TSI given the stand elicits an attenuated reaction compared to that of BP (i.e., TSI changes of 1–2% whereas BP changes by around 25%; O'Connor et al., 2020). This smaller change of TSI is expected given the action of cerebral autoregulation and changes of this magnitude have shown an association with outcomes, such as depression, in TILDA (Briggs et al., 2019). The remaining components (FPC2–FPC5) cover a small proportion of the variance (3.2%) but represent the changes in cerebral oxygenation brought about by standing. These components were useful in predicting standing time in the study and previous works have shown that components with low variance explained can still be crucially important (Jolliffe, 1982).

Notwithstanding these advantages, the analysis of NIRS data in this manner does have some limitations. Firstly, the selection of the number of components to use to adequately represent the data can be challenging, although objective criteria do exist to assist in making this selection (Jolliffe, 1982). Additional methods such as cross-validation or parallel analysis may be required for robust identification of the number of components needed for identifying relationships with specific outcomes. Secondly, the direct effect of NIRS measurements at a specific time-point on the dependent variable may be difficult to decipher given that several components (which are influenced by the same time point) can be related to the outcome measure simultaneously. Clinicians may find it difficult to identify modes of variation in the data by eye; however, this can be facilitated by the development of a software package to fit the fPCA model described herein to a particular individual's measured response profile. The principal components presented here are generated from those who attended TILDA's health center and had valid data for NIRS. To ensure population representativeness, considering sample loss and non-attendance, weighting should be considered when the method is applied to specific outcome measures. Systems using NIRS can produce reliable estimates of CAR during orthostasis and have shown utility in clinical studies using head-up tilt testing (Bachus et al., 2018; Mol et al., 2020). There is a high correlation between CAR assessed using either TCD or NIRS (Zweifel et al., 2010). The NIRS system in this study has an estimated penetration depth of 20 mm into the prefrontal cortex, future work assessing the correlation between oxygenation in this area and other regions of the brain is of interest.

CONCLUSION

In conclusion, fPC regression shows utility in identifying independent parameters which characterize the cerebral oxygenation reaction to standing. A demonstration of principal component regression was provided to allow for continued use and development of data-driven approaches to high-resolution data analysis in aging research.

DATA AVAILABILITY STATEMENT

The datasets for this study are available upon reasonable request to The Irish Longitudinal Study on Ageing (email: tilda@tcd.ie)

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Trinity College Dublin. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JO'C: drafting manuscript. JO'C, SK, and LN: data analysis. LN and RK: data acquisition. JO'C, MO'C, BH, LN, RR-O, RR, RK, and SK: study concept, interpretation of data, critical revision of the article for important intellectual content, and final approval for submission.

FUNDING

The Irish longitudinal study on aging (TILDA) is funded by Irish Life, The Atlantic Philanthropies and the Department of Health in Ireland. This work was further supported by an Aging Research Leadership Fellowship awarded to MO'C from the Centre for Aging Research and Development in Ireland, which became the Aging Research and Development Division within the Institute of Public Health in Ireland (IPH) in September 2015, sponsored by the American Federation for Aging Research Paul B. Beeson Career Development Awards in Aging Research for the Island of Ireland. RR-O is funded by Science Foundation Ireland grant number: 18/FRL/6188.

ACKNOWLEDGMENTS

We would like to acknowledge the continued commitment and cooperation of the TILDA participants and research team.

REFERENCES

Aaslid, R., Lindegaard, K. F., Sorteberg, W., and Nornes, H. (1989). Cerebral autoregulation dynamics in humans. *Stroke* 20, 45–52. doi: 10.1161/01.str.20.1.45

Bachus, E., Holm, H., Hamrefors, V., Melander, O., Sutton, R., Magnusson, M., et al. (2018). Monitoring of cerebral oximetry during head-up tilt test in adults with history of syncope and orthostatic intolerance. *Europace* 20, 1535–1542. doi: 10.1093/europace/eux298

Brandon, S. C. E., Graham, R. B., Almosnino, S., Sadler, E. M., Sadler, E. M., and Stevenson, J. M. (2013). Interpreting principal components in biomechanics: representative extremes and single component reconstruction. *J. Electromyogr. Kinesiol.* 23, 1304–1310. doi: 10.1016/j.jelekin.2013.09.010

Briggs, R., Carey, D., Claffey, P., McNicholas, T., Newman, L., Nolan, H., et al. (2019). The association between frontal lobe perfusion and depressive symptoms in later life. *Br. J. Psychiatry* 214, 230–236. doi: 10.1192/bjp.2018.288

Briggs, R., Carey, D., Kennelly, S. P., and Kenny, R. A. (2018). Longitudinal association between orthostatic hypotension at 30 seconds post-standing and late-life depression. *Hypertension* 71, 946–954. doi: 10.1161/hypertensionaha.117.10542

Colier, W. N., Binkhorst, R. A., Hopman, M. T., and Oeseburg, B. (1997). Cerebral and circulatory haemodynamics before vasovagal syncope induced by orthostatic stress. *Clin. Physiol.* 17, 83–94. doi: 10.1046/j.1365-2281.1997.01414.x

de Bruïne, E. S., Reijnierse, E. M., Trappenburg, M. C., Pasma, J. H., de Vries, O. J., Meskers, C. G. M., et al. (2017). Standing up slowly antagonises initial blood pressure decrease in older adults with orthostatic hypotension. *Gerontology* 63, 137–143. doi: 10.1159/000450642

Donoghue, O. A., Harrison, A. J., Coffey, N., and Hayes, K. (2008). Functional data analysis of running kinematics in chronic achilles tendon injury. *Med. Sci. Sports Exerc.* 40, 1323–1335. doi: 10.1249/mss.0b013e31816c4807

Donoghue, O. A., McGarrigle, C. A., Foley, M., Fagan, A., Meaney, J., and Kenny, R. A. (2018). Cohort profile update: the irish longitudinal study on aging (TILDA). *Int. J. Epidemiol.* 47, 1398–1398l. doi: 10.1093/ije/dyy163

Finucane, C., O'Connell, M. D. L., Donoghue, O., Richardson, K., Savva, G. M., and Kenny, R. A. (2017). Impaired orthostatic blood pressure recovery is associated with unexplained and injurious falls. *J. Am. Geriatr. Soc.* 65, 474–482. doi: 10.1111/jgs.14563

Finucane, C., O'Connell, M. D. L., Fan, C. W., Savva, G. M., Soraghan, C. J., Nolan, H., et al. (2014). Age-related normative changes in phasic orthostatic blood pressure in a large population study: findings from the Irish longitudinal study on aging (TILDA). *Circulation* 130, 1780–1789. doi: 10.1161/circulationaha.114.009831

Finucane, C., Savva, G. M., and Kenny, R. A. (2017). Reliability of orthostatic beat-to-beat blood pressure tests: implications for population and clinical studies. *Clin. Auton. Res.* 27, 31–39. doi: 10.1007/s10286-016-0393-3

Freeman, R., Wieling, W., Axelrod, F. B., Benditt, D. G., Benarroch, E., Biaggioni, I., et al. (2011). Consensus statement on the definition of orthostatic hypotension, neurally mediated syncope and the postural tachycardia syndrome. *Clin. Auton. Res.* 21, 69–72. doi: 10.1007/s10286-011-0119-5

Hadley, E. C., Kuchel, G. A., Newman, A. B., Allore, H. G., Bartley, J. M., Bergeman, C. S., et al. (2017). Report: NIA workshop on measures of physiologic resiliencies in human aging. *J. Gerontol. A Biol. Sci. Med. Sci.* 72, 980–990. doi: 10.1093/gerona/glx015

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Appl. Stat.* 31, 300–333. doi: 10.2307/2348005

Kaufmann, H. (1996). Consensus statement on the definition of orthostatic hypotension, pure autonomic failure and multiple system atrophy. *Clin. Auton. Res.* 6, 125–126. doi: 10.1007/BF02291236

Kawaguchi, T., Uyama, O., Konishi, M., Nishiyama, T., and Iida, T. (2001). Orthostatic hypotension in elderly persons during passive standing: a comparison with young persons. *J. Gerontol. A Biol. Sci. Med. Sci.* 56, 273–280. doi: 10.1093/gerona/56.5.m273

Kenny, R. A., Ingram, A., Bayliss, J., and Sutton, R. (1986). Head-up tilt: a useful test for investigating unexplained syncope. *Lancet* 327, 1352–1355. doi: 10.1016/s0140-6736(86)91665-x

Kim, Y.-S., Bogert, L. W. J., Immink, R. V., Harms, M. P. M., Colier, W. N. J., Van Lieshout, J. J., et al. (2011). Effects of aging on the cerebrovascular orthostatic response. *Neurobiol. Aging* 32, 344–353. doi: 10.1016/j.neurobiolaging.2009.02.019

Kim, J. M., Choi, J. K., Choi, M., Ji, M., Hwang, G., Ko, S. B., et al. (2018). Assessment of cerebral autoregulation using continuous-wave near-infrared spectroscopy during squat-stand maneuvers in subjects with symptoms of orthostatic intolerance. *Sci. Rep.* 8:13257. doi: 10.1038/s41598-018-31685-y

Klem, G. H., Lüders, H. O., Jasper, H. H., and Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 3–6.

Kuchel, G. A. (2018). Frailty and resilience as outcome measures in clinical trials and geriatric care: are we getting any closer? *J. Am. Geriatr. Soc.* 66, 1451–1454. doi: 10.1111/jgs.15441

Lucas, S. J. E., Tzeng, Y. C., Galvin, S. D., Thomas, K. N., Ogoh, S., and Ainslie, P. N. (2010). Influence of changes in blood pressure

on cerebral perfusion and oxygenation. *Hypertension* 55, 698–705. doi: 10.1161/HYPERTENSIONAHA.109.146290

McCrory, C., Berkman, L. F., Nolan, H., O'Leary, N., Foley, M., and Kenny, R. A. (2016). Speed of heart rate recovery in response to orthostatic challenge. *Circ. Res.* 119, 666–675. doi: 10.1161/circresaha.116.308577

Mehagnoul-Schipper, D. J., Vloet, L. C. M., Colier, W. N., Hoefnagels, W. H., and Jansen, R. W. (2000). Cerebral oxygenation declines in healthy elderly subjects in response to assuming the upright position. *Stroke* 31, 1615–1620. doi: 10.1161/01.str.31.7.1615

Mol, A., Maier, A. B., van Wezel, R. J. A., and Meskers, C. G. M. (2020). Multimodal monitoring of cardiovascular responses to postural changes. *Front. Physiol.* 11:168. doi: 10.3389/fphys.2020.00168

Mol, A., Woltering, J. H. H., Colier, W. N. J. M., Maier, A. B., Meskers, C. G. M., and van Wezel, R. J. A. (2019). Sensitivity and reliability of cerebral oxygenation responses to postural changes measured with near-infrared spectroscopy. *Eur. J. Appl. Physiol.* 119, 1117–1125. doi: 10.1007/s00421-019-04101-0

O'Connell, M. D., Savva, G. M., Finucane, C., Romero-Ortuno, R., Fan, C. W., and Kenny, R. A. (2018). Impairments in hemodynamic responses to orthostasis associated with frailty: results from the irish longitudinal study on aging (TILDA). *J. Am. Geriatr. Soc.* 66, 1475–1483. doi: 10.1111/jgs.15327

O'Connor, J. D., O'Connell, M. D. L., Nolan, H., Newman, L., Knight, S. P., and Kenny, R. A. (2020). Impact of standing speed on the peripheral and central hemodynamic response to orthostasis. *Hypertension* 75, 524–531. doi: 10.1161/hypertensionaha.119.14040

Odden, M. C., and Melzer, D. (2019). Machine learning in aging research. *J. Gerontol. A Biol. Sci. Med. Sci.* 74, 1901–1902. doi: 10.1093/gerona/glz074

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. New York, NY: Springer.

Ramsay, J. O., and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. New York, NY, Springer.

Ramsay, J. O., and Silverman, B. W. (2005). *Functional Data Analysis*. 2nd Edn. New Jersey, NJ: Springer Verlag.

Romero-Ortuno, R., Cogan, L., O'shea, D., Lawlor, B. A., and Kenny, R. A. (2011). Orthostatic haemodynamics may be impaired in frailty. *Age Aging* 40, 576–583. doi: 10.1093/Aging/afr076

Romero-Ortuno, R., O'Connell, M. D., Finucane, C., Soraghan, C., Fan, C. W., and Kenny, R. A. (2013). Insights into the clinical management of the syndrome of supine hypertension—orthostatic hypotension (SH-OH): the irish longitudinal study on aging (TILDA). *BMC Geriatr.* 13:73. doi: 10.1186/1471-2318-13-73

van Beek, A. H., Claassen, J. A., Rikkert, M. G., and Jansen, R. W. (2008). Cerebral autoregulation: an overview of current concepts and methodology with special focus on the elderly. *J. Cereb. Blood Flow Metab.* 28, 1071–1085. doi: 10.1038/jcbfm.2008.13

van Wijnen, V. K., Finucane, C., Harms, M. P. M., Nolan, H., Freeman, R. L., Westerhof, B. E., et al. (2017). Noninvasive beat-to-beat finger arterial pressure monitoring during orthostasis: a comprehensive review of normal and abnormal responses at different ages. *J. Intern. Med.* 282, 468–483. doi: 10.1111/joim.12636

Varadhan, R., Walston, J. D., and Bandeen-Roche, K.. (2018). Can a link be found between physical resilience and frailty in older adults by studying dynamical systems? *J. Am. Geriatr. Soc.* 66, 1455–1458. doi: 10.1111/jgs.15409

Wallace, M. L., Buysse, D. J., Redline, S., Stone, K. L., Ensrud, K., Leng, Y., et al. (2019). Multidimensional sleep and mortality in older adults: a machine-learning comparison with other risk factors. *J. Gerontol. A Biol. Sci. Med. Sci.* 74, 1903–1909. doi: 10.1093/gerona/glz044

Wilcox, R. (2012). "Some Multivariate Methods. Introd. to Robust Estim," in *Hypothesis Test*, (New York, NY: Elsevier), 215–289.

Woods, C. M., and Edwards, M. C. (2007). Factor analysis and related methods. *Handbook of Statistics 27. The Netherlands*, eds C. R. Rao, J. P. Miller and D. C. Rao (Elsevier: BV Japan Journal of Nursing Science), 367–394.

Xing, C.-Y., Tarumi, T., Meijers, R. L., Turner, M., Repshas, J., Xiong, L., et al. (2017). Arterial pressure, heart rate, and cerebral hemodynamics across the adult life span. *Hypertension* 69, 712–720. doi: 10.1161/hypertensionaha.116.08986

Zweifel, C., Castellani, G., Czosnyka, M., Carrera, E., Brady, K. M., Kirkpatrick, P. J., et al. (2010). Continuous assessment of cerebral autoregulation with near-infrared spectroscopy in adults after subarachnoid hemorrhage. *Stroke* 41, 1963–1968. doi: 10.1161/STROKEAHA.109.577320

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.

APPENDIX

Appendix 1

This section includes a Github link (https://github.com/JohnDOConnor/TILDA_bioengineering) to code for performing principal component regression. Simulated data is used here for demonstration. Each simulated signal has a random trough time, magnitude and baseline. The component reconstructions demonstrate the ability to identify these trends in the first 3 components, the scree point also shows little benefit from including more than 3 components.

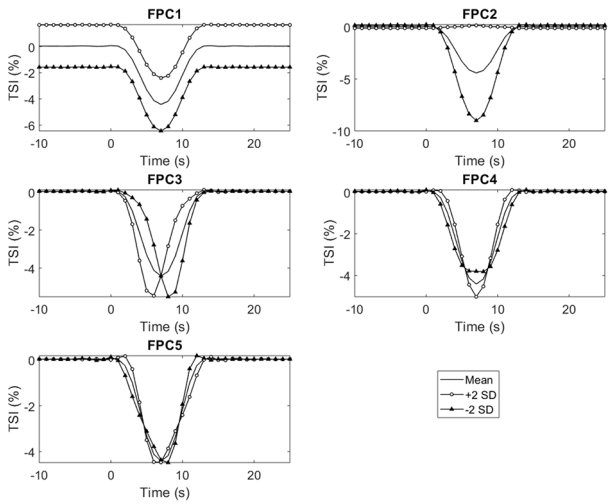


FIGURE 5 | Single component reconstructions for the first five functional principal modes from simulated data.

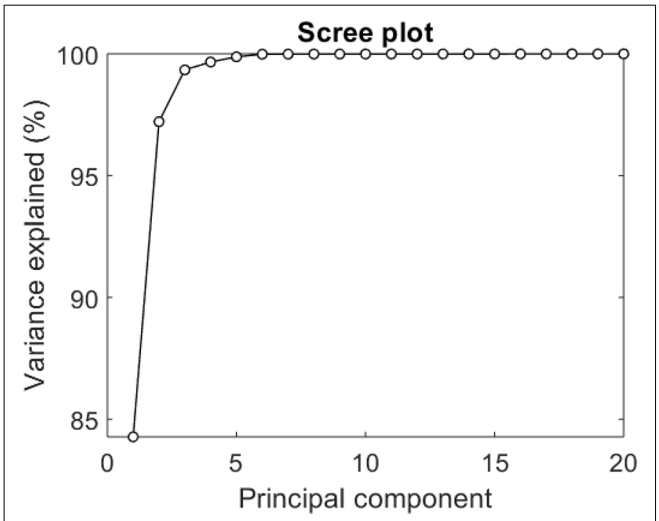


FIGURE 6 | Scree plot showing the cumulative variance explained by each additional principal component in the simulated data.

Appendix 2

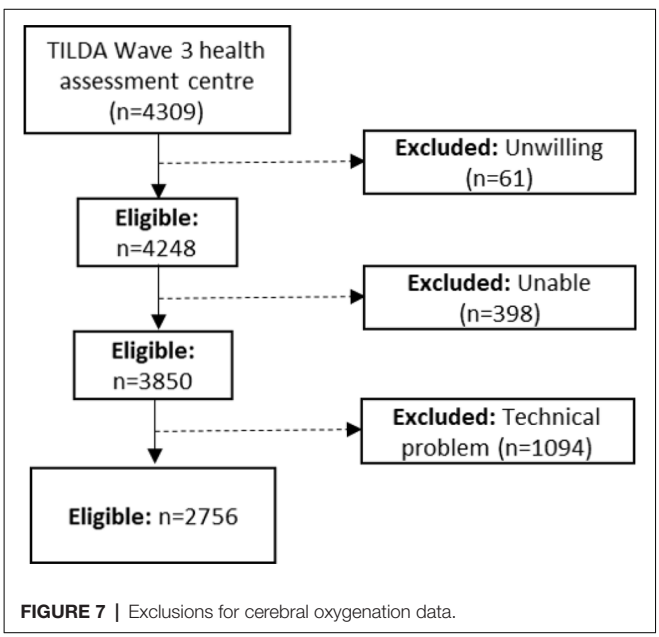


FIGURE 7 | Exclusions for cerebral oxygenation data.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 04 September 2020
doi: 10.3389/fnhum.2020.00363

Decline Variability of Cortical and Subcortical Regions in Aging: A Longitudinal Study

Silvano Sele[1,2], Franziskus Liem[2], Susan Mérillat[2] and Lutz Jäncke[1,2]**

[1] Division Neuropsychology, Department of Psychology, University of Zurich, Zurich, Switzerland, [2] University Research Priority Program (URPP), "Dynamics of Healthy Aging", University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Claudia Buss, Charité – Universitätsmedizin Berlin, Germany

Reviewed by:
Jerod Rasmussen, University of California, Irvine, United States
Gerard Nisal Bischof, Helmholtz Association of German Research Centres (HZ), Germany

***Correspondence:**
Silvano Sele
silvano.sele@uzh.ch
Lutz Jäncke
lutz.jaencke@uzh.ch;
l.jaencke@psychologie.uzh.ch

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

Received: *03 April 2020*
Accepted: *10 August 2020*
Published: *04 September 2020*

Citation:
Sele S, Liem F, Mérillat S and Jäncke L (2020) Decline Variability of Cortical and Subcortical Regions in Aging: A Longitudinal Study. Front. Hum. Neurosci. 14:363. doi: 10.3389/fnhum.2020.00363

Describing the trajectories of age-related change for different brain structures has been of interest in many recent studies. However, our knowledge regarding these trajectories and their associations is still limited due to small sample sizes and low numbers of repeated measures. For the present study, we used a large longitudinal dataset (four measurements over 4 years) comprising anatomical data from a sample of healthy older adults (N = 231 at baseline). This dataset enables us to gain new insights about volumetric cortical and subcortical changes and their associations in the context of healthy aging. Brain structure volumes were derived from T1-weighted MRI scans using FreeSurfer segmentation tools. Brain structure trajectories were fitted using mixed models and latent growth curve models to gain information about the mean extent and variability of decline trajectories for different brain structures as well as the associations between individual trajectories. On the group level, our analyses indicate similar linear changes for frontal and parietal brain regions, while medial temporal regions showed an accelerated decline with advancing age. Regarding subcortical regions, some structures showed strong declines (e.g., hippocampus), others showed little decline (e.g., pallidum). Our data provide little evidence for sex differences regarding the aforementioned trajectories. Between-person variability of the person-specific slopes (random slopes) was largest in subcortical and medial temporal brain structures. When looking at the associations between the random slopes from each brain structure, we found that the decline is largely homogenous across the majority of cortical brain structures. In subcortical and medial temporal brain structures, however, more heterogeneity of the decline was observed, meaning that the extent of the decline in one structure is less predictive of the decline in another structure. Taken together, our study contributes to enhancing our understanding of structural brain aging by demonstrating (1) that average volumetric change differs across the brain and (2) that there are regional differences with respect to between-person variability in the slopes. Moreover, our data

suggest (3) that random slopes are highly correlated across large parts of the cerebral cortex but (4) that some brain regions (i.e., medial temporal regions) deviate from this homogeneity.

Keywords: aging, structural MRI, latent growth curve model, longitudinal data analysis, cortical and subcortical brain structure trajectories

INTRODUCTION

Growing old is accompanied by physical and cognitive declines. It is of great importance to know how these declines can be kept at a minimum so that the quality of life can remain high. By using brain imaging techniques like magnetic resonance imaging (MRI), one can observe that the brain does undergo pronounced changes with aging (Oschwald et al., 2019). Many recent studies have described aging-related change for different brain structures (Du et al., 2006; Leonard et al., 2008; Driscoll et al., 2009; Fjell and Walhovd, 2010; Raz et al., 2010, 2005; Walhovd et al., 2011; Ziegler et al., 2012; Fjell et al., 2013, 2009; Pfefferbaum et al., 2013; Storsve et al., 2014; Fraser et al., 2015; Jäncke et al., 2015; Potvin et al., 2016; Coupé et al., 2017; Narvacan et al., 2017; Vinke et al., 2018). For most brain structures investigated in these studies, an average annual volumetric decline ranging between 0.2 and 0.8% has been reported. This volumetric decline comes with some regional variability, meaning that some brain structures are more sensitive to age decline than others, with no clear consensus about which brain region or cortical lobe shows the largest volumetric losses. Furthermore, the results of these studies were heterogeneous, whether the aging-related changes differ between men and women. Besides the extent of decline, many of the aforementioned studies were also interested in whether the trajectories of volumetric decline follow linear or non-linear shapes. Accelerated decline with advancing age has repeatedly been reported in some subcortical and medial temporal lobe structures, while the decline of total gray matter volume has been reported to follow a more linear shape (e.g., Walhovd et al., 2011; Ziegler et al., 2012; Vinke et al., 2018). In general, however, the decline patterns of subcortical structures in the last decades of life seem to be rather diverse (e.g., Walhovd et al., 2011; Ziegler et al., 2012). Generally, estimates of average trajectories seem to be quite noisy because information about average annual decline rates often comes from suboptimal sources, such as (a) cross-sectional studies (e.g., Fjell and Walhovd, 2010; Walhovd et al., 2011; Ziegler et al., 2012), (b) longitudinal studies with only two repeated measures (e.g., Raz et al., 2005; Fjell et al., 2009; Storsve et al., 2014; Narvacan et al., 2017), or (3) longitudinal studies that did not separate the between-subject from the within-subject effects by using between-subject information in estimation of average trajectories (e.g., Driscoll et al., 2009; Narvacan et al., 2017; Vinke et al., 2018). Estimation of average trajectories from cross-sectional data is very noisy because there is a lot of between-subject variance in these brain structures. In addition, the estimates may be more easily biased by other factors (e.g., cohort effects). Secondly, because longitudinal MRI studies are resource-intensive, the number of repeated measures usually falls short. This is problematic because three repeated measures are minimally required to properly estimate a linear slope model with random intercepts and random slopes, while four or more repeated measures are preferred (Muthén and Curran, 1997; Curran et al., 2010). Although there was already a body of research dedicated to structural brain changes in aging, data from longitudinal studies with more than three repeated measures would be beneficial for more accurate estimation. Lastly, by separating between-subject from within-subject variances, longitudinal studies result in more precise estimates for longitudinal changes.

Most of the previous studies (Du et al., 2006; Leonard et al., 2008; Driscoll et al., 2009; Fjell and Walhovd, 2010; Walhovd et al., 2011; Ziegler et al., 2012; Fjell et al., 2013, 2009; Pfefferbaum et al., 2013; Storsve et al., 2014; Fraser et al., 2015; Jäncke et al., 2015; Potvin et al., 2016; Coupé et al., 2017; Narvacan et al., 2017; Vinke et al., 2018) focused on population average annual structural brain changes. While these average structural changes surely are of interest, they provide only a part of the characterization of aging-related change because aging affects people differently. Consequently, we would expect that the changes in brain structure also vary between individuals, even in healthy aging people. There may be brain structures that show high variability in the amount of decline between individuals, while others may show only marginal variability. Brain structures that show high between-person variability are of special interest because these structures may be more vulnerable to environmental factors and lifestyle choices (Raz et al., 2005). In addition, estimating the associations between the person-specific brain structure trajectories would provide important information about the heterogeneity of the aging process in the brain. By this, we could investigate if a person with a faster-than-average decline in brain structure A also has a faster-than-average decline in brain structure B. Associations between the trajectories of two brain structures can be classified into level-level, level-change, change-change associations (Oschwald et al., 2019). A level-level association represents the covariance parameter between random intercepts, level-change represents the covariance between random intercepts and random slopes, and change-change represents the covariance between random slopes. So far, slope variability of different brain structures and slope associations between brain structures have only been described in the study by Raz et al. (2005). The authors reported similar random slope variability for almost all of the brain structures they studied. Evidence for an association, however, was only observed for about one-quarter of the studied slope-to-slope associations (Raz et al., 2005). In a later study, Raz et al. (2010) confirmed their previous results regarding slope variability but did not report any associations between the trajectories (Raz et al., 2010, 2005). Due to the limited number of repeated

measures, small sample sizes in combination with wide age ranges in these two studies, the resulting estimates should be taken with caution.

Using a longitudinal dataset with four measurement occasions over the span of 4 years, our main goal therefore was (1) to provide more information on the average volumetric decline of lobular and subcortical structures as well as its shape (acceleration of decline with advancing age) and predictors (e.g., potential sex differences) and, most importantly, (2) to gain new insights about the variations of and the associations between person-specific brain structure trajectories in healthy aging.

MATERIALS AND METHODS

Sample Description

Structural MRI data were taken from the Longitudinal Healthy Aging Brain Database Project (LHAB; Switzerland) – an ongoing project conducted at the University of Zurich (Zöllig et al., 2011). We used data from the first four measurement occasions (baseline, 1-year follow-up, 2-year follow-up, 4-year follow-up). For 24 subjects, additional 3-year follow-up data were collected. The baseline LHAB dataset included 232 participants, of which 231 had MRI data and were therefore included in the current analysis (age at baseline: M = 70.8, range = 64–87; females: 113). At each measurement occasion, participants completed an extensive battery of neuropsychological and psychometric cognitive tests and underwent brain imaging. The brain imaging data was usually acquired in the same week as the behavioral assessments. Inclusion criteria for study participation at baseline were age $\geq$64, right-handedness, fluent German language proficiency, a score of $\geq$26 on the Mini Mental State Examination (MMSE; Folstein et al., 1975), no self-reported neurological disease of the central nervous system and no contraindications to MRI. Participation was voluntary and all participants gave written informed consent in accordance with the declaration of Helsinki. Self-reported physical and mental health of the sample at baseline, as measured by the SF-12 (Ware et al., 1996), were 50.9 $\pm$ 7.4 (M $\pm$ SD) and 54.8 $\pm$ 6.3, respectively, which indicates above-average health compared to a norm population (Ware et al., 1995). As expected, sample means for these general health indicators slightly declined over time, but still indicated above-average health at 4-year follow-up (physical health score: 50.5 $\pm$ 6.9, mental health score: 53.1 $\pm$ 8.0, MMSE = 28.3 $\pm$ 1.3). At 4-year follow-up, the structural MRI dataset still comprised 72% of the baseline sample (N = 166). The exact attrition pattern is shown in **Supplementary Table S1**. We were assuming that the missing mechanism was missing at random, meaning that - given the covariates and observed values – missingness should not depend on unobserved values (Bhaskaran and Smeeth, 2014). By using the normalization of the age category of 70–90 years for the entire sample, the mean IQ of the sample was 120.6 (SD = 6.7) at baseline. In follow-up measurements, the participants achieved on average similar IQ scores (Jäncke et al., 2019). Finally, acquisition and processing of MRI data is prone to unwanted influences and errors. We excluded subjects who had rather large influence on parameter estimation as indicated by the Cook's distance and the loglikelihood contribution of observations (Cook and Weisberg, 1982; Cook, 1986). These Influence measures were obtained with latent growth curve models using the time window approach for the time intervals. Subjects having a Cook's distance $>$ 0.5 and a likelihood contribution of $<$ $-$7.5 or a Cook's distance $>$ 1 and a likelihood contribution of $<$ $-$4 were excluded. These values were chosen based on visual inspection of the excluded subjects. Depending on the brain structure, 0 to 4 subjects were excluded. The LHAB sample has been used in previous publications of our group (e.g., Jäncke et al., 2019; Oschwald et al., 2019).

Image Acquisition

Magnetic resonance imaging scanning was carried out at the University Hospital of Zurich on a 3.0T Philips Ingenia scanner (Philips Medical Systems, Best, Netherlands). T1-weighted images were recorded with a gradient echo sequence (3D turbo field echo, 160 sagittal slices, slice thickness = 1 mm, in-plane resolution = 1 $\times$ 1 mm, FOV = 240 $\times$ 240 mm, repetition time = 8.18 ms, echo time = 3.80 ms, flip angle = 8).

Image Processing

FreeSurfer (v5.3, Fischl, 2012) as implemented in the FreeSurfer BIDS-App (Gorgolewski et al., 2017) was used to obtain volumetric measurements of cortical and subcortical structures using the Desikan-Killiany parcellation scheme (Desikan et al., 2006). As part of our data processing pipeline, the structural MR images were visually inspected for good SNR and obvious artifacts (such as motion). In addition, the surfaces, created by the FreeSurfer software, were carefully visually checked for gross deviations. Only very few images (N = 24) had to be excluded from the sample due to insufficient data quality. In addition, for some images (chosen at random), the surfaces, created by the FreeSurfer software, were visually checked for gross deviations.

We deliberately refrained from applying manual correction of the reconstructed surfaces given that previous research showed limited applicability of manual corrections (McCarthy et al., 2015). Furthermore, we believe that manual corrections, particularly in case of longitudinal structural MRI data, can bias the images, not only through between-rater discrepancies but also through inconsistencies in applying corrections across time points.

The sum of the left and the right hemisphere volumes were used for each brain structure. We used the sum of the left and the right hemisphere volumes because we think that the noise in volumetric MRI data is rather high to quantify more fine-grained changes between the hemispheres. Over the chosen brain parcellation, we expect the slopes between the hemispheres to be highly correlated. Furthermore, by using bi-lateral sums over both hemispheres, we could reduce the number of variables. Fewer variables facilitates visualization and pattern detection. Likewise, to keep the number of variables low, we decided to focus on volumetric changes, as one of the most important and widely used structural measures. Additionally, the sole focus on

volumetric changes in cortical regions facilitates the comparison with changes seen in subcortical regions.

Statistical Modeling

The trajectories of the brain structures were fitted with linear random slope models, allowing for person-specific intercepts and slopes. To separate the between-subject variance from the within-subject variance, the age predictor was separated in the age-at-study-entry predictor (entry-age) and in the time-difference-between-the-baseline-and-the-subsequent-measurements predictor (slope) (Grimm et al., 2016). Because there were some small deviations from the planned time interval between measures for some subjects, we used the exact follow-up time in the mixed model framework and used a time-window approach in the latent growth curve framework (Grimm et al., 2016). The models were slightly modified depending on the predictors and parameters of interest as described below.

Average population trajectories were fitted with linear random slope models, which included as predictors: slope, intracranial volume (ICV), entry-age, $(\text{entry-age})^2$, slope × entry-age, and slope × ICV interaction. Main parameters of interest were the slope and the slope × entry-age interaction. These two parameters determine the average volumetric decline. The slope × entry-age interaction indicates whether a larger decline with advancing age at study entry is expected. The expected volumetric loss over the span of 20 years (starting at age 65 with an updated slope every 4 years until age 85) was extrapolated from the linear combination of the slope and the slope × entry-age interaction parameter. Bootstrap was used to gain information about the variability of these estimates, as well as about fastest declining brain structures by comparing the extrapolated volumetric declines between each brain structure to each other brain structure in a pairwise fashion. A bootstrapped mass of Region A > Region B was obtained by calculating the percent of bootstrap samples where the decline in region A was estimated to be larger than the decline in region B.

To investigate the effect of sex, average trajectories were fitted for men and women using the same basic model as described above with additional predictors for sex, sex × entry-age, $(\text{sex} \times \text{entry-age})^2$, sex × slope, and sex × slope × entry-age interactions. Main parameters of interest were the sex × slope and sex × slope × entry-age interactions, as well as the random slope variance parameter. The sex × slope and sex × slope × entry-age interactions quantify the difference between the slopes of men and the slopes of women. The sex × slope × entry-age interaction parameter provides information whether one of the sexes declines faster with advancing age. The random slope variance parameter quantifies how much individuals deviate from the average slopes and therefore shows how heterogenous the decline is in the population.

To evaluate how the decline of brain structures is related to changes in other brain structures, we fitted associations between the trajectories of the different brain structures with bivariate growth models. We applied basically the same (univariate) model as described in the last paragraph but allowing for associations between random intercepts, random slopes, and within-subject errors (at the same time point) between the models. In this analysis, we were mainly interested in the correlation between random slopes. Consequently, only these correlations are reported in the corresponding results section. The results of other analyzed types of associations are provided in the **Supplementary Figures S8, S9**.

We used a time window approach to approximate the exact time difference between measurements. The number of time bins and the distances were chosen with a k-median algorithm implemented in the R package *Ckmeans.1d.dp* (Wang and Song, 2011). This resulted in eleven time-bins (0, 1.0, 1.1, 2.0, 2.1, 2.2, 3.0, 4.0, 4.1, 4.2, 4.5 years). Models were estimated using Bayesian estimation with the default priors implemented in Mplus [prior for variance-covariance matrices of size p $\sim$ Inverse-Wishart (0, -p-1), priors for intercepts and slopes $\sim$ Normal $(0,10^{10})$]. This default covariance prior corresponds to a uniform prior on (-infinity to infinity) for all elements of the covariance matrix (Asparouhov and Muthén, 2010a). A graphical representation of the fitted bivariate models is shown in **Figure 1**. Bayesian estimation was used (instead of maximum likelihood) because the estimated random slope variance parameters were very small (compared to the within-subject error) for some brain structures. The small random slope parameters may have led to estimation difficulties in the maximum likelihood framework.

To obtain a simple estimate of the multivariate correlation structure on which it was possible to do principal component analysis (PCA), we sampled the plausible values (Asparouhov and Muthén, 2010b) of the random slopes from each of the bivariate models, connecting each structure to each other structure. To obtain just one random slope estimate per person for each structure, we used the mean of the means from the distributions of the plausible values of the bivariate models. The random slope estimates were adjusted for the covariates ICV, sex, and entry-age using regression. Note that this simple estimate will be biased and may underestimate or overestimate the correlations. However, the general covariance pattern should be retained. Visualizing the principal components may reveal some patterns that might not be apparent by looking at the bivariate correlations.

Data Scaling

The various brain regions differ in their total size. To put the amount of volumetric loss on comparable scales between the different brain structures, each brain structure was divided by its intercept multiplied by 100 (at an age of 65 years estimated with linear random slope models including as covariates on the intercept: ICV (grand mean-centered), entry-age, $(\text{entry-age})^2$. Using this scaling, the slope parameter corresponds to the annual percent change of the intercept (at age 65).

Software

The mixed models were fitted with the R package *lme4* (v. 1.1-21, Bates et al., 2015). Univariate (for influence measures) and bivariate latent growth curve models were fitted in Mplus (v. 8.4, Muthén and Muthén, 1998–2017)[1]. The R package *MplusAutomation* (v. 0.7-3, Hallquist and Wiley, 2018) was used

[1]https://www.statmodel.com

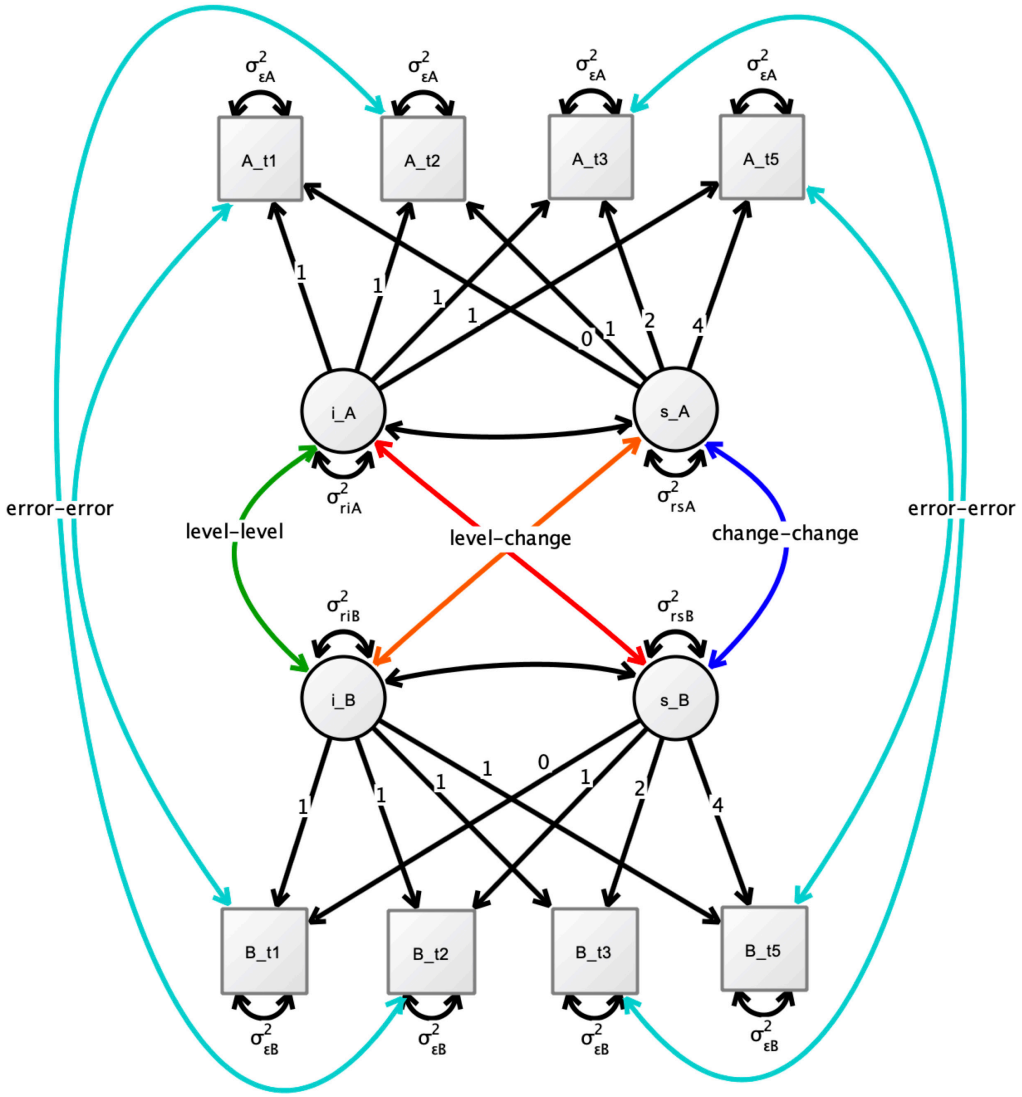


FIGURE 1 | Simplified illustration of the used bivariate growth model for two brain structures A and B. i_ = random intercept. s_ = random slope. Not shown are the included covariates on the intercept and slopes. Also, the separation of the time intervals into eleven time bins from 0 to 4.5 years is not illustrated. The main interest in our study was on the change-change associations between brain structures. t1: baseline, t2 = 1-year follow-up, t3 = 2-year follow-up, 5 = 4-year follow-up.

to fit the bivariate models in a loop. Principle components were plotted with the R package *factorextra* (v. 1.0.5)[2].

RESULTS

The raw data trajectories of the different brain structures are shown **Figure 2**. Excluded subjects are shown in **Supplementary Figure S1**.

Average Volumetric Declines

The expected volumetric declines over the span of 20 years are shown in **Figure 3**. The expected slopes were extrapolated from the linear combination of the slope and the slope × entry-age interaction parameter, starting at age 65 with an updated slope every 4 years.

We observed some variability in the average volumetric decline of brain structures belonging to the frontal, temporal, and parietal lobe. However, generally, the expected declines from age 65 to age 85 of these regions seem to be in similar ranges (ranging from about 12 to 15% for the majority of structures belonging to the frontal, temporal or parietal lobe). Slightly smaller volumetric declines were estimated for the lateral orbitofrontal cortex and medial orbitofrontal cortex (10%), and the paracentral lobule, rostral middle frontal gyrus and superior frontal gyrus (∼11%). Slightly larger declines were estimated for the fusiform gyrus, entorhinal cortex, temporal

[2]https://CRAN.R-project.org/package=factoextra

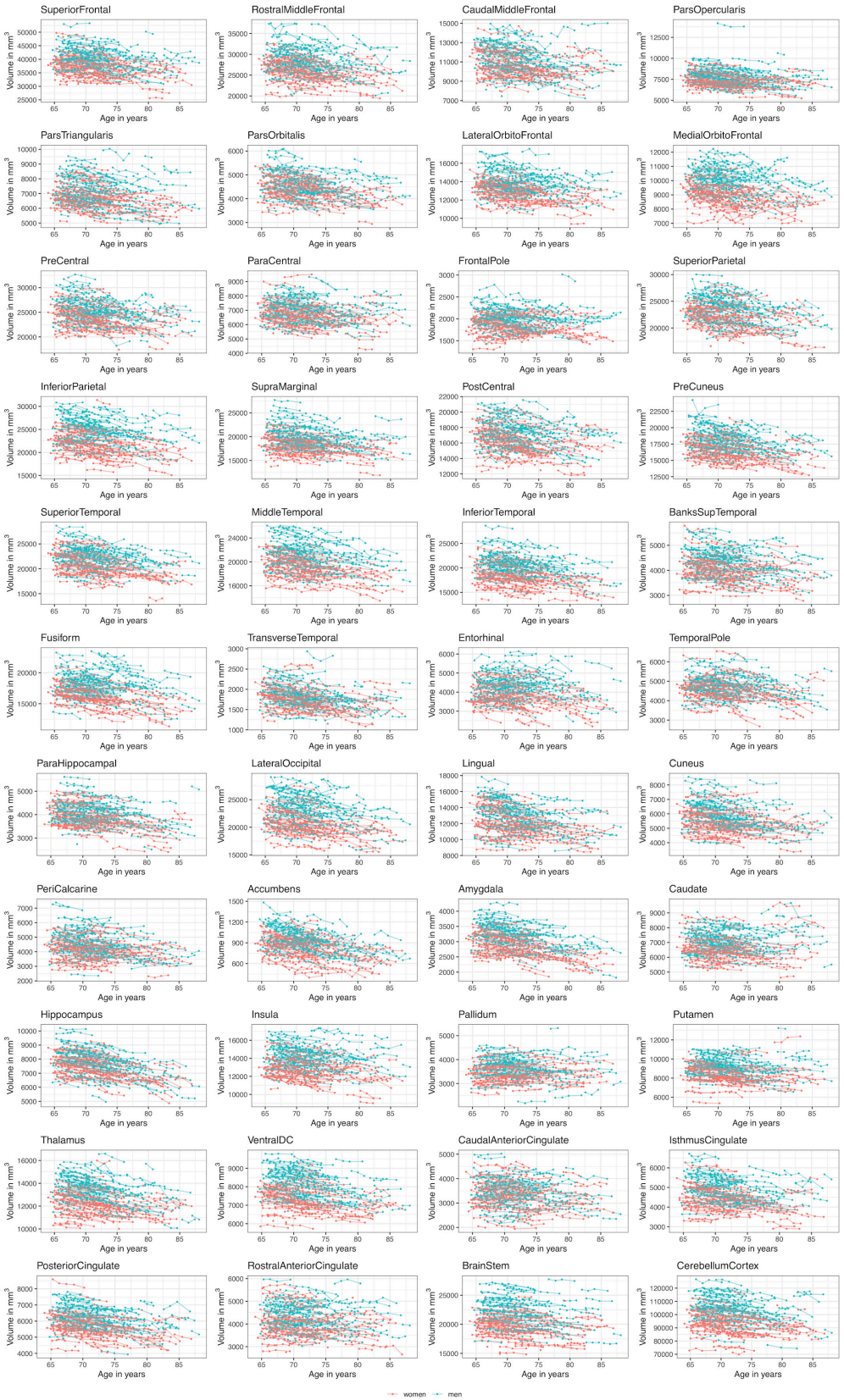


FIGURE 2 | Volumetric brain structure trajectories (sum of left and right hemisphere) colored by sex.

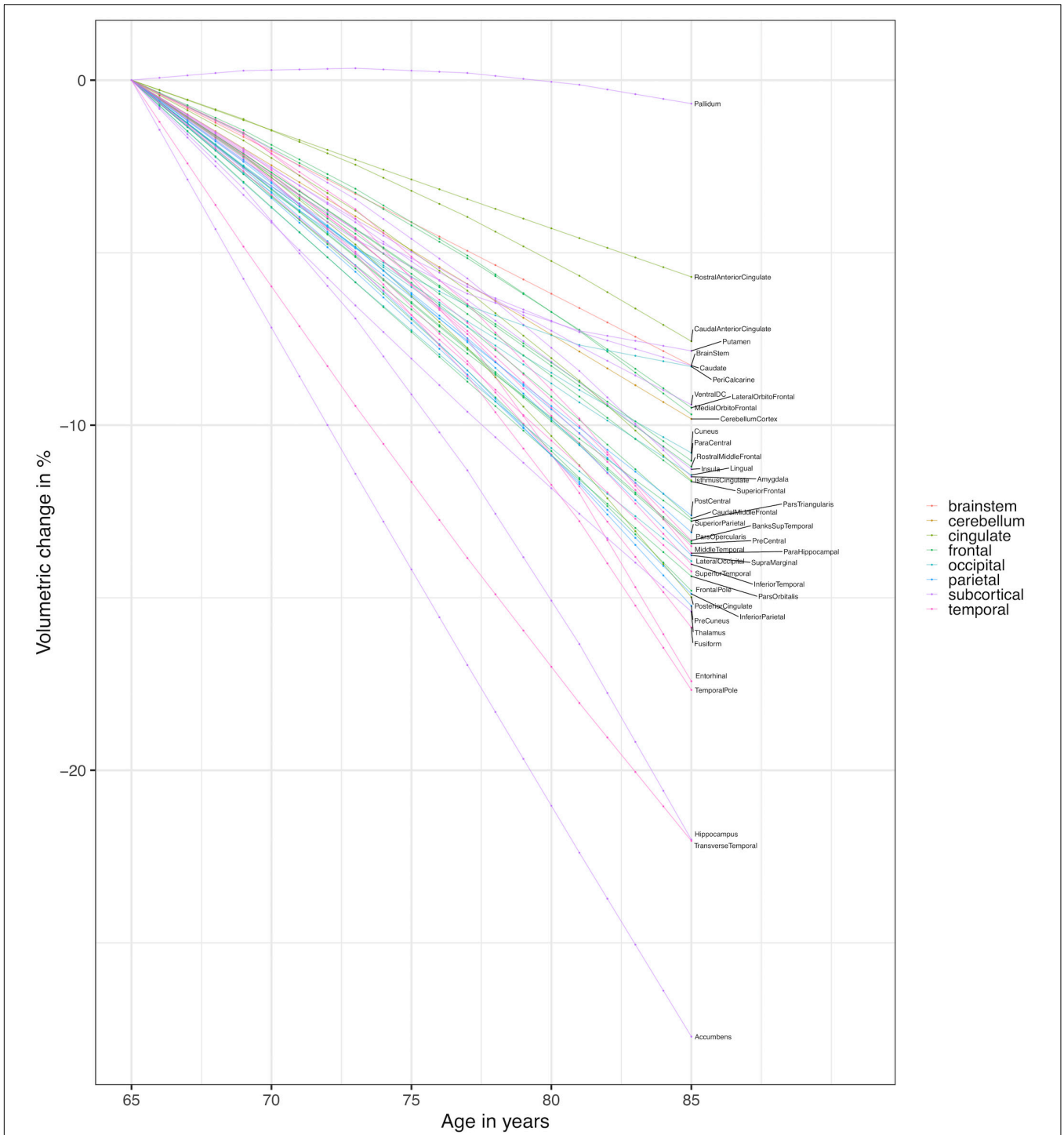


FIGURE 3 | Expected structural slopes. The expected volumetric loss (of the average volumes at age 65) over the span of 20 years (starting at age 65 with an updated slope every 4 years until age 85) was extrapolated from the linear combination of the slope and the slope × entry-age interaction parameters.

pole, and transverse temporal gyrus (∼16 to 22%). Of the occipital regions, the peri-calcarine cortex showed the smallest decline (8%), followed by the cuneus and lingual gyrus (∼11%). The lateral occipital cortex showed a similar decline (14%) as structures of other lobes. Parameter estimates are shown in **Table 1**. Pairwise comparisons using bootstrap are shown in **Supplementary Figure S2**.

In most of the temporal, frontal, and parietal regions, an accelerating decline with advancing age was estimated (slope × entry-age parameter in **Table 1**). This effect was

particularly large in the medial temporal lobe regions (fusiform gyrus, temporal pole, entorhinal cortex, and parahippocampal gyrus). In frontal and parietal regions, the acceleration was less pronounced. In these regions, zero was included in all of the 95% confidence intervals of the slope × entry-age parameter. In occipital regions, the decline seemed

TABLE 1 | Volumetric changes in % of brain structures average volumes (at age 65) with 95% confidence intervals.

Regions	Slope	Slope × Entry-Age	Loss 20 years	Random Slope SD	Random Slope *P* value	Sex *P* value
Caudal Middle Frontal	−0.556 (−0.398, −0.707)	−0.010 (0.013, −0.034)	−12.726 (−10.593, −14.891)	0.125 (0.008, 0.356)	0.563	0.074
Frontal Pole	−0.663 (−0.443, −0.871)	−0.010 (0.022, −0.044)	−14.831 (−11.681, −18.038)	0.458 (0.197, 0.656)	0.076	0.479
Lateral Orbitofrontal	−0.385 (−0.260, −0.508)	−0.011 (0.008, −0.031)	−9.546 (−7.719, −11.354)	0.108 (0.003, 0.288)	0.887	0.125
Medial Orbitofrontal	−0.363 (−0.178, −0.559)	−0.015 (0.012, −0.042)	−9.726 (−7.288, −12.093)	0.162 (0.007, 0.427)	0.937	0.371
Para-Central	−0.533 (−0.376, −0.686)	−0.002 (0.019, −0.025)	−11.065 (−8.933, −13.215)	0.173 (0.005, 0.360)	0.912	0.044
Pars Opercularis	−0.626 (−0.503, −0.751)	−0.005 (0.013, −0.024)	−13.394 (−11.692, −15.086)	0.154 (0.024, 0.303)	0.347	0.026
Pars Orbitalis	−0.737 (−0.597, −0.873)	0.002 (0.023, −0.020)	−14.437 (−12.487, −16.419)	0.001 (0.000, 0.291)	1.000	0.367
Pars Triangularis	−0.682 (−0.546, −0.812)	0.005 (0.026, −0.016)	−12.808 (−10.902, −14.795)	0.212 (0.085, 0.354)	0.003	0.168
Pre-Central	−0.631 (−0.486, −0.770)	−0.005 (0.016, −0.027)	−13.446 (−11.422, −15.498)	0.214 (0.013, 0.373)	0.359	0.051
Rostral Middle Frontal	−0.533 (−0.378, −0.690)	−0.003 (0.019, −0.027)	−11.261 (−9.235, −13.200)	0.043 (0.000, 0.109)	0.633	0.024
Superior Frontal	−0.548 (−0.401, −0.691)	−0.004 (0.017, −0.026)	−11.655 (−9.734, −13.585)	0.178 (0.012, 0.341)	0.342	0.010
Inferior Parietal	−0.677 (−0.563, −0.791)	−0.008 (0.008, −0.026)	−14.904 (−13.277, −16.545)	0.190 (0.011, 0.312)	0.487	0.113
Post-Central	−0.627 (−0.502, −0.749)	−0.000 (0.017, −0.018)	−12.633 (−10.944, −14.291)	0.194 (0.013, 0.338)	0.560	0.630
Pre-Cuneus	−0.641 (−0.533, −0.750)	−0.015 (0.002, −0.033)	−15.256 (−13.451, −17.065)	0.186 (0.012, 0.323)	0.548	0.015
Superior Parietal	−0.592 (−0.457, −0.725)	−0.008 (0.010, −0.027)	−13.117 (−11.419, −14.860)	0.072 (0.013, 0.299)	0.203	0.010
Supra Marginal	−0.577 (−0.463, −0.688)	−0.014 (0.002, −0.031)	−13.769 (−12.304, −15.320)	0.209 (0.023, 0.303)	0.143	0.015
Banks Sup Temporal	−0.505 (−0.390, −0.621)	−0.020 (−0.001, −0.040)	−13.354 (−11.381, −15.374)	0.223 (0.062, 0.341)	0.110	0.019
Entorhinal	−0.379 (−0.141, −0.629)	−0.062 (−0.027, −0.095)	−17.423 (−13.676, −21.288)	0.782 (0.613, 0.965)	3e-06	0.025
Fusiform	−0.564 (−0.457, −0.674)	−0.029 (−0.012, −0.046)	−15.856 (−14.099, −17.708)	0.210 (0.038, 0.328)	0.188	0.206
Inferior Temporal	−0.525 (−0.399, −0.647)	−0.022 (−0.003, −0.041)	−14.035 (−12.156, −15.990)	0.267 (0.099, 0.388)	0.105	0.062
Middle Temporal	−0.536 (−0.420, −0.652)	−0.017 (−0.001, −0.035)	−13.515 (−11.896, −15.192)	0.266 (0.168, 0.353)	0.019	0.069
Para-Hippocampal	−0.395 (−0.270, −0.522)	−0.036 (−0.015, −0.057)	−13.702 (−11.561, −15.930)	0.341 (0.224, 0.445)	0.004	0.033
Superior Temporal	−0.660 (−0.535, −0.786)	−0.006 (0.011, −0.024)	−14.250 (−12.696, −15.852)	0.224 (0.057, 0.341)	0.172	0.021
Temporal Pole	−0.545 (−0.357, −0.738)	−0.042 (−0.013, −0.073)	−17.623 (−14.605, −20.960)	0.598 (0.475, 0.721)	1e-08	0.300
Transverse Temporal	−1.206 (−1.032, −1.376)	0.013 (0.038, −0.013)	−22.089 (−19.577, −24.551)	0.211 (0.024, 0.422)	0.374	0.004
Cuneus	−0.624 (−0.492, −0.758)	0.011 (0.033, −0.012)	−10.844 (−8.562, −13.146)	0.213 (0.028, 0.399)	0.152	0.128
Lateral Occipital	−0.743 (−0.616, −0.869)	0.006 (0.026, −0.015)	−13.958 (−12.028, −15.979)	0.257 (0.103, 0.361)	0.068	0.433
Lingual	−0.618 (−0.502, −0.736)	0.006 (0.024, −0.012)	−11.457 (−9.581, −13.374)	0.340 (0.245, 0.446)	4e-04	0.101
Peri-calcarine	−0.674 (−0.446, −0.899)	0.032 (0.063, 0.003)	−8.330 (−5.440, −11.149)	0.490 (0.277, 0.668)	0.016	0.616
Caudal Anterior Cingulate	−0.282 (−0.154, −0.407)	−0.012 (0.005, −0.030)	−7.575 (−5.864, −9.296)	0.005 (0.004, 0.290)	0.992	0.010
Isthmus Cingulate	−0.437 (−0.309, −0.566)	−0.018 (0.005, −0.041)	−11.589 (−9.262, −13.941)	0.328 (0.216, 0.432)	0.002	0.026
Posterior Cingulate	−0.544 (−0.415, −0.672)	−0.026 (−0.004, −0.047)	−15.013 (−12.996, −17.032)	0.200 (0.018, 0.325)	0.457	0.005
Rostral Anterior Cingulate	−0.289 (−0.144, −0.439)	0.001 (0.022, −0.022)	−5.709 (−3.663, −7.796)	0.021 (0.001, 0.093)	0.900	0.013
Accumbens	−1.439 (−1.142, −1.750)	0.007 (0.047, −0.034)	−27.643 (−23.398, −32.031)	0.429 (0.105, 0.843)	0.063	0.956
Amygdala	−0.383 (−0.250, −0.517)	−0.024 (−0.003, −0.043)	−11.455 (−9.269, −13.621)	0.481 (0.392, 0.578)	6e-14	0.573
Caudate	−0.585 (−0.450, −0.718)	0.021 (0.044, −0.002)	−8.305 (−5.956, −10.606)	0.463 (0.385, 0.555)	3e-12	0.568
Hippocampus	−0.784 (−0.666, −0.904)	−0.040 (−0.018, −0.061)	−21.999 (−19.656, −24.339)	0.441 (0.366, 0.518)	1e-13	0.286
Insula	−0.499 (−0.374, −0.623)	−0.008 (0.011, −0.028)	−11.262 (−9.342, −13.373)	0.307 (0.198, 0.423)	0.003	0.016
Pallidum	0.069 (0.161, −0.024)	−0.013 (0.003, −0.029)	−0.703 (0.991, −2.446)	0.203 (0.065, 0.316)	0.230	0.986
Putamen	−0.639 (−0.529, −0.748)	0.031 (0.046, 0.016)	−7.850 (−6.318, −9.386)	0.326 (0.254, 0.405)	1e-05	0.697
Thalamus	−0.832 (−0.745, −0.920)	0.008 (0.021, −0.006)	−15.409 (−14.043, −16.777)	0.240 (0.170, 0.304)	3e-04	0.385
Ventral DC	−0.516 (−0.438, −0.593)	0.006 (0.018, −0.006)	−9.409 (−8.180, −10.632)	0.210 (0.146, 0.276)	0.001	0.016
Brain Stem	−0.411 (−0.362, −0.461)	−0.000 (0.007, −0.007)	−8.268 (−7.493, −9.028)	0.112 (0.042, 0.164)	0.075	0.607
Cerebellum Cortex	−0.494 (−0.418, −0.572)	0.000 (0.012, −0.011)	−9.818 (−8.654, −10.960)	0.147 (0.027, 0.219)	0.193	0.827

Slope = expected volumetric annual changes at age 65. Slope × Entry-Age = interaction slope with entry-age. Loss 20 years = expected volumetric loss over the span of 20 years (starting at age 65 with an updated slope every 4 years until age 85) extrapolated from the linear combination of the slope and the slope × entry-age interaction parameters. Random Slope P value = p-value comparing a random slope model to a random intercept model with a likelihood ratio test. Sex P value = p-value obtained with a likelihood ratio test comparing a model including a separate slope and a separate slope × entry-age interaction parameter for men and women to a model without separate parameters.

more steady (lateral occipital cortex, lingual gyrus) or even decelerating with advancing age (peri-calcarine cortex, cuneus). The slope × entry-age parameter estimates are plotted in **Supplementary Figure S3**.

Volumetric declines of the subcortical regions were heterogeneous. The pallidum showed almost no decline (1%), the putamen and the caudate showed small declines (∼8%), the amygdala and the insula showed a moderate decline (11%), the thalamus a slightly larger decline (15%), and the hippocampus (22%) and the nucleus accumbens (28%) showed a rather large decline in comparison to other regions. The nucleus accumbens showed the largest decline of all the observed brain structures, with a steady decline of about 1.4% per year. The declines of the hippocampus and the amygdala were clearly accelerating with advancing age, while the decline of the putamen and caudate was decelerating with advancing age.

Sex Differences in Trajectories

Given that the estimated total ICV was already included (**Table 1**, the *x*-axis of **Supplementary Figure S4**), allowing for different slopes and different slope × entry-age interactions for men and women did not substantially improve the model fit for most of the brain structures. We would like to emphasize that using ICV as covariate usually eliminates most of the sex influences on brain volume measures (Jäncke et al., 2015, 2019). As expected, the estimated influence of sex were small in comparison with the general slopes at age 70 (slope parameter of the *y*-axis of **Supplementary Figure S4** and **Supplementary Table S2**). In sum, the data provide little (or even no) evidence that one of the sexes declines faster than the other in the examined age range considering the sample size, the number of comparisons, the observational nature of the study, and the freedom in modeling such trajectories.

Random Slope Variances

The estimated random slope variance parameters from the random slope models (with the following covariates for the slope parameter: entry-age, total-ICV, sex and sex × entry-age) are shown in **Table 1** and in **Supplementary Figure S5**. Reliable ($p < 0.05$, **Table 1**) random slope variability was observed in subcortical regions (amygdala, caudate, hippocampus, putamen, thalamus, insula), in medial temporal lobe structures and temporal pole, and in lingual gyrus and peri-calcarine cortex with random slope standard deviations typically ranging between 0.2 and 0.5%. The largest variability was observed in the entorhinal cortex (0.8%), followed by the temporal pole (0.6%), followed by the amygdala, caudate, and hippocampus (∼0.5%). Model fit (evaluated with the likelihood ratio test) clearly improved in these regions when including the random slope parameter. Large variability was further observed in the peri-calcarine cortex, frontal pole, and nucleus accumbens (∼0.5%). However, the model fit did not improve clearly for these regions due to large uncertainty in the estimated random slopes. In most regions of the frontal lobe, the parietal lobe, and the cingulum the random slope standard deviation estimate ranged from 0 to 0.2%. Allowing for random slopes did not substantially improve the model fit in these regions. To illustrate the random slope variation in the expected annual changes, the estimated random slopes are plotted alongside the average slopes at an age of 70 years in **Figure 4**. In frontal and parietal regions, some subjects were expected to have little decline (∼1%) while others seem to be clearly declining (∼4%) over the span of 4 years. In regions with reliable random slope differences (e.g., hippocampus), some subjects showed almost no decline while others were declining about twice as fast as the average slope.

Associations Between Random Slopes

Associations between the random slopes were estimated in bivariate models using Bayesian estimation with the default priors as provided by *Mplus*. **Figure 5** shows the medians of the posterior distributions of the correlation parameter between random slopes from the bivariate models. The exact numbers are shown in **Supplementary Figure S6**.

In general, there are strong associations between the random slopes within and between the frontal, temporal, parietal and to a lesser extent occipital lobe with estimated correlations typically ranging between 0.5 and 0.9. However, many estimates are not reliable because associations between the slopes estimated for brain structures with little estimated random slope variance (compared to the within-subject error term) resulted in posterior distributions with a large variance under the assumed model. As can be seen in **Figure 5**, this was the case for a lot of structures.

There are a few exceptions to this homogenous correlation pattern. While the entorhinal cortex, parahippocampal gyrus and temporal pole showed strong and reliable associations with each other, with the hippocampus, and the amygdala, and strong but less reliable associations with other regions (the insula, lateral orbitofrontal and medial orbitofrontal cortex), they were only weakly correlated with most of the other brain structures. Regions from the lateral medial orbital and lateral orbitofrontal cortex showed rather strong but unreliable correlations with each other and further regions (with the temporal pole, with the entorhinal cortex, with the amygdala and with the hippocampus), but weak correlations with most other brain structures. The peri-calcarine cortex showed moderate correlations with other occipital structures and rather weak correlations to other lobe structures.

The correlation pattern involving subcortical structures was in general more diverse. Amygdala, hippocampus, and to a slightly lesser extent insula and thalamus, were strongly and reliably correlated with each other and with temporal lobe structures, but also quite strongly with most other brain structures. The pallidum showed weak correlations with other structures. Putamen and caudate showed strong and reliable correlations with each other but weak or unreliable correlations with other structures. Nucleus accumbens stands out of the pattern because it was negatively correlated with most brain structures. Positive correlations for the nucleus accumbens were limited to the putamen, caudate and ventral DC (diencephalon). However, the estimates involving the nucleus accumbens were unreliable.

To obtain a simple estimate of the multivariate correlation structure on which it was possible to do PCA, we used the mean of the means from the plausible values distributions of the

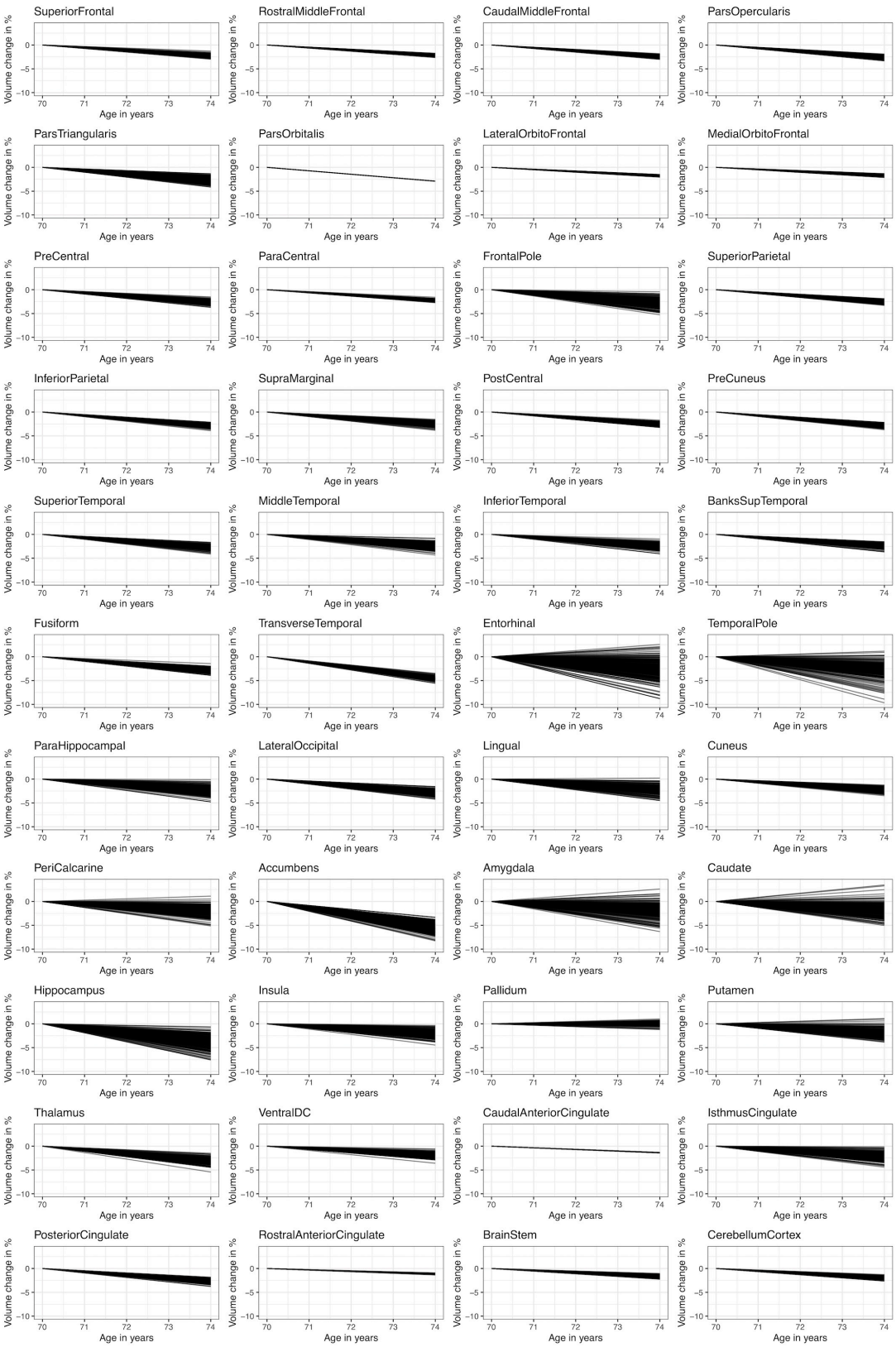


FIGURE 4 | Estimated person-specific slopes in comparison to the average slope at an age of 70 years. *Y*-axis represents the% of average volume (at age 65) change. The random slopes were added to the average slope at an age of 70 years.

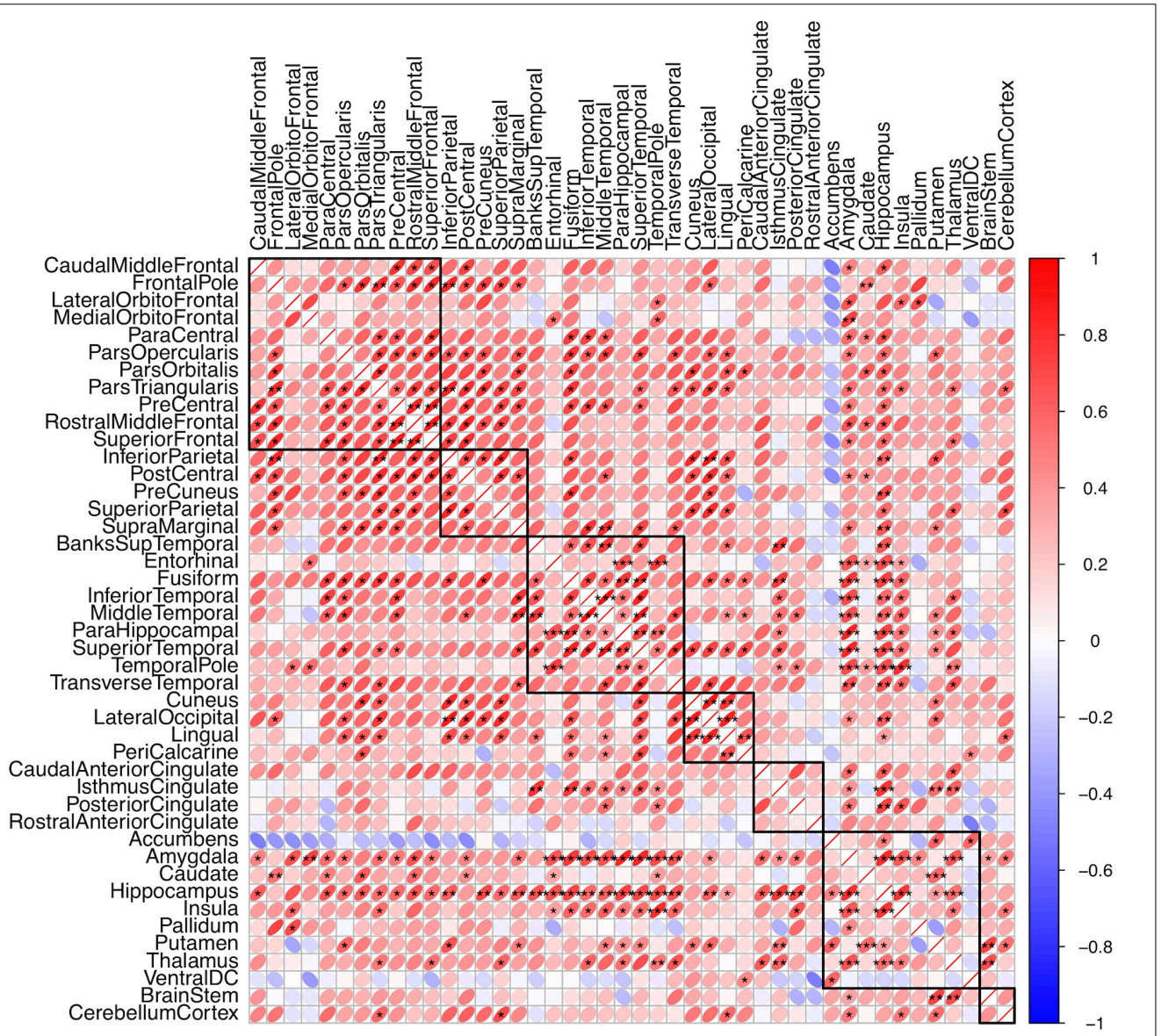


FIGURE 5 | Medians of posterior distributions of correlations between the random slopes of different brain structures. The variance of the posteriors are quite large for many brain structures (as can indirectly be seen when the median is rather far away from zero, but a substantial mass of the posterior distribution is on the other side of zero; * indicates a posterior mass on the other side of zero < 0.05, ** corresponds to < 0.005, *** corresponds to < 0.0005).

bivariate models. The estimated multivariate correlation matrix of these random slope estimates is shown in **Supplementary Figure S7**. In general, the correlations were slightly smaller than the ones estimated in the bivariate models but the pattern looked similar. This suggests that the explained variance in PCA may be underestimated, but that the direction of the principle components was reasonably approximated.

Because the data were on similar scales already through the scaling by the intercept in each region, we applied the PCA on the unscaled person-specific slope estimates. The first principle component explained about 35% of the between-person variance of the random slopes and can be seen as a weighted average of most brain regions. It loads strongly on the entorhinal cortex, temporal pole, hippocampus, and amygdala because these regions had large random slope variances. The second principle component (explaining 12% of the slope variances), separates the entorhinal cortex, the temporal pole and the nucleus accumbens from most other structures, indicating that after the general decline has been accounted for, there is still a lot of variability in our sample in these regions (**Figure 6** and **Supplementary Table S3**). The third principle component loads strongly on caudate, and nucleus accumbens, but also on putamen, entorhinal cortex, peri-calcarine cortex, and lingual gyrus, and separates these regions from most other regions (especially amygdala). The fourth component separates the nucleus accumbens, amygdala, and hippocampus from

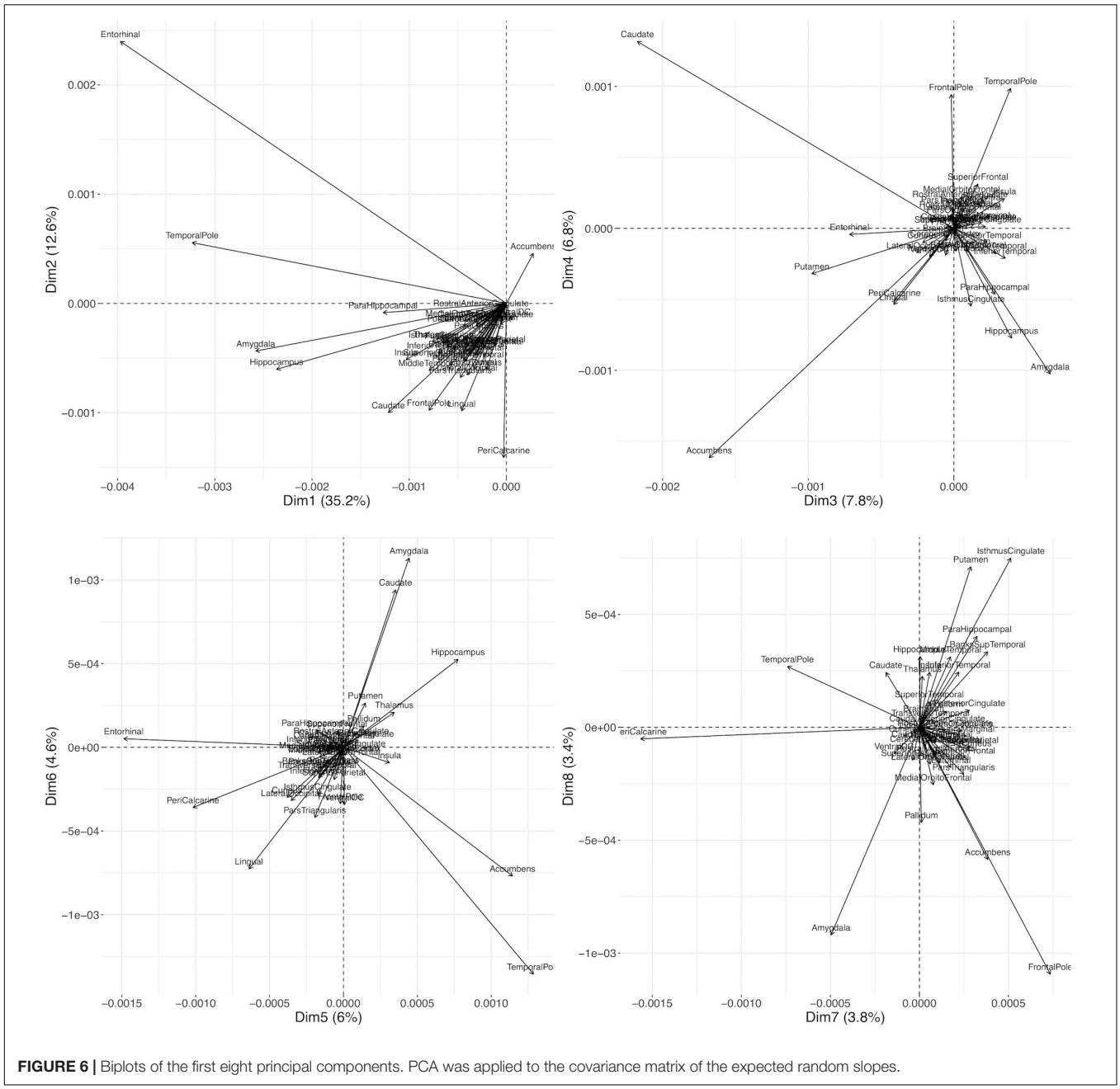


FIGURE 6 | Biplots of the first eight principal components. PCA was applied to the covariance matrix of the expected random slopes.

caudate, frontal pole, and temporal pole. The first 10 principal components explained already 85% of the 44 (brain structure) random slope variances.

DISCUSSION

Our study extends the current knowledge about brain structure trajectories by providing information about healthy aging from a longitudinal study with four repeated measurements and large sample size. In addition to describing the average brain volume trajectories, we put particular emphasis on the variation of the person-specific trajectories and the associations between the person-specific trajectories.

Brain Regions With Largest Decline

Previous research presents a heterogeneous picture about which brain structures are most affected by the volumetric decline in aging. Frontal and temporal lobes were reported to show the largest decline in a large sample size study by Vinke et al. (2018). Other studies report some regional variability in the average volumetric decline but not a clear preference regarding largest volumetric losses for one lobe or lobular region, whereby regions of the occipital lobe were often found to show the

smallest decline of the lobular regions (Driscoll et al., 2009; Fjell et al., 2009; Raz et al., 2010, 2005; Pfefferbaum et al., 2013; Storsve et al., 2014). This is in accordance with our data that showed some variability in the average decline in regions of the frontal, temporal and parietal lobes, but in general, these brain regions showed similar volumetric reductions. Entorhinal cortex, temporal pole and transverse temporal gyrus showed slightly larger declines, while half of the regions belonging to the frontal lobe (lateral and medial orbitofrontal cortex, paracentral lobule, rostral middle frontal gyrus and superior frontal gyrus) and regions belonging to the occipital lobe (peri-calcarine gyrus, cuneus, lingual gyrus) showed slightly smaller volumetric reductions than other brain structures.

Subcortical brain structures seem to differ most with respect to their estimated decline trajectories over 20 years. The hippocampus has often been reported to be one of the brain regions with the largest decline and with an accelerating decline with increasing age (Raz et al., 2010, 2005; Fjell et al., 2013, 2009; Narvacan et al., 2017; Vinke et al., 2018). In our data, the hippocampus showed one of the largest volumetric reductions of all the observed brain regions, and the decline was clearly accelerating with advancing age. Further, the amygdala has often been reported to show an accelerated decline with increasing age (Fjell et al., 2013, 2009; Narvacan et al., 2017), which is also consistent with our data.

In our analyses, the nucleus accumbens showed the steepest decline of all the observed brain structures with no acceleration with age. While this is in line with the studies by Coupé et al. (2017) and Fjell et al. (2013), at least one other study proposed an alternative trajectory shape (U-shaped curve; Vinke et al., 2018). The putamen and the caudate have been reported to show slight increases in very old age (Fjell et al., 2013; Coupé et al., 2017). In our data, the slopes of these two regions (especially of the putamen) were decelerating with increasing age. The pallidum showed almost no decline in our study, which was similarly reported in Narvacan et al. (2017), but which is in contrast to the large study by Vinke et al. (2018).

Of note, studies differ in their ways of data scaling. We scaled our data by the average volume at age 65. The percent of volumetric decline depends on data scaling and on the studied age range. This complicates direct comparison to other studies. Another important aspect is that we did not observe subjects over the span of 20 years but estimated the expected decline over 20 years based on the slope and the slope × entry-age parameters. The estimates should be less noisy than in cross-sectional studies, but there is still more noise in these estimates than it would be the case if subjects had been observed over the span of 20 years.

Sex Differences

Previous studies yielded mixed results regarding the question of whether men differ from women in their aging trajectories. Vinke et al. (2018) report statistically significant age × sex interactions in all studied brain regions. Coupé et al. (2017) found sex differences in individuals beyond age 70 when using brain volumes normalized by the ICV. The few studies that separated the between-subject from the within-subject effects generally found no or only little evidence for sex differences in the studied regions (Raz et al., 2010, 2005; Pfefferbaum et al., 2013; Narvacan et al., 2017). Our analyses, in which ICV was included as an additional covariate, do not provide convincing evidence for sex differences in the slopes in almost all studied brain regions. Based on our data and on previous studies, it seems that the sex differences in the expected trajectories are rather small. The estimates, however, may depend on the assumed model e.g., whether we assume a linear and additive model including the covariates ICV and sex or whether the brain regions submitted to the models are normalized by ICV (e.g., Narvacan et al., 2017 used both measures). Nevertheless, there may be other factors, such as different lifestyle choices, which have a bigger influence on the trajectories (Raz et al., 2010, 2005). We included the sex covariate in the models that were used to estimate the correlation parameters based on theoretical considerations, even though they had little or almost no impact on the estimated correlations.

Random Slope Variances

Quantifying random slope variances of the different brain structures has not been of major interest in most studies to date, except for two studies by Raz et al. (2010, 2005). Reliable random slope variances have been reported in almost all of their studied regions (except the primary occipital and the orbital-frontal cortex). However, the random slope variances in the studies by Raz et al. might have been overestimated given that a clean separation of the within-person error from the random slope term is difficult in case of only two repeated measures (Grimm et al., 2017). Our data, with four measurement occasions, allow to more precisely disentangle the within-person error from the random slope and, thus, lead to more accurate estimates of the random slope variances.

Evaluating our data, reliable random slope variability was observed in subcortical regions, in medial temporal structures and temporal pole, and in lingual gyrus and peri-calcarine cortex. Because the within-subject error is quite large compared to the annual volumetric changes and the person-specific variations thereof, the person-specific slopes and the random slope variance term were estimated with large uncertainty in a lot of regions. Therefore, the conclusion that some regions have little random slope variability and other regions have large variability needs to be taken with caution. Nevertheless, regions belonging to the medial temporal cortex (hippocampus, entorhinal cortex, parahippocampal gyrus), temporal pole, amygdala, and caudate showed reliable random slope variability, and in the population, the random slope variance may be rather large. In these regions, some subjects showed almost no decline while other subjects were declining twice as fast as the average person. However, such a variation of the decline rate may be present in most other brain structures as well, including those with unreliably estimated random slope variances.

Associations Between Random Slopes

One of the main objectives of this study was to investigate the associations of the slopes of different brain structures – an important aspect that was analyzed by Raz et al. (2005), but

that has not been of much interest since then. In their study, estimates were mostly positive but also rather small in magnitude. Reliable associations were only found for 13 of 55 studied slope-to-slope associations. However, we think that these correlations were probably underestimated because the correlation estimates are influenced by the noise of the within-subject error term. Our data suggest that the slope-to-slope correlation pattern is, in general, quite homogenous, especially within but also between most regions of the frontal, temporal and parietal lobe and, to a lesser extent, of the occipital lobe, while the correlation pattern of subcortical regions was more diverse. Besides subcortical structures, there are a few lobular structures that divert this homogenous correlation pattern. Of special interest are the entorhinal cortex and the temporal pole, which were strongly and reliably associated with each other, with the amygdala, with the hippocampus, and with the insula but showed only weak correlations with other lobular structures. However, correlations involving structures with unreliable random slope variances (the majority of regions of frontal and parietal lobe structures) are often broad and often substantially overlap zero, even when the median indicates a large correlation. In the maximum likelihood estimation framework, these models would often lead to non-positive definite covariance matrix warnings, which may then be resolved by setting the random slope variance to zero.

Modeling Considerations

There is a lot of between-subject variability in the size of brain volumes. Separating the between-subject from the within-subject effect is important (Grimm et al., 2016). It can be achieved by separating the age of the participants into the age at the study entry and the difference from the age at study entry to the age at subsequent measurements (e.g., as done in Pfefferbaum et al., 2013). Such an approach was not always applied in previous longitudinal studies, and it seems to be worth considering it in further analyses. Aside from this, the estimation of non-linear trajectories is only reasonable with four or more time points, which highlights the need for longitudinal studies with more time points. In our analyses, we refrained from including a quadratic slope because it would have made the interpretation of the slope parameters more difficult (e.g., a linear and a quadratic random slope may cancel each other out in a person) and because the first and the last measurement were only 4 years apart. Also, four time points are still at the lower limit to appropriately capture person-specific trajectories. We are aware that our rather simple linear models did not capture the true trajectories in great detail, but we expect them to give a reasonable approximation. Model fit indices further indicated no grossly miss-specified models. Further, simpler models are less prone to overfitting. However, more time points and a longer observational time span would clearly be beneficial in many ways.

Measurement Error Term and Image Quality

A large amount of the uncertainty in the random slope variances and covariances is due to the large within-subject error term. This within-subject error consists of true deviations from the assumed linear shape and of the measurement errors related to the structural MRI scanning procedure (e.g., artifacts) and data processing. The measurement errors are probably responsible for the main part of the within-subject variance. Therefore, the reduction of the measurement errors would substantially improve estimation. Motion artifacts and deterioration of scan quality have been proposed to confound estimates attributed to age effects in cross-sectional studies (Alexander-Bloch et al., 2016; Ducharme et al., 2016; Rosen et al., 2018; Klapwijk et al., 2019). Systematic changes in image quality during repeated measures may also confound parameter estimates in longitudinal studies. In addition, inclusion of such motion and scan quality regressors in the models may be able to reduce the measurement error variance. If no such movement data were recorded during T1 acquisition then an approximate movement regressor may be estimated from functional MRI data acquired during the same scanning session (Alexander-Bloch et al., 2016). However, we think that the inclusion of such approximate motion and scan quality regressors are of little help to improve the estimation of the slope parameters. First, these approximate regressors are relatively imprecise. Second, in a longitudinal study, it may be reasonable to assume that individuals show similar motion and rather a stable image quality during repeated measures. Finally, inclusion of time varying regressors that show systematic change may distort the estimates, making interpretation more difficult. An option would be to model these regressors as growth processes and to regress their random intercepts and slopes on the random intercepts and slopes of the brain structures. We tested this approach on some brain structures by using the Euler number as a proxy for image quality (Rosen et al., 2018). Modeling the Euler number as a growth process, it showed decreases over the 4 years and a slope × entry-age interaction (larger decreases with advancing entry-age) with unreliable random slopes. The inclusion of the Euler number as a growth process in the models lead to slightly higher uncertainty but similar covariance estimates as in the reported models. The slope × entry-age interaction estimates were slightly lower in magnitude as in the reported models. Worthy of note, this approach increases model complexity by a large amount and may lead to convergence issues. We conclude from these additional analyses, that the estimates of this study should be taken with caution, and that the uncertainty of these estimates might be even higher than our fitted models propose.

Definition of Healthy Aging

The population of interest in our study was healthy elderly people, where the definition of "healthy" was based on the inclusion criteria. It seems clear that there is still a lot of variability in the term healthy aging. Lifestyle factors (like physical activity) that may influence the volumetric trajectories have not been considered in this analysis. For example, there is some evidence that physical activity positively influences changes in hippocampal volume (Erickson et al., 2011; Varma et al., 2015). Pre-symptomatic Alzheimer pathology may influence trajectories negatively (Fjell et al., 2013). Estimates of brain structure trajectories − especially about the variability and the

associations between random slopes — may critically depend on the definition of healthy aging.

CONCLUSION

Expected volumetric brain region changes, as well as variations and associations of random slopes, have been described in this paper. The focus on random slope variations and associations thereof provided insights into the heterogeneity of the aging process. We think this is valuable information that should not be missed when analyzing such data. Our data present a picture of a rather homogenous aging process of volumetric brain region changes. Subcortical regions, medial temporal structures, and the temporal pole showed reliable random slope variances and a more diverse correlation pattern. Data from further studies with more repeated measures and a longer observational time span would be beneficial.

DATA AVAILABILITY STATEMENT

The R analysis code and the associated data tables will be made available by the authors. The raw MR image data underlying this article are not publicly available and can only be accessed via collaborations with the URPP Dynamics of Healthy Aging.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Kantonale Ethikkommission Zürich (EK 2010-0267/3). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SM, FL, and LJ contributed to the design, set-up, maintenance, and support of the Longitudinal Healthy Aging Brain (LHAB) database. SS performed the statistical analysis and wrote the first draft of the manuscript. All authors contributed to manuscript revision and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00363/full#supplementary-material

REFERENCES

Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., et al. (2016). Subtle in-scanner motion biases automated measurement of brain anatomy from in vivo MRI. *Hum. Brain Mapp.* 37, 2385–2397. doi: 10.1002/hbm.23180

Asparouhov, T., and Muthén, B. (2010a). *Bayesian Analysis Using Mplus: Technical Implementation (Mplus Technical Report)*. Available online at: http://statmodel.com/download/Bayes3.pdf

Asparouhov, T., and Muthén, B. (2010b). *Plausible Values for Latent Variables Using Mplus (Mplus Technical Report)*. Available online at: http://www.statmodel.com/download/Plausible.pdf

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bhaskaran, K., and Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *Int. J. Epidemiol.* 43, 1336–1339. doi: 10.1093/ije/dyu080

Cook, R. D. (1986). Assessment of local influence. *J. R. Stat. Soc. Ser. B* 48, 133–155. doi: 10.1111/j.2517-6161.1986.tb01398.x

Cook, R. D., and Weisberg, S. (1982). Criticism and influence analysis in regression. *Sociol. Methodol.* 13, 313–361. doi: 10.2307/270724

Coupé, P., Catheline, G., Lanuza, E., and Manjón, J. V. (2017). Towards a unified analysis of brain maturation and aging across the entire lifespan: a MRI analysis. *Hum. Brain Mapp.* 38, 5501–5518. doi: 10.1002/hbm.23743

Curran, P. J., Obeidat, K., and Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *J. Cogn. Dev.* 11, 121–136. doi: 10.1080/15248371003699969

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Driscoll, I., Davatzikos, C., An, Y., Wu, X., Shen, D., Kraut, M., et al. (2009). Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology* 72, 1906–1913. doi: 10.1212/WNL.0b013e3181a82634

Du, A. T., Schuff, N., Chao, L. L., Kornak, J., Jagust, W. J., Kramer, J. H., et al. (2006). Age effects on atrophy rates of entorhinal cortex and hippocampus. *Neurobiol. Aging* 27, 733–740. doi: 10.1016/j.neurobiolaging.2005.03.021

Ducharme, S., Albaugh, M. D., Nguyen, T. V., Hudziak, J. J., Mateos-Pérez, J. M., Labbe, A., et al. (2016). Trajectories of cortical thickness maturation in normal brain development - The importance of quality control procedures. *Neuroimage* 125, 267–279. doi: 10.1016/j.neuroimage.2015.10.010

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. U.S.A.* 108, 3017–3022. doi: 10.1073/pnas.1015950108

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., and Walhovd, K. B. (2013). Brain changes in older adults at very low risk for Alzheimer's disease. *J. Neurosci.* 33, 8237–8242. doi: 10.1523/JNEUROSCI.5506-12.2013

Fjell, A. M., and Walhovd, K. B. (2010). Structural brain changes in aging: courses, causes and cognitive consequences. *Rev. Neurosci.* 21, 187–221.

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.* 29, 15223–15231. doi: 10.1523/JNEUROSCI.3252-09.2009

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Fraser, M. A., Shaw, M. E., and Cherbuin, N. (2015). A systematic review and meta-analysis of longitudinal hippocampal atrophy in healthy human ageing. *Neuroimage* 112, 364–374. doi: 10.1016/j.neuroimage.2015.03.035

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotã, M., Chakravarty, M. M., et al. (2017). BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol.* 13:e1005209. doi: 10.1371/journal.pcbi.1005209

Grimm, K. J., Ram, N., and Estabrook, R. (2016). *Growth Modeling: Structural Equation and Multilevel Modeling Approaches, Methodology in the Social Sciences*. New York, NY: Guilford Press.

Grimm, K. J., Ram, N., and Estabrook, R. (2017). *Growth Modeling: Structural Equation and Multilevel Modeling Approaches*. New York, NY: Guilford Press.

Hallquist, M. N., and Wiley, J. F. (2018). MplusAutomation: an R package for facilitating large-scale latent variable analyses in mplus. *Struct. Equ. Model.* 25, 621–638. doi: 10.1080/10705511.2017.1402334

Jäncke, L., Mérillat, S., Liem, F., and Hänggi, J. (2015). Brain size, sex, and the aging brain. *Hum. Brain Mapp.* 36, 150–169. doi: 10.1002/hbm.22619

Jäncke, L., Sele, S., Liem, F., Oschwald, J., and Merillat, S. (2019). Brain aging and psychometric intelligence: a longitudinal study. *Brain Struct. Funct.* 225, 519–536. doi: 10.1007/s00429-019-02005-5

Klapwijk, E. T., van de Kamp, F., van der Meulen, M., Peters, S., and Wierenga, L. M. (2019). Qoala-T: a supervised-learning tool for quality control of FreeSurfer segmented MRI data. *Neuroimage* 189, 116–129. doi: 10.1016/j.neuroimage.2019.01.014

Leonard, C. M., Towler, S., Welcome, S., Halderman, L. K., Otto, R., Eckert, M. A., et al. (2008). Size matters: cerebral volume influences sex differences in neuroanatomy. *Cereb. Cortex* 18, 2920–2931. doi: 10.1093/cercor/bhn052

McCarthy, C. S., Ramprashad, A., Thompson, C., Botti, J. A., Coman, I. L., and Kates, W. R. (2015). A comparison of FreeSurfer-generated data with and without manual intervention. *Front. Neurosci.* 9:379. doi: 10.3389/fnins.2015.00379

Muthén, B. O., and Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. *Psychol. Methods* 2, 371–402. doi: 10.1037/1082-989X.2.4.371 doi: 10.1037/1082-989x.2.4.371

Muthén, L. K., and Muthén, B. O. (1998–2017). Mplus User's Guide, 8th Edn. Los Angeles, CA: Muthén and Muthén.

Narvacan, K., Treit, S., Camicioli, R., Martin, W., and Beaulieu, C. (2017). Evolution of deep gray matter volume across the human lifespan. *Hum. Brain Mapp.* 38, 3771–3790. doi: 10.1002/hbm.23604

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., et al. (2019). Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev. Neurosci.* 31, 1–57. doi: 10.1515/revneuro-2018-0096

Pfefferbaum, A., Rohlfing, T., Rosenbloom, M. J., Chu, W., Colrain, I. M., and Sullivan, E. V. (2013). Variation in longitudinal trajectories of regional brain volumes of healthy men and women (ages 10 to 85years) measured with atlas-based parcellation of MRI. *Neuroimage* 65, 176–193. doi: 10.1016/j.neuroimage.2012.10.008

Potvin, O., Mouiha, A., Dieumegarde, L., and Duchesne, S. (2016). Normative data for subcortical regional volumes over the lifetime of the adult human brain. *Neuroimage* 137, 9–20. doi: 10.1016/j.neuroimage.2016.05.016

Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K. M., and Lindenberger, U. (2010). Trajectories of brain aging in middle-aged and older adults: regional and individual differences. *Neuroimage* 51, 501–511. doi: 10.1016/j.neuroimage.2010.03.020

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, individual differences and modifiers. *Cereb. Cortex* 15, 1676–1689. doi: 10.1093/cercor/bhi044

Rosen, A. F. G., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., et al. (2018). Quantitative assessment of structural image quality. *Neuroimage* 169, 407–418. doi: 10.1016/j.neuroimage.2017.12.059

Storsve, A. B., Fjell, A. M., Tamnes, C. K., Westlye, L. T., Overbye, K., Aasland, H. W., et al. (2014). Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: regions of accelerating and decelerating change. *J. Neurosci.* 34, 8488–8498. doi: 10.1523/jneurosci.0391-14.2014

Varma, V. R., Chuang, Y. F., Harris, G. C., Tan, E. J., and Carlson, M. C. (2015). Low-intensity daily walking activity is associated with hippocampal volume in older adults. *Hippocampus* 25, 605–615. doi: 10.1002/hipo.22397

Vinke, E. J., de Groot, M., Venkatraghavan, V., Klein, S., Niessen, W. J., Ikram, M. A., et al. (2018). Trajectories of imaging markers in brain aging: the Rotterdam Study. *Neurobiol. Aging* 71, 32–40. doi: 10.1016/J.NEUROBIOLAGING.2018.07.001

Walhovd, K. B., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., et al. (2011). Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiol. Aging* 32, 916–932. doi: 10.1016/j.neurobiolaging.2009.05.013

Wang, H., and Song, M. (2011). Ckmeans.1d.dp: optimal k-means clustering in one dimension by dynamic programming. *R J.* 3, 29–33.

Ware, J. E., Kosinski, M., and Keller, S. D. (1995). *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*, 2nd Edn. Boston: Health Institute, New England Medical Center.

Ware, J. E., Kosinski, M., and Keller, S. D. (1996). A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med. Care* 34, 220–233. doi: 10.1097/00005650-199603000-00003

Ziegler, G., Dahnke, R., Jäncke, L., Yotter, R. A., May, A., and Gaser, C. (2012). Brain structural trajectories over the adult lifespan. *Hum. Brain Mapp.* 33, 2377–2389. doi: 10.1002/hbm.21374

Zöllig, J., Mérillat, S., Eschen, A., Röcke, C., Martin, M., and Jäncke, L. (2011). Plasticity and imaging research in healthy aging: core ideas and profile of the International Normal Aging and plasticity imaging center (INAPIC). *Gerontology* 57, 190–192. doi: 10.1159/000324307

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 29 October 2020
doi: 10.3389/fnhum.2020.552111

Differential Brain Activity in Regions Linked to Visuospatial Processing During Landmark-Based Navigation in Young and Healthy Older Adults

Stephen Ramanoël[1,2,3]†, Marion Durteste[1†], Marcia Bécu[1], Christophe Habas[4] and Angelo Arleo[1]*

[1]Sorbonne Université, INSERM, CNRS, Institut de la Vision, Paris, France, [2]Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland, [3]University of Côte d'Azur, LAMHESS, Nice, France, [4]CHNO des Quinze-Vingts, INSERM-DGOS CIC 1423, Paris, France

OPEN ACCESS

Edited by:
Susan Mérillat, University of Zurich, Switzerland

Reviewed by:
Joshua B. Julian, Princeton University, United States
Stephen Duncan Auger, Imperial College London, United Kingdom

***Correspondence:**
Stephen Ramanoël
stephen.ramanoel@univ-cotedazur.fr

†These authors share first authorship

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

Received: *15 April 2020*
Accepted: *22 September 2020*
Published: *29 October 2020*

Citation:
Ramanoël S, Durteste M, Bécu M, Habas C and Arleo A (2020) Differential Brain Activity in Regions Linked to Visuospatial Processing During Landmark-Based Navigation in Young and Healthy Older Adults. Front. Hum. Neurosci. 14:552111. doi: 10.3389/fnhum.2020.552111

Older adults have difficulties in navigating unfamiliar environments and updating their wayfinding behavior when faced with blocked routes. This decline in navigational capabilities has traditionally been ascribed to memory impairments and dysexecutive function, whereas the impact of visual aging has often been overlooked. The ability to perceive visuospatial information such as salient landmarks is essential to navigating efficiently. To date, the functional and neurobiological factors underpinning landmark processing in aging remain insufficiently characterized. To address this issue, functional magnetic resonance imaging (fMRI) was used to investigate the brain activity associated with landmark-based navigation in young and healthy older participants. The performances of 25 young adults (μ = 25.4 years, σ = 2.7; seven females) and 17 older adults (μ = 73.0 years, σ = 3.9; 10 females) were assessed in a virtual-navigation task in which they had to orient using salient landmarks. The underlying whole-brain patterns of activity as well as the functional roles of specific cerebral regions involved in landmark processing, namely the parahippocampal place area (PPA), the occipital place area (OPA), and the retrosplenial cortex (RSC), were analyzed. Older adults' navigational abilities were overall diminished compared to young adults. Also, the two age groups relied on distinct navigational strategies to solve the task. Better performances during landmark-based navigation were associated with increased neural activity in an extended neural network comprising several cortical and cerebellar regions. Direct comparisons between age groups revealed that young participants had greater anterior temporal activity. Also, only young adults showed significant activity in occipital areas corresponding to the cortical projection of the central visual field during landmark-based navigation. The region-of-interest analysis revealed an increased OPA activation in older adult participants during the landmark condition. There were no significant between-group differences in

PPA and RSC activations. These preliminary results hint at the possibility that aging diminishes fine-grained information processing in occipital and temporal regions, thus hindering the capacity to use landmarks adequately for navigation. Keeping sight of its exploratory nature, this work helps towards a better comprehension of the neural dynamics subtending landmark-based navigation and it provides new insights on the impact of age-related visuospatial processing differences on navigation capabilities.

Keywords: healthy aging, spatial navigation, landmark, fMRI, scene-selective regions

INTRODUCTION

The 21st century is characterized by an unprecedented increase in the number of older adults within the worldwide population. There were 703 million people aged 65 years or over in 2019 and this number is projected to more than double by 2050 (United Nations, Departement of Economic and Social Affairs, Population Division, 2019). In parallel, we can expect a significant rise in the prevalence of neurodegenerative diseases such as Alzheimer's and Parkinson's diseases in the older population. To identify appropriate biomarkers of age-related sensori-cognitive alterations, it is critical to gain a better understanding of brain changes in healthy aging. In this context, spatial navigation as a complex behavior encompassing perceptual and cognitive processes provides an ideal framework for the study of normal and pathological aging (Gazova et al., 2012; Lithfous et al., 2013; Allison et al., 2016; Laczó et al., 2017, 2018; Coughlan et al., 2018).

An extensive body of literature has highlighted a robust age-related decline in navigation ability in various species including rodents as well as non-human and human primates (Foster et al., 2012; Lester et al., 2017). Healthy older adults exhibit impairments in their capacity to navigate efficiently, reorient or update their wayfinding behavior when faced with obstacles (Iaria et al., 2009; Moffat, 2009; Harris et al., 2012; Daugherty and Raz, 2017; Merhav et al., 2019). In real-world settings, they are impaired at rapidly acquiring information about their surroundings and are thus slower and more error-prone than young adults when navigating (Kirasic, 1991; Wilkniss et al., 1997). In virtual reality (VR) paradigms older adults also choose inefficient routes, underestimate distances, and make frequent turning errors (Adamo et al., 2012). Cross-sectional studies in VR have shed light on an age-related shift in the use of navigation strategies: older adults favor response over place-based strategies (Bohbot et al., 2012; Rodgers et al., 2012). A place-based strategy involves the formation of mental map-like representations of the absolute position of the goal concerning spatial cues in the environment. A response-based strategy refers to the process whereby an association between a specific stimulus and the goal location is formed. The choice of a navigation strategy critically depends on the visual information present in the environment (Foo et al., 2005; Ratliff and Newcombe, 2008). Indeed, successful navigation requires the perception and the integration of relevant visual-spatial cues such as buildings or monuments, and the binding of these salient elements to directional information (Ekstrom, 2015; Epstein et al., 2017; Julian et al., 2018).

Visual-spatial cues can be salient objects used as navigational landmarks or characteristics about the geometric shape of space (Lester et al., 2017; Bécu et al., 2020). Landmarks can be conceptualized as discrete objects that are independent of the environment's layout, such as a tree or a monument (Epstein and Vass, 2014). Landmarks' size, stability, and proximity to the goal are among the key factors that influence their use for navigation (Stankiewicz and Kalia, 2007; Auger et al., 2012; Auger and Maguire, 2018). Geometric cues encompass all the elements that are intrinsic to and continuous with the external limits of space. These elements include the overall layout, boundaries of the environment, wall lengths, and angle dimensions (Cheng and Newcombe, 2005; Tommasi et al., 2012; Giocomo, 2016). Several studies in virtual environments have emphasized the idea that old age hinders the ability to use landmark information for navigation (Picucci et al., 2009; Harris et al., 2012; Wiener et al., 2012; Zhong and Moffat, 2016; Hartmeyer et al., 2017). More recently, Bécu et al. (2020) extended these findings by unveiling an age-related preference for geometric cues during real-world navigation, when both types of spatial cues were informative.

Despite the extensive body of literature characterizing the neural underpinnings of human spatial navigation (for recent reviews see Chersi and Burgess, 2015; Spiers and Barry, 2015; Epstein et al., 2017; Herweg and Kahana, 2018; Julian et al., 2018), few experiments have explored this question in the context of healthy aging. To date, only 15 neuroimaging studies have focused on spatial processing in normal aging and the majority of these have used structural analyses (Li and King, 2019). These studies, both cross-sectional and longitudinal, have highlighted an age-related decline in place-based navigation associated with structural changes to the hippocampus mainly (Lövdén et al., 2012; Daugherty et al., 2016; Korthauer et al., 2016; Daugherty and Raz, 2017). Age-related changes in other structures of the medial temporal lobe such as the entorhinal cortex, as well as changes in the prefrontal cortices and cerebellum have also been reported (for recent reviews see Lester et al., 2017; Li and King, 2019). Only one cross-sectional functional magnetic resonance imaging (fMRI) study has investigated the link between the use of visual-spatial cues and the navigational skills of young and older adults (Schuck et al., 2015). The authors combined computational modeling and fMRI during a virtual-navigation task to examine how participants learned object locations relative to a circular enclosure or a salient landmark. Young participants used a hippocampal-dependent system for the representation of geometry (circular arena) and a striatal-dependent system for the representation of a landmark (traffic cone). On the other

hand, older participants relied on hippocampal structures for landmark-based navigation and were insensitive to geometric information provided by the environmental boundaries. This absence of reliance on geometric information is surprising considering the behavioral findings mentioned above, and it could be related to the small field of view inside the scanner (Sturz et al., 2013).

Several other brain regions, known to be altered in healthy aging (Lester et al., 2017; Zhong and Moffat, 2018), have also been identified as crucial for visuospatial processing during navigation (Epstein and Vass, 2014; Julian et al., 2018). Recently, there has been growing interest in unearthing the roles of the parahippocampal place area (PPA), the occipital place area (OPA), and the retrosplenial cortex (RSC). These regions are all involved in the processing of visual scenes such as landscapes or urban environments. In particular, they have been speculated to integrate incoming visual inputs with higher-level cognitive processes (for reviews see Epstein et al., 2017; Julian et al., 2018). The PPA is sensitive to navigationally relevant cues (Janzen and van Turennout, 2004; Epstein, 2008) and it may be implicated in the recognition of spatial contexts (Marchette et al., 2015). The OPA has been associated with the processing of local elements in scenes (Kamps et al., 2016) as well as with the representation of environmental boundaries (Julian et al., 2016). The RSC is suggested to play a role in anchoring heading information to local visual cues (see Mitchell et al., 2018 for a recent review of RSC functions). The exploration of the neural activity of these scene-selective regions in the context of aging has just begun. Nevertheless, some evidence exists about functional changes in the PPA and RSC of older adults that have been linked to impaired processing of visual scenes and difficulties in switching between navigation strategies, respectively (Ramanoël et al., 2015; Zhong and Moffat, 2018). The impact of aging on the OPA has not been fully characterized yet, but recent findings have hinted at its preserved connectivity with other navigational brain structures in healthy aging (Ramanoël et al., 2019).

Thus, although several behavioral studies have provided evidence for differential use of landmarks across the lifespan, there is a lack of knowledge on the functional and neurobiological factors responsible for the deterioration of landmark information processing in older age. To address this caveat, the present study used fMRI to investigate to what extent healthy aging influences behavior and neural activity associated with landmark-based navigation. A second objective consisted in deciphering the role played by scene-selective regions (namely, PPA, OPA, RSC) in age-related landmark-based navigation deficits as these areas appear to be critical for the integration of relevant visual information for navigation.

MATERIALS AND METHODS

Participants

Overall, 25 young adults (18 males, seven females) and 21 older adults (eight males, 13 females) completed the experiment, but four older adults were excluded: two (females) for a lack of task understanding and two (one male, one female) for in-scanner motion (movements >5 mm across trials). Thus, 25 young adults (18 males, seven females; 25.4 ± 2.7 years) and 17 older adults (seven males, 10 females; 73.0 ± 3.9 years) were included in the analyses. The participants were part of the French cohort study *SilverSight* (∼350 subjects) established in 2015 at the Vision Institute, Quinze-Vingts National Ophthalmology Hospital, Paris (Lagrené et al., 2019). The battery of clinical and functional examinations used to enroll participants comprised an ophthalmological and functional visual screening, a neuropsychological evaluation, an oculomotor screening, an audio-vestibular assessment as well as a static/dynamic balance examination. The neuropsychological evaluation included the Mini-Mental State Examination (MMSE; Folstein et al., 1975) and computerized versions of the 3D mental rotation test (Vandenberg and Kuse, 1978), perspective-taking test (Kozhevnikov and Hegarty, 2001), and forward and backward spans of the Corsi block-tapping task (Corsi, 1973). Enrolled older participants had a score of 24[1] or higher on the MMSE. All subjects were right-handed, they had a normal or corrected-to-normal vision, and they had no history of neurological or psychiatric disorders. Centration measurements and acuity were evaluated at least 2 weeks before the experimental session to order MRI-compatible glasses for participants requiring visual correction (manufactured by Essilor International). Participants gave their written informed consent to participate in the study. All screening and experimental procedures were compliant with the tenets of the Declaration of Helsinki, and they were approved by the Ethical Committee "CPP Ile de France V" (ID_RCB 2015-A01094-45, CPP N°: 16122).

Virtual Navigation Task and Experimental Protocols

The Virtual Navigation Task

The virtual navigation task was displayed on an MRI-compatible liquid crystal display monitor (NordicNeuroLab, Bergen, Norway) positioned at the head of the scanner bore. Participants viewed the screen [size: 69.84 cm (H) × 39.26 cm (V); pixels: 1,920 × 1,080] at a distance of 115 cm *via* a mirror fixed above the head-coil. The visible part of the screen subtended approximately 34 × 20° of visual angle.

The virtual environment was programmed with the Unity3D game engine (Unity Technologies SF; San Francisco, CA, USA[2]) and it allowed participants to navigate actively from a first-person perspective. The virtual environment was a three-arm maze (Y-maze) consisting of three corridors radiating out from a center delimited by homogenous wooden-like walls. Two configurations were designed. In the *landmark condition*, all arms were 18 virtual meters (vm) long and equiangular. Three 3D light gray-colored objects (a square, a triangle, and a circle) were placed in between the arms at the center of the maze (**Figure 1A**).

[1]All older participants scored 28 or above on the MMSE except one participant who scored 24. We nonetheless decided to include this subject as the extended neuropsychological evaluation was within normal range and no significant changes were detected when removing the participant from the fMRI analyses.
[2]https://unity.com/

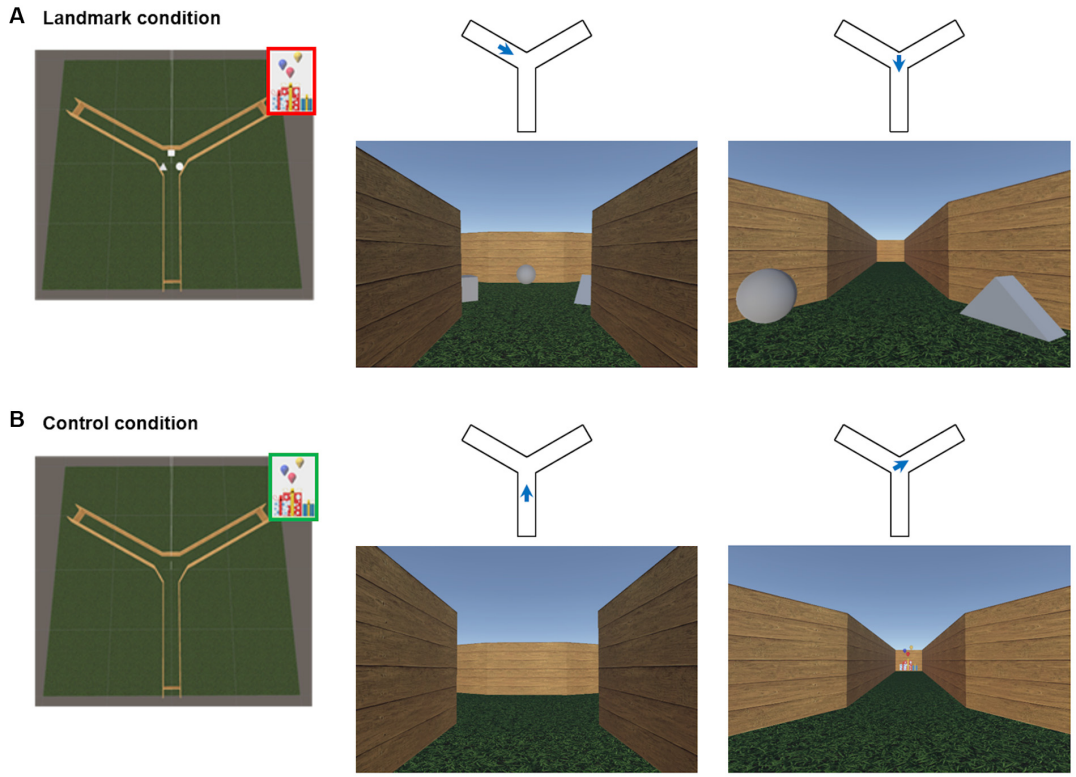


FIGURE 1 | The virtual environment. **(A)** An overhead perspective of the environment for the landmark condition and two example views representing a first-person perspective within the maze. Blue arrows represent the position and the orientation associated with the example views in the landmark condition. **(B)** An overhead perspective of the environment for the control condition and example views within the maze. Blue arrows represent the position and the orientation associated with the example views in the control condition. Red: hidden goal; Green: visible goal. The aerial view was never seen by participants.

In the *control condition*, the arms were still 18 vm long and equiangular, but the maze was devoid of objects (**Figure 1B**).

Participants navigated actively through the virtual environment with an MRI-compatible ergonomic two-grip response device (NordicNeuroLab, Bergen, Norway). They could move forward (thumb press), turn right (right index press), and turn left (left index press). A single finger press was necessary to initiate or stop the movement. The forward speed of movement was set at 3 vm/s and the turning speed at 40°/s.

The Spatial Navigation Paradigm

Before scanning, all participants familiarized themselves with the response device in an unrelated virtual space both outside and inside the scanner. They were required to navigate within a square open-field environment and to walk over a wooden board that appeared at different locations.

The scanning session during the navigation task was divided into three runs: an encoding phase and a retrieval phase for the landmark condition and a control condition. At the beginning of the *encoding phase*, participants were positioned at the center of the maze randomly facing one of the three arms. They were instructed to find a goal (gifts) hidden at the end of one corridor and to remember its location using the visual information available in the center of the environment (i.e., the three 3D light gray-colored objects; **Figure 1A**). The goal appeared when the subject arrived at the correct location. The encoding phase lasted 3 min to ensure that participants could explore all corridors. Then, the *retrieval phase* began and it consisted of seven trials. In each trial, participants were placed at the end of a non-rewarded corridor with their back against the wall. The starting positions were pseudo-randomized across both trials and subjects. Participants were asked to navigate to the previously encoded goal location. Upon arrival at the end of the correct arm, the gifts appeared to indicate successful completion of the trial, and a fixation cross on a gray screen was presented for an inter-trial interval of 3–8 s. Finally, the control condition began and it included four trials. At the beginning of each control trial, subjects started at the end of an arm and moved to the center of the maze. The target was readily visible from the center and participants were instructed to navigate towards it. The control condition was designed to account for potential confounding factors such as motor and simple perceptual aspects of the task.

A short debriefing phase concluded the experimental session. Participants were probed on the strategy they used to orient in

the landmark condition. They were asked to report how they solved the task: (i) using one object; (ii) using at least two objects; (iii) randomly; and (iv) another strategy. Participants were deemed to be using a *place-based* strategy when their decision was based on two landmarks or more and to be using a *response-based* strategy when their decision was based on a single visual-spatial cue (Iaria et al., 2003; Iglói et al., 2010, 2015; Chrastil, 2013; Gazova et al., 2013; Packard and Goodman, 2013; Colombo et al., 2017; Laczó et al., 2017). No participants answered that they oriented randomly or that they used a different strategy.

Functional Localizer Experiment

Following the spatial navigation task, a block fMRI paradigm similar to that used by Ramanoël et al. (2019) was used to locate the scene-selective regions: the PPA, the OPA, and the RSC. Participants were presented with blocks of 900 × 900-pixel grayscale photographs (18 × 18° of visual angle) representing scenes, faces, everyday objects, and scrambled objects. The functional run lasted 4 min 40 s and it was composed of 14 20-s task blocks (four blocks of scenes, two blocks of faces, two blocks of objects, two blocks of scrambled objects, and four blocks of fixation). Each stimulus was presented for 400 ms followed by a 600 ms inter-stimulus interval. Participants performed a "one-back" repetition detection task.

Statistical Analyses of Behavioral Data

During all phases of the virtual-navigation task, the trial durations (to calculate the average navigation time, i.e., the mean time to reach the goal), the grip responses (i.e., the number of times each button of the MRI-compatible two-grip response device was pressed), and the error rate (i.e., the number of times a participant chose the wrong corridor across the seven trials) were recorded. These behavioral measures were used to quantify the spatial navigation performance of subjects, which was compared across age group and sex. The normality of the behavioral data was assessed graphically with quantile-quantile plots and numerically with the Shapiro–Wilk test. Descriptive characteristics and behavioral measures were compared using independent samples *t*-test for normally distributed continuous data, Mann–Whitney *U* tests for non-normally distributed continuous data, and chi-square test of independence for categorical data. To investigate the potential influence of strategy use on navigation performance, simple logistic regressions were conducted in both age groups. Also, the links between sex and strategy use were examined with Fisher's exact test. Finally, associations between neuropsychological test scores and behavioral performance were computed using simple linear regression analyses with a statistical threshold set at $p < 0.005$ after adjusting for multiple comparisons [$p = 0.05/(5 \times 2)$].

Acquisition and Statistical Analyses of MRI Data

MRI Acquisition

Data were collected using a 3 Tesla Siemens MAGNETOM Skyra whole-body MRI system (Siemens Medical Solutions, Erlangen, Germany) equipped with a 64-channel head coil at the Quinze-Vingts National Ophthalmology Hospital in Paris, France. T2*-weighted echo-planar imaging (EPI) sequences, optimized to minimize signal dropout in the medial temporal region (Weiskopf et al., 2006), were acquired for functional imaging during the navigation task (voxel size = 3 × 3 × 2 mm, TR/TE/flip angle = 2685 ms/30 ms/90°, interslice gap = 1 mm, slices = 48, matrix size = 74 × 74, FOV = 220 × 220 mm). For the localizer experiment, 284 volumes from 64 slices were acquired using a T2*-weighted simultaneous multi-slice echo planar sequence (SMS-EPI; voxel size = 2.5 × 2.5 × 2.4 mm, TR/TE/flip angle = 1,000 ms/30 ms/90°, matrix size = 100 × 100, SMS = 2, GRAPPA = 2). The anatomical volume consisted of a T1-weighted, high-resolution three-dimensional MPRAGE sequence (voxel size = 1 × 1 × 1.2 mm, TR/TE/IT/flip angle = 2,300 ms/2.9 ms/900 ms/9°, matrix size = 256 × 240 × 176).

Whole-Brain Analyses

FMRI data analysis was performed using a combination of SPM12 release 7487 (Wellcome Department of Imaging Neuroscience, London, UK) and ArtRepair toolbox (Mazaika et al., 2009) implemented in MATLAB 2015 (Mathworks Inc., Natick, MA, USA). The first five functional volumes of the encoding, retrieval, and control runs were discarded to allow for equilibration effects. A slice-timing correction was applied and functional images were realigned to the mean functional image using a rigid body transformation. Artifacts related to the motion were then examined with ArtRepair. Volumes displaying elevated global intensity fluctuation (>1.3%) and movement exceeding 0.5 mm/TR were repaired using interpolation from adjacent scans. The T1-weighted anatomical volume was then realigned to match the mean functional image of each participant and normalized to the Montreal Neurological Institute (MNI) space using a 4th-degree B-Spline interpolation. The anatomical normalization parameters were subsequently used for the normalization of functional volumes. Each functional scan was smoothed with an 8 mm FWHM (Full Width at Half Maximum) Gaussian kernel. The preprocessed images were visually inspected to ensure that there were no realignment or normalization issues. Statistical analysis was performed using the general linear model for block design at the single-participant level (Friston et al., 1995). The seven trials of the retrieval phase in the landmark condition, the four trials of the control condition, and fixation times were modeled as regressors, constructed as box-car functions, and convolved with the SPM hemodynamic response function (HRF). The encoding phase was not included in the analysis as the time taken to find the goal for the first time differed greatly both within and between age groups. Time to reach the goal by trial, grip responses during navigation, and movement parameters derived from the realignment correction (three translations and three rotations) were entered in the design matrix as regressors of no-interest. The time series for each voxel was high-pass-filtered (1/128 Hz cutoff) to remove low-frequency noise and signal drift. Individual contrasts were submitted to multiple regression and a two-samples *t*-test. The main contrast of

interest for all analyses was [Landmark > Control]. Sex and total brain volume (gray and white matter) were included as covariates in the regression and the total brain volume was included as a covariate in the two-samples *t*-test (see "Behavioral Results" section). Areas of activation were tested for significance using a statistical threshold of $p < 0.001$ uncorrected at voxel-level, with a minimum cluster extent of $k = 10$ voxels (Iglói et al., 2010, 2015; Sutton et al., 2010; Schuck et al., 2015; Javadi et al., 2017).

Of note, a control analysis that excluded all error trials was performed to identify the potential impact of errors on variability in brain activity. No significant change to the fMRI results was observed after removing error trials.

Region-of-Interest Analyses

Data from the localizer experiment were analyzed using SPM12. For each participant, the first four functional localizer volumes were discarded and the remaining images were realigned, co-registered to the T1-weighted anatomical image, normalized to the MNI space, and smoothed using an 8 mm FWHM Gaussian kernel. The slice-timing correction was not applied, following recommendations from the Human Connectome Project functional preprocessing pipeline for multi-slice sequences (Glasser et al., 2013). The localizer images were analyzed using a single participant general linear model for block design. Five categories of interest (scenes, faces, objects, scrambled objects, fixation) were modeled as five regressors and convolved with a canonical HRF. Movement parameters were included in the model as regressors of no interest and each voxel's time-series was high-pass-filtered (1/128 Hz cutoff).

PPA, OPA, and RSC were located independently for each participant using the fMRI contrast [Scenes > (Faces + Objects)]. Significant voxel clusters on individual *t*-maps were identified using family-wise error correction (FWE) for multiple comparisons ($\alpha = 0.05$, *t*-value > 4.8). Mask ROIs were created as the 40 contiguous voxels with the highest *t*-values around the peaks of activation from the left and right hemispheres. The two 40-voxel regions from each hemisphere were subsequently summed into a single 80-voxel ROI. Mean parameter estimates were extracted from the three mapped ROIs using the REX MATLAB-based toolkit. Analogously to the whole-brain analyses, the main contrast of interest for the ROI analyses was [Landmark > Control].

RESULTS

Behavioral Results

The navigation performances across age groups are presented in **Figure 2**. Older subjects were significantly slower to reach the goal than younger subjects (i.e., longer navigation time) in the landmark condition (19.85 s ± 1.67 vs. 11.97 s ± 0.13; $U_{(40)} = 418$, $p = 10^{-6}$, $r = 0.81$, 95% confidence interval (CI) of the difference [3.40, 8.36]; **Figure 2Ai**) and in the control condition (20.83 s ± 0.95 vs. 14.07 s ± 0.27; $U_{(40)} = 418$, $p = 10^{-6}$, $r = 0.81$, 95% CI of the difference [4.70, 8.15]; **Figure 2Aii**). In addition, older adults made more navigation errors than young adults (mean ± SEM: 1.0 ± 0.41 vs. 0 ± 0.00; $U_{(40)} = 125$, $p = 0.0007$, $r = 0.44$, 95% CI of the difference [0.00, 0.00]). On average, older adults chose the wrong corridor in 10% of trials (**Figure 2B**). No sex effect was observed on navigation time in each age group separately. However, when data from both age groups were pooled, women's navigation time was significantly longer than men's (17.75 s ± 1.88 s vs. 13.40 s ± 0.64 s; $U_{(40)} = 2.29$, $p = 0.022$, $r = 0.35$, 95% CI of the difference [0, 15, 7.05]). Sex and total intracranial volume were therefore included as covariates in the fMRI multiple regression analyses. The post-scanner debriefing phase revealed a significant difference in the type of navigational strategy used between age groups ($X^2_{(1,N=42)} = 4.06$, $p = 0.044$, $\phi = 0.67$). Indeed, older adults relied less on place strategies during landmark-based navigation than younger adults (**Figure 2C**). However, in neither age group did strategy use predict navigation performance. Finally, there was no sex effect on strategy use.

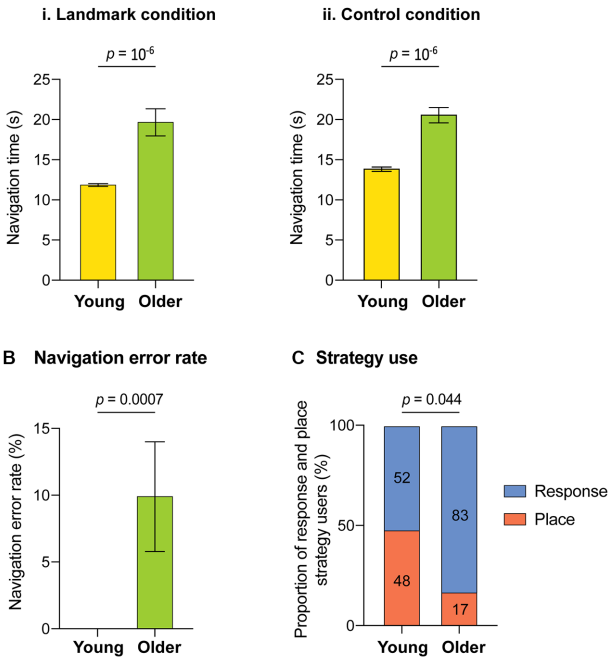


FIGURE 2 | Behavioral results for the virtual-navigation task across age groups. **(A)** Time taken to reach the goal (navigation time): **(i)** averaged across seven trials in the landmark condition, **(ii)** averaged across four trials in the control condition. **(B)** Time taken to reach the goal averaged across seven trials in the landmark condition (navigation time). **(C)** The proportion of participants who used a place-based or response-based strategy in the landmark condition (strategy use). Error bars represent the standard errors of the mean.

Neuropsychological assessments showed that older adults had significantly poorer performance than young adults across all measures. Descriptive and cognitive characteristics are summarized in **Table 1**. Finally, longer navigation time was associated with poorer performance on the forward span of the Corsi block-tapping task ($R^2 = 0.31$, $p = 0.0001$, $f^2 = 0.45$), on the backward span of the Corsi block-tapping task ($R^2 = 0.19$,

TABLE 1 | Descriptive characteristics and cognitive performance of young and older participants.

	Groups Mean (± SEM)				
Sex (M/F)	**Young 18/7**	**Older 7/10**	***p*-value**	**ES***	**95% CI of the difference**
Age[1]	25.4 (± 0.5)	73.0 (± 0.9)	$p < 0.001$	14.8	[45.6, 49.7]
Total brain volume[1] (cm^3)	1301 (± 18)	1061 (± 23)	$p < 0.001$	−2.67	[−297.6, −183.3]
MMSE[2]	30.0 (± 0.0)	28.8 (± 0.4)	$p < 0.001$	−0.61	[−2.0, −0.0]
3D mental rotation[1]	18.3 (± 0.9)	12.7 (± 1.2)	$p < 0.001$	−1.20	[−8.8, −2.7]
Corsi forward[2]	7.2 (± 0.2)	4.4 (± 0.2)	$p < 0.001$	−0.80	[−4.0, −2.0]
Corsi backward[2]	6.2 (± 0.3)	4.6 (± 0.2)	$p < 0.001$	−0.54	[−2.0, −1.0]
Perspective taking[2]	15.3 (± 1.7)	46.1 (± 6.7)	$p < 0.001$	0.65	[16.8, 35.7]

*M, male; F, female; SEM, standard error of the mean; ES, effect size (*Hedges' g was computed for independent samples t-tests and r was computed for Mann–Whitney U tests); CI, confidence interval (95% CI of the difference between means was computed for independent samples t-tests and 95% CI of the difference between medians for Mann–Whitney U tests); MMSE, mini-mental state examination; [1]independent samples t-test; [2]Mann–Whitney U test.*

$p = 0.004, f^2 = 0.23$) and on the perspective-taking test ($R^2 = 0.55$, $p = 10^{-8}, f^2 = 1.22$). A significant association was also reported between the number of errors during the virtual-navigation task and performance on the perspective-taking test ($R^2 = 0.58$, $p = 10^{-9}, f^2 = 1.38$).

Whole-Brain fMRI Results

Multiple Regression Analyses

The brain regions related to navigation performance were located by examining the association between both groups' navigation time, age, and patterns of brain activity for the fMRI contrast [Landmark > Control]. A negative association was observed between navigation time and neural activity in multiple clusters across the brain (**Table 2** and **Figure 3**) including frontal (right superior and middle gyri), temporal (middle, inferior, lingual, and parahippocampal gyri), parietal (angular gyrus including the inferior parietal lobule), and occipital cortices (left superior occipital gyrus) as well as the cerebellum (lobule VI and vermis). Temporal activations in the left hemisphere (LH) comprised the posterior part of the hippocampus ($x = -24$, $y = -46$, $z = 5$). Other activations included the ventral temporal cortex ($x = 45$, $y = 8, z = -37$) and the visual area V3A ($x = -21, y = -97, z = 23$),

TABLE 2 | Cerebral regions whose activity for the contrast [Landmark > Control] was predicted by navigation time across all participants and across age groups (sex and intracranial volume were included as covariates).

	H	BA	*k*	*x*	*y*	*z*	*t*	R^2	ES [95% CI]
Multiple Regression Navigation time/[Landmark > Control]									
Inferior temporal gyrus	R	21	40	45	-22	−16	5.08	0.33	0.24 [0.16, 0.32]
				51	−31	−13	4.80		0.28 [0.19, 0.38]
Lingual gyrus	L	-	27	−30	−58	5	4.69	0.14	0.40 [0.26, 0.54]
				−24	−46	5	4.66		0.46 [0.30, 0.62]
Lingual gyrus	R	-	15	30	−49	2	4.64	0.19	0.54 [0.35, 0.73]
[Middle temporal gyrus]				42	−49	5	3.65		0.22 [0.12, 0.32]
Middle frontal gyrus	R	9	35	18	44	26	4.59	0.09	0.19 [0.13, 0.26]
Cerebellum Vermis	R	-	67	3	−55	−4	4.59	0.15	0.39 [0.25, 0.53]
				3	−43	−13	3.92		0.40 [0.24, 0.57]
Middle temporal gyrus	R	20	10	45	8	−37	4.56	0.29	0.23 [0.15, 0.32]
Parahippocampal gyrus	L	19	13	−33	−40	−4	4.47	0.16	0.36 [0.23, 0.49]
Middle frontal gyrus	R	10	23	15	65	20	4.37	0.09	0.65 [0.41, 0.90]
[Superior frontal gyrus]				6	68	14	3.75		0.61 [0.34, 0.88]
Cerebellum lobule VI	R	-	13	27	−55	−31	4.26	0.24	0.20 [0.13, 0.28]
Superior occipital gyrus	L	18	11	−21	−97	23	4.26	0.11	0.50 [0.31, 0.70]
Angular gyrus	L	39	53	−54	−61	29	4.01	0.22	0.53 [0.31, 0.75]
				−45	−79	35	4.00		0.37 [0.22, 0.52]
				−54	−70	26	3.94		0.30 [0.18, 0.43]
Angular gyrus	R	39	13	45	−58	32	4.01	0.11	0.51 [0.30, 0.72]
Superior frontal gyrus	L	8	10	−9	35	47	3.97	0.13	0.22 [0.13, 0.31]
Multiple Regression Age/[Landmark > Control]									
Superior temporal gyrus	L	22	31	−45	−1	−16	5.61	0.26	0.16 [0.12, 0.21]
		38		−51	8	−16	3.54		0.08 [0.05, 0.12]
Brainstem	R	-	17	3	−7	−1	5.40	0.11	0.20 [0.14, 0.26]

The statistical threshold was defined as $p < 0.001$ uncorrected for multiple comparisons at voxel-level with an extent voxel threshold set at 10 voxels. For each cluster, the region with the maximum t-value is listed first and other regions in the cluster are listed [in square brackets]. Montreal Neurological Institute (MNI) coordinates (x, y, z) of the peak activation and the number of voxels (k) in a cluster are also shown. All values including t-values, effect sizes, and 95% confidence intervals were provided for the peak voxel coordinate but the R^2 values were provided for the entire cluster. H, hemisphere; R, right hemisphere; L, left hemisphere; BA, Brodmann area; ES, effect size; CI, confidence interval.

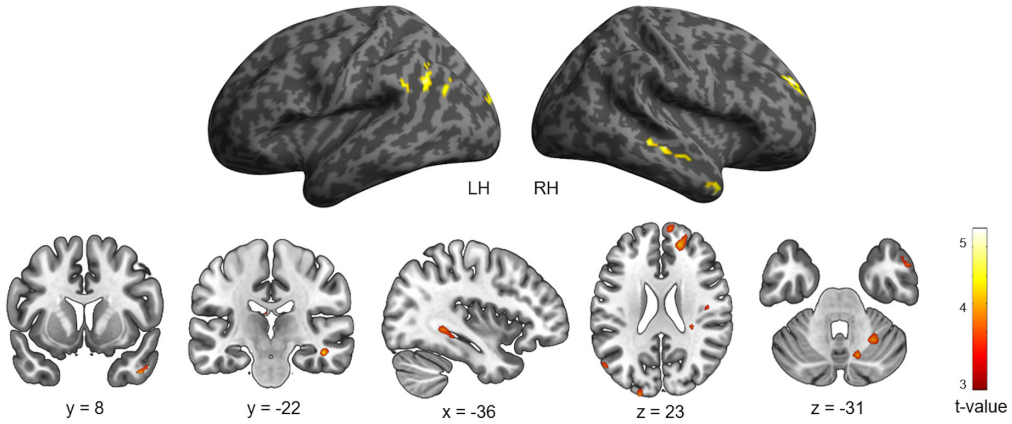


FIGURE 3 | Cerebral regions whose activity for the contrast [Landmark > Control] was predicted by navigation time projected onto 3D inflated anatomical templates and 2D slices ($p < 0.001$ uncorrected, $k = 10$ voxels). LH, left hemisphere; RH, right hemisphere.

TABLE 3 | Cerebral regions whose activity for the contrast [Landmark > Control] was elicited by within-group or between-group analyses (total intracranial volume was included as a covariate).

Group analyses [Landmark > Control]		H	BA	*k*	*x*	*y*	*z*	*t*	ES [95% CI]
Within-group									
[Young]	Inferior Occipital Gyrus	R	18	65	30	−88	−10	5.07	4.49 [3.04, 5.95]
	Superior Temporal Gyrus	L	38	21	−48	20	−25	4.54	2.55 [1.63, 3.47]
					−42	11	−22	4.01	2.52 [1.49, 3.55]
	Cerebellum Crus I-II	R	-	19	33	−82	−34	4.49	2.75 [1.74, 3.76]
	Middle Occipital Gyrus	L	18	37	−30	−97	−7	4.42	3.38 [2.12, 4.64]
	[Inferior Occipital Gyrus]		19		−39	−85	−13	3.54	3.25 [1.74, 4.76]
	Inferior Temporal Gyrus (Amygdala/Hippocampus)	L	53/54	15	−30	−1	−22	4.37	1.43 [0.89, 1.97]
	Middle Occipital Gyrus	R	19	16	48	−79	2	4.20	1.85 [1.13, 2.57]
[**Older**]	*No significant activation*								
Between-group									
[**Young > Older**]	Inferior Temporal Gyrus	L	-	12	−33	2	−25	3.86	2.49 [1.43, 3.56]
	(Amygdala/Hippocampus)				−36	−7	−25	3.82	2.13 [1.21, 3.05]
[**Older > Young**]	*No significant activation*								

The statistical threshold was defined as $p < 0.001$ uncorrected for multiple comparisons at voxel-level with an extent voxel threshold set at 10 voxels. For each cluster, the region with the maximum t-value is listed first and other regions in the cluster are listed [in square brackets]. Montreal Neurological Institute (MNI) coordinates (x, y, z) of the peak activation and the number of voxels (k) in a cluster are also shown. H, hemisphere; R, right hemisphere; L, left hemisphere; BA, Brodmann area; ES, effect size; CI, confidence interval.

which is part of the dorsal visual stream. The inverse association did not elicit any significant brain activations. Complementary analyses revealed a negative association between age and patterns of neural activity in the left superior temporal gyrus ($x = -45$, $y = -1$, $z = -16$) and the brainstem ($x = 3$, $y = -7$, $z = -1$) for the fMRI contrast [Landmark > Control; **Table 2**].

Two-Sample Analyses

First, the brain activity elicited by the contrast [Landmark > Control] was examined in each age group individually. In a second step, between-group analyses were conducted to compare activity between young and older participants. Results for the within-group and between-group analyses are shown in **Table 3** and **Figure 4**. For young participants, significant clusters were found in the superior and inferior temporal gyri of the LH which included the amygdala and hippocampus, in the middle and inferior occipital gyri bilaterally, and in the right cerebellum (Crus I-II). In the older participant group, no significant activation was elicited by the fMRI contrast [Landmark > Control]. Through the direct comparisons between age groups, significant results were found for the fMRI contrast [Young > Older] but not for the contrast [Older > Young]. Young participants exhibited stronger activity in the left inferior temporal gyrus, comprising the amygdala and hippocampal regions, compared to older subjects.

The absence of activation in the older adult group prompted us to conduct further exploratory analyses. First, the cognitive cost of the control condition was investigated both between and within-age groups using the fMRI contrast [Control > Fixation] with a conservative cluster-level FWE correction $p < 0.05$. The latter statistical threshold is applied here because the navigational condition is not compared to an active control condition (i.e., that involves virtual spatial navigation). No significant activation was found for the group comparison

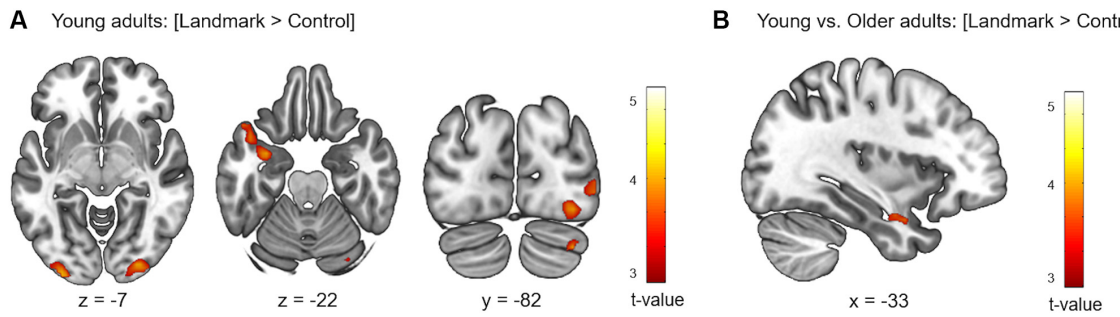


FIGURE 4 | Cerebral regions whose activity for the contrast [Landmark > Control] was elicited by within- **(A)** or between-group **(B)** analyses projected onto 2D slices ($p < 0.001$ uncorrected, $k = 10$ voxels).

TABLE 4 | Cerebral regions whose activity for the contrast [Control > Fixation] was elicited by between-group analyses (total intracranial volume was included as a covariate).

Group analyses [Control > Fixation]		H	BA	*k*	*x*	*y*	*z*	*t*	ES [95% CI]
[Young > Older]	*No significant activation*								
[Older > Young]	Superior Frontal Gyrus	L	-	30	−27	32	56	5.56	3.18 [2.06, 4.30]
	[Middle Frontal Gyrus]		8		−36	23	56	4.58	2.44 [1.39, 3.48]

The statistical threshold was defined as $p < 0.05$ FWE-corrected for multiple comparisons at cluster-level with an extent voxel threshold set at 10 voxels. For each cluster, the region with the maximum t-value is listed first and other regions in the cluster are listed [in square brackets]. Montreal Neurological Institute (MNI) coordinates (x, y, z) of the peak and number of voxels (k) of clusters are also shown. H, hemisphere; R, right hemisphere; L, left hemisphere; BA, Brodmann area; ES, effect size; CI, confidence interval.

TABLE 5 | Cerebral regions whose activity for the contrast [Landmark > Fixation] was elicited by between-group analyses (total intracranial volume was included as a covariate).

Group analyses [Landmark > Fixation]		H	BA	*k*	*x*	*y*	*z*	*t*	ES [95% CI]
[Young > Older]	*No significant activation*								
[Older > Young]	Middle Frontal Gyrus	R	10	22	24	47	−1	6.06	0.95 [0.69, 1.21]
	Angular Gyrus	L	7	223	−30	−64	47	5.66	2.43 [1.72, 3.14]
	[Superior Parietal Gyrus]		39		−33	−55	44	5.59	2.26 [1.59, 2.93]
	[Supramarginal Gyrus]		40		−48	−43	44	5.32	1.61 [1.11, 2.11]
	Middle Frontal Gyrus	R	10	34	39	38	17	5.37	1.51 [1.05, 1.97]
					33	47	20	4.47	1.40 [0.88, 1.92]
	Cerebellum	R	-	23	36	−73	−22	5.19	2.90 [1.98, 3.82]
	Middle Frontal Gyrus	L	6	127	−42	5	41	5.14	1.78 [1.21, 2.35]
			9		−45	26	26	5.13	1.60 [1.09, 2.11]
			6		−45	5	50	5.12	1.73 [1.18, 2.29]
	Superior Parietal Gyrus	R	39	36	30	−55	44	4.89	1.87 [1.24, 2.50]
	[Angular Gyrus]		7		30	−64	53	4.64	3.34 [2.16, 4.53]

The statistical threshold for the cluster was defined as $p < 0.05$ FWE-corrected for multiple comparisons at the cluster-level with an extent voxel threshold set at 10 voxels. For each cluster, the region with the maximum t-value is listed first and other regions in the cluster are listed [in square brackets]. Montreal Neurological Institute (MNI) coordinates (x, y, z) of the peak and number of voxels (k) of clusters are also shown. H, hemisphere; R, right hemisphere; L, left hemisphere; BA, Brodmann area; FEW, family-wise error; ES, effect size; CI, confidence interval.

[Young > Older]. However, by using the inverse group comparison [Older > Young] activations were revealed in the left superior and middle frontal gyri (**Table 4**). Within-group analyses revealed a widespread pattern of activity in young adults encompassing the superior occipital, superior frontal, and superior parietal gyri. In older adults, similar activations were noted along with some lingual and precentral activity (**Supplementary Table 1**). Of note, we observed smaller clusters of activation in the older adult group compared with the younger adult group. Secondly, between- and within-group analyses using the fMRI contrast [Landmark > Fixation] were performed with the same cluster-level FWE correction $p < 0.05$. Significant activations were reported for the [Older > Young] comparison only, and they included the middle frontal and angular gyri in each hemisphere as well as the right cerebellum (**Table 5**). Concerning the within-group analyses, young participants displayed activity in the middle frontal, superior parietal, and occipital gyri while older participants showed

activations of the superior and middle frontal gyri, superior parietal gyrus, precentral gyrus, fusiform gyrus, and cerebellum (**Supplementary Table 2**).

Association Between Neuropsychological Evaluation and fMRI Activity

Exploratory regression analyses were also conducted to test the associations between cognitive scores and brain activations for the contrast [Landmark > Control] across all participants. Significant associations were found for the 3D mental rotation test, the forward span of the Corsi block-tapping task, and the perspective-taking test. Specifically, higher scores on the 3D mental rotation test were correlated with enhanced activity in the brainstem ($x = 9$, $y = -25$, $z = -22$; effect size = 0.20, 95% CI [0.13, 0.27]; $R^2 = 0.05$). Better visuo-spatial working memory was associated with increased activity in the left parahippocampal gyrus and the posterior part of the hippocampus ($x = -24$, $y = -31$, $z = 16$; 1.06 [0.77, 1.36]; $R^2 = 0.17$). Finally, perspective-taking ability was associated with activation of the superior frontal gyrus bilaterally (left: $x = -9$, $y = -10$, $z = 83$; 0.21 [0.14, 0.28]; $R^2 = 0.13$; right: $x = 15$, $y = -13$, $z = 83$; 0.26 [0.17, 0.35]; $R^2 = 0.13$), the left caudate nucleus ($x = -18$, $y = 14$, $z = 23$; 0.07 [0.05, 0.10]; $R^2 = 0.29$), and the brainstem ($x = 3$, $y = -25$, $z = -46$, 0.21 [0.13, 0.29]; $R^2 = 0.21$). Of note, effect sizes (ESs) and 95% CIs were computed for the peak voxel coordinate but the R^2 values were computed for the entire cluster.

Region-of-Interest Results

The PPA, OPA and RSC ROIs were defined for each individual based on the independent localizer experiment. First, age-related effects were examined in average parameter activity for the fMRI contrast [Landmark > Control]. No significant differences were found. It is nonetheless interesting to note the presence of OPA activity in older adults only (1.40 ± 1.20 vs. −0.18 ± 0.44, $p = 0.234$; **Supplementary Figure 1**). Considering the latter result and the age-related differences observed during the control condition, further exploratory analyses were performed using the fMRI contrasts [Landmark > Fixation] and [Control > Fixation]. A three-way ANOVA was conducted with scene-selective region, age and condition as factors using the fMRI contrasts [Landmark > Fixation] and [Control > Fixation] (**Figures 5A,B**). Scene-selective region ($F_{(2,240)} = 64.62$, $p < 0.001$, partial $\eta^2 = 0.35$, 95% CI [0.26, 0.43]) and age ($F_{(1,240)} = 16.13$, $p < 0.001$, partial $\eta^2 = 0.06$ [0.02, 0.13]) but not condition had an effect on fMRI parameter activity. Moreover, a significant interaction between scene-selective region and age was uncovered ($F_{(2,240)} = 4.18$, $p = 0.016$, partial $\eta^2 = 0.03$ [0.00, 0.09]). *Post hoc* tests revealed that the OPA was more activated than the PPA during the landmark condition in young (1.07 ± 0.21 vs. 0.25 ± 0.09, $p = 0.040$, Hedges' $g = -0.91$, 95% CI [−1.50, −0.33]) and older adults (2.19 ± 0.27 vs. 0.60 ± 0.14, $p < 0.001$, Hedges' $g = -1.52$ [−2.28, −0.76]) and during the control condition in young (1.55 ± 0.27 vs. 0.51 ± 0.13, $p = 0.001$, Hedges g' = −0.81 [−1.38, −0.23]) and older adults (2.19 ± 0.29 vs. 0.70 ± 0.12, $p < 0.001$, Hedges' $g = -1.15$ [−1.19, −0.43]). The OPA was also significantly more activated than the RSC during the landmark condition in young (1.07 ± 0.21 vs. 0.21 ± 0.09, $p = 0.022$, Hedges' $g = -0.83$ [−1.41, −0.25]) and older adults (2.19 ± 0.37 vs. 0.43 ± 0.15, $p < 0.001$, Hedges' $g = -1.72$ [−2.50, −0.93]) and during the control condition in young (1.55 ± 0.27 vs. 0.34 ± 0.17, $p < 0.001$, Hedges' $g = -0.99$ [−1.58, −0.41]) and older adults (2.19 ± 0.29 vs. 0.45 ± 0.18, $p < 0.001$, Hedges' $g = -1.33$ [−2.07, −0.59]). Finally, significantly enhanced OPA activity was found in older adults compared with young adults during the landmark condition (2.19 ± 0.27 vs. 1.07 ± 0.21, $p = 0.002$, Hedges' $g = -1.04$ [−1.70, −0.39]) but not the control condition.

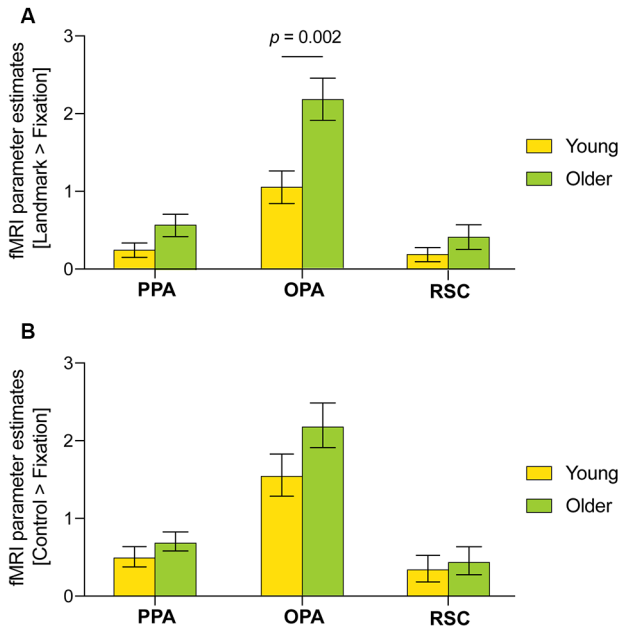


FIGURE 5 | Functional magnetic resonance imaging (fMRI) parameter estimates in the parahippocampal place area (PPA), occipital place area (OPA), and retrosplenial cortex (RSC) for the fMRI contrasts **(A)** [Landmark > Fixation] and **(B)** [Control > Fixation] across age groups. A three-way ANOVA was performed with scene-selective region, age and condition as factors. Only significant differences between age groups are displayed. Error bars reflect standard errors of the mean.

DISCUSSION

In the present exploratory fMRI study, age-related differences in landmark-based navigation were investigated using a simple Y-maze paradigm. The task was designed to limit the influence of mnemonic and motor components to a maximum to gain a preliminary understanding of the neural bases subtending visual-spatial cue reliance in young and healthy older adults.

Behavior

Well-established findings were corroborated by our results, showing that older adults had lower scores than their younger counterparts on visuospatial cognitive neuropsychological measures, including the perspective-taking, 3D mental rotation, and Corsi block-tapping tasks (Ohta et al., 1981; Clancy Dollinger, 1995; Iachini et al., 2005; Techentin et al., 2014).

Moreover, visuospatial memory and perspective-taking ability were associated with measures of navigational behavior. These tests are indeed known to be good predictors of general navigation skills and their decline with age may account in part for older adults' deficient navigation performance (Zhong and Moffat, 2016). Notably, perspective-taking, mental rotation, and visuospatial memory are important abilities for spatial learning and the dynamic manipulation of sensory information during navigation (Allen et al., 1996; Kozhevnikov et al., 2006; Meneghetti et al., 2018; Muffato et al., 2020).

Consistent with past literature, older subjects' navigation performances were significantly poorer than young subjects', with a bias for response-based strategies in older adults (Moffat and Resnick, 2002; Bohbot et al., 2012; Harris and Wolbers, 2012; Rodgers et al., 2012; Gazova et al., 2013; Wiener et al., 2013; Schuck et al., 2015; van der Ham et al., 2015; Zhong and Moffat, 2016; Kimura et al., 2019; Merhav and Wolbers, 2019; Bécu et al., 2020). Indeed, in our study, 52% of younger adults and 83% of older adults preferred a response-based strategy. Our findings are strikingly similar to those from Rodgers et al. (2012) who found that 46% of younger adults and 82% of older adults favored a response-based strategy in a sample of 86 participants. Our results are also in line with those reported by Bohbot et al. (2012) on a large sample of young ($n = 175$) and older ($n = 125$) participants. This navigation strategy preference was associated with older age. However, the possibility that the differential proportion of women in the two age groups (young: 28% vs. old: 59%) partly accounted for the increased use of response-based strategies in older adults cannot be excluded (Perrochon et al., 2018). Notwithstanding these age-related differences, it is important to mention that older participants achieved a high level of performance on the task and made few errors. It can be argued that this result stemmed from the simplicity of the virtual environment that contained a unique junction and three proximal landmarks (Moffat and Resnick, 2002; Caffò et al., 2018). Moreover, both place- and response-based strategies could be used to complete the task.

Whole-Brain Analyses

Following previous neuroimaging studies looking at the neural bases of spatial navigation, landmark-based navigation recruited an extended network of brain regions (Kuhn and Gallinat, 2014; Spiers and Barry, 2015; Coughlan et al., 2018; Cona and Scarpazza, 2019).

The multiple regression analyses showed that this network spanned posterior structures linked to visuospatial processing. Activation of the left superior occipital gyrus was reported which corresponds to visual area V3A and which is involved in optic flow tracking for visual path integration (Sherrill et al., 2015; Zajac et al., 2019). Also, our landmark-based navigation paradigm elicited activity in the ventral temporal cortex. The latter is known to process high-level visual information such as object quality (Kravitz et al., 2013; Nau et al., 2018). The recruited network also encompassed the posterior section of the hippocampus and the parahippocampal gyrus, brain areas that play a central role in spatial navigation and that are particularly active during immediate retrieval phases of navigation paradigms (Kuhn and Gallinat, 2014; Cona and Scarpazza, 2019). Furthermore, significant activity was found in the angular gyrus, a region of the posterior parietal cortex known to encode landmarks in the environment concerning the self (Ciaramelli et al., 2010; Auger and Maguire, 2018). Our task prompted activation of the prefrontal cortex which is thought to contribute to spatial working memory during active navigation (Wolbers and Hegarty, 2010; Ito, 2018). It thus appears that accurate landmark-based navigation required the integration of objects within a first-person framework and the maintenance of such representations in working memory (Sack, 2009; Seghier, 2012; Miniaci and De Leonibus, 2018). Finally, lobule VI and the vermis of the right cerebellum were found to be activated. This finding is in accordance with the cerebellum's postulated role in cognitive aspects of spatial navigation (Rochefort et al., 2013). We must nonetheless acknowledge the eventuality that cerebellar activity reflected sensory-motor processing such as the degree of motor learning or eye and finger movements (Bo et al., 2011; Iglói et al., 2015).

Of interest, whole-brain analyses for the contrast [Landmark > Control] revealed that young adults recruited the cortical projections of the central visual field in posterior occipital regions (MNI coordinates left: $x = -30$, $y = -97$, $z = -7$ and right: $x = 30$, $y = -88$, $z = -10$). The latter brain area is dedicated to fine-grained visual perception such as object recognition (Wandell et al., 2005; Kauffmann et al., 2014). Additionally, group comparisons revealed that young subjects had more activity in the anterior section of the inferior temporal gyrus than older subjects. As mentioned previously, the anterior temporal cortex is critical for perceptual recognition and visual object processing (Litman et al., 2009). Our findings resonate with recent evidence highlighting deficient fine-grained processing of sensory information in older adults and emphasize the importance of acute object discrimination for landmark-based navigation and episodic memory (Burke et al., 2018; Greene and Naveh-Benjamin, 2020). Taken together, the above results suggest that brain regions involved in the representation of fine-grained information may be disrupted in older age. Further research is warranted to determine whether the age-related decline in orientation skills could stem from the less efficient processing of visual-spatial cues.

Worthy of note, the differential patterns of neural activity observed in the young and older participant groups may be partially due to age-related cognitive and motor differences. Although the duration of the familiarization phase was tailored to each subject's needs and we controlled for response device use, the potential influence of older adults' lesser familiarity with new technologies and declining executive functions cannot be omitted. For example, the lack of activity elicited by the contrast [Landmark > Control] in older subjects could reflect the deficient integration of new instructions when switching between tasks (Hirsch et al., 2016). Additionally, the longer time necessary to reach the visible goal in the control condition along with the greater frontal activations elicited by the contrast [Control > Fixation] in older participants suggests a possible contribution of age-related executive impairment. These

results hint at the possibility that the control condition was cognitively more demanding for older participants than for young participants.

A higher level of recruitment of superior parietal regions was detected in older adults compared to young participants from the contrast [Landmark > Fixation]. The aforementioned association between angular gyrus activity and first-person navigation may provide a plausible explanation for the response strategy bias in the older adult group. This is consistent with the observed age-related reduction in temporal activity. Indeed, changes in strategy preference with advancing age have been extensively documented and they are thought to be mediated by a shift from the hippocampal regions towards other cerebral structures such as the parietal cortex (Rodgers et al., 2012; Wiener et al., 2013). Within this framework of interpretation, older adults' increased cerebellar activity could also reflect a change in strategy preference as recent evidence has implicated the cerebellum in the mediation of response-based strategies (Iglói et al., 2015). Older participants further displayed enhanced activation of frontal cortices. Various authors have stressed the impact of age-related modifications in the prefrontal cortex on hippocampal and striatal dynamics, which could contribute to impaired strategy implementation and switching (Lester et al., 2017; Goodroe et al., 2018; Zhong and Moffat, 2018). In contrast to previous studies that reported striatal activity during response-based navigation, our results did not show increased striatal activity in the older adult group (Konishi et al., 2013; Schuck et al., 2015). Such a difference may be explained by the high proportion of young adults using response-based strategies in our task.

Scene-Selective Regions Analyses

Given the predominant role of visual perception in human spatial navigation (Ekstrom, 2015; Nau et al., 2018), there has been a heightened interest in the PPA, RSC, and OPA and their respective contributions to landmark processing (Epstein et al., 2017; Julian et al., 2018). It is key to specify that there were no age-related differences in the activity of scene-selective regions when looking at the contrast [Landmark > Control]. However, a seemingly augmented activation of the OPA in older adults along with age-related differences during the control condition led us to conduct further exploratory analyses with the fMRI contrasts [Landmark > Fixation] and [Control > Fixation]. Interestingly, the OPA was more activated than the PPA and RSC in both landmark and control conditions across age groups. This finding fits well with the OPA's postulated implication in coding navigational affordances and visible paths in the environment (Bonner and Epstein, 2017; Patai and Spiers, 2017). Our results pointed to greater OPA activity in older participants compared with younger participants, which is in line with recent work showing higher functional connectivity around the OPA in older adults (Ramanoël et al., 2019). It is essential to note that the OPA activation in our study cannot be attributed to landmark-based navigation *per se* as it was uncovered by comparing neural activity during the landmark condition and fixation. The OPA is known to be sensitive to self-perceived distance and motion (Persichetti and Dilks, 2016) and the extraction of navigational affordances of the local visual scene from a first-person perspective (Bonner and Epstein, 2017). Critically, these self-centered navigation skills are relatively well preserved in healthy aging (Moffat, 2009). In line with the over-activation of the parietal cortex in older adults, one could conceive that the increased OPA activation in the older adult group reflects a compensatory mechanism to offset the reduced activity in the temporal cortex, thus mitigating age-related place learning deficits. As a side note, considering that the OPA has been causally linked to the processing of environmental boundaries (Julian et al., 2016), our result offers a potential explanation for older adults' preferential reliance on geometric information in an ecological cue conflict paradigm (Bécu et al., 2020). Such possibilities remain highly speculative and further studies are necessary to test these hypotheses specifically.

Surprisingly, no differences were found in the activity of the RSC across age groups. Previous work has demonstrated an age-related decline in RSC activation during spatial navigation tasks (Meulenbroek et al., 2004; Moffat et al., 2006; Antonova et al., 2009). The RSC is known to mediate several cognitive functions pertaining to spatial navigation (Vann et al., 2009; Mitchell et al., 2018) including translation between reference frames and recollection of visual landmarks (Auger et al., 2015). The discrepancy between our results and those from the literature could be explained by the relative simplicity of our task. In contrast to previous research conducted with young adults, our paradigm strived to restrict mnemonic processing and comprised only three stable, salient and simple landmarks located at a single intersection (Wolbers and Büchel, 2005; Auger et al., 2012; Auger and Maguire, 2018). Another probable explanation lies in the idea that functional and structural changes to the RSC could be more pronounced in pathological aging than in normal aging (Fjell et al., 2014; Dillen et al., 2016).

Finally, weak recruitment of the PPA was observed during landmark-based navigation, with no significant difference across age groups. Previous studies have found the PPA to be involved in the encoding of the navigational relevance of objects for orientation (Janzen and van Turennout, 2004) and in landmark recognition (Epstein and Vass, 2014). As previously noted, our virtual environment comprised a small number of simple and non-ambiguous objects, the lack of activity in the PPA is thus unsurprising. Also, two recent studies de-emphasized PPA's contribution to active navigation and highlighted its specificity for place recognition (Persichetti and Dilks, 2018, 2019).

Research exploring the neural activity within scene-selective regions in the context of aging is still in its infancy. Future studies are needed to better characterize age-related changes in brain areas implicated in processing both the visual and cognitive properties of spatial cues.

Limitations and Perspectives

The current study has limitations. First and foremost, there is a possibility that the age-related differences in behavior and neural activity during the control condition may have biased the secondary regression analyses. The latter are of

exploratory nature and are to be taken with great caution. Second, although the spatial memory component of the landmark condition was limited to a maximum, the idea that the observed neural activations reflect differences in spatial memory processing or task difficulty between the landmark and control conditions cannot be omitted. Furthermore, the age-related differences observed in the control condition emphasize the plausible impact of more general cognitive deficits such as executive dysfunction on landmark-based navigation. Further work specifically designed to disambiguate the role of age-related differences in visuospatial function and other cognitive dimensions should be conducted. These results also highlight the importance of the task chosen as the control condition in virtual spatial navigation paradigms to be relevant to the population(s) of interest. Critically, our findings put forward the question of whether control tasks are still appropriate in studies comparing complex behavior between young and older adults. Such a topic of research demands closer attention. Third, only the retrieval phase was used for the analyses, as the encoding phase proved to be too heterogeneous across participants. It would be of immediate interest to assess the influence of various visuospatial modulations, such as the visibility of spatial cues, on the quality of spatial encoding. FMRI spatial navigation paradigms only allow for visual input signals, however, "real-world" spatial navigation is reliant upon multiple sources of sensory information. Active walking as part of ecological study designs would provide proprioceptive and self-motion feedback signals as well as an improved field of view to participants. Such studies are necessary to complement the present findings. Previous research has indeed shown that navigation performance in older subjects is tightly coupled to the availability of multiple sources of sensory information (Adamo et al., 2012). Finally, future studies should take into consideration the role of sex and they should include an intermediate age group to gain a finer understanding of the neural dynamics subtending spatial navigation across the lifespan (Grön et al., 2000; van der Ham and Claessen, 2020).

CONCLUSION

To conclude, the present study sheds light on the possibility that navigational deficits in old age are linked to functional differences in brain areas involved in visual processing and to impaired representations of landmarks in temporal regions. This work helps towards a better comprehension of the neural dynamics subtending landmark-based navigation and it provides new insights on the impact of age-related spatial processing changes on navigation capabilities. We argue that approaching the study of spatial navigation in healthy and pathological aging from the perspective of visuospatial abilities is a critical next step in the field. Neuroimaging methods coupled with VR paradigms open up promising avenues to investigate age-related changes in navigation ability and to evaluate the benefits of training programs on older adults' autonomy and mobility.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the datasets generated for this study are available on reasonable request to the corresponding author. Requests to access the datasets should be directed to stephen.ramanoel@univ-cotedazur.fr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee "CPP Ile de France V" (ID_RCB 2015-A01094-45, CPP N°: 16122). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SR, MB, CH, and AA: study design. SR and MD: data acquisition and data processing. SR, MD, MB, and AA: manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Chair SILVERSIGHT Agence Nationale de la Recherche (ANR-18-CHIN-0002), the LabEx LIFESENSES (ANR-10-LABX-65) and by the IHU FOReSIGHT (ANR-18-IAHU-01).

ACKNOWLEDGMENTS

We would like to express our gratitude to the men and women who took part in this study. We thank the Quinze-Vingts Hospital for allowing us to acquire the MRI data. We thank Konogan Baranton and Isabelle Poulain (Essilor International) for the manufacturing of MRI-compatible glasses. A preliminary version of this manuscript has been released as a pre-print on bioRxiv (Ramanoël et al., 2020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.552111/full#supplementary-material.

REFERENCES

Adamo, D. E., Briceno, E. M., Sindone, J. A., Alexander, N. B., and Moffat, S. D. (2012). Age differences in virtual environment and real world path integration. *Front. Aging Neurosci.* 4:26. doi: 10.3389/fnagi.2012.00026

Allen, G. L., Kirasic, K. C., Dobson, S. H., Long, R. G., and Beck, S. (1996). Predicting environmental learning from spatial abilities: an indirect route. *Intelligence* 22, 327–355. doi: 10.1016/s0160-2896(96)90026-4

Allison, S. L., Fagan, A. M., Morris, J. C., and Head, D. (2016). Spatial navigation in preclinical Alzheimer's disease. *J. Alzheimers. Dis.* 52, 77–90. doi: 10.3233/JAD-150855

Antonova, E., Parslow, D., Brammer, M., Dawson, G. R., Jackson, S. H. D., and Morris, R. G. (2009). Age-related neural activity during allocentric spatial memory. *Memory* 17, 125–143. doi: 10.1080/09658210802077348

Auger, S. D., and Maguire, E. A. (2018). Retrosplenial cortex indexes stability beyond the spatial domain. *J. Neurosci.* 38, 1472–1481. doi: 10.1523/JNEUROSCI.2602-17.2017

Auger, S. D., Mullally, S. L., and Maguire, E. A. (2012). Retrosplenial cortex codes for permanent landmarks. *PLoS One* 7:e43620. doi: 10.1371/journal.pone.0043620

Auger, S. D., Zeidman, P., and Maguire, E. A. (2015). A central role for the retrosplenial cortex in de novo environmental learning. *Elife* 4:e09031. doi: 10.7554/eLife.09031

Bécu, M., Sheynikhovich, D., Tatur, G., Agathos, C. P., Bologna, L. L., Sahel, J.-A., et al. (2020). Age-related preference for geometric spatial cues during real-world navigation. *Nat. Hum. Behav.* 4, 88–99. doi: 10.1038/s41562-019-0718-z

Bo, J., Peltier, S. J., Noll, D. C., and Seidler, R. D. (2011). Age differences in symbolic representations of motor sequence learning. *Neurosci. Lett.* 504, 68–72. doi: 10.1016/j.neulet.2011.08.060

Bohbot, V. D., McKenzie, S., Konishi, K., Fouquet, C., Kurdi, V., Schachar, R., et al. (2012). Virtual navigation strategies from childhood to senescence: evidence for changes across the life span. *Front. Aging Neurosci.* 4:28. doi: 10.3389/fnagi.2012.00028

Bonner, M. F., and Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proc. Natl. Acad. Sci. U S A* 114, 4793–4798. doi: 10.1073/pnas.1618228114

Burke, S. N., Gaynor, L. S., Barnes, C. A., Bauer, R. M., Bizon, J. L., Roberson, E. D., et al. (2018). Shared functions of perirhinal and parahippocampal cortices: implications for cognitive aging. *Trends Neurosci.* 41, 349–359. doi: 10.1016/j.tins.2018.03.001

Caffò, A. O., Lopez, A., Spano, G., Serino, S., Cipresso, P., Stasolla, F., et al. (2018). Spatial reorientation decline in aging: the combination of geometry and landmarks. *Aging Ment. Health* 22, 1372–1383. doi: 10.1080/13607863.2017.1354973

Cheng, K., and Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychon. Bull. Rev.* 12, 1–23. doi: 10.3758/bf03196346

Chersi, F., and Burgess, N. (2015). The cognitive architecture of spatial navigation: hippocampal and striatal contributions. *Neuron* 88, 64–77. doi: 10.1016/j.neuron.2015.09.021

Chrastil, E. R. (2013). Neural evidence supports a novel framework for spatial navigation. *Psychon. Bull. Rev.* 20, 208–227. doi: 10.3758/s13423-012-0351-6

Ciaramelli, E., Rosenbaum, R. S., Solcz, S., Levine, B., and Moscovitch, M. (2010). Mental space travel: damage to posterior parietal cortex prevents egocentric navigation and reexperiencing of remote spatial memories. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 619–634. doi: 10.1037/a0019181

Clancy Dollinger, S. M. (1995). Mental rotation performance: age, sex, and visual field differences. *Dev. Neuropsychol.* 11, 215–222. doi: 10.1080/87565649509540614

Colombo, D., Serino, S., Tuena, C., Pedroli, E., Dakanalis, A., Cipresso, P., et al. (2017). Egocentric and allocentric spatial reference frames in aging: a systematic review. *Neurosci. Biobehav. Rev.* 80, 605–621. doi: 10.1016/j.neubiorev.2017.07.012

Cona, G., and Scarpazza, C. (2019). Where is the "where" in the brain? A meta-analysis of neuroimaging studies on spatial cognition. *Hum. Brain Mapp.* 40, 1867–1886. doi: 10.1002/hbm.24496

Corsi, P. M. (1973). Human memory and the medial temporal region of the brain. *Dissert. Abstrac. Int.* 34:891.

Coughlan, G., Laczó, J., Hort, J., Minihane, A.-M., and Hornberger, M. (2018). Spatial navigation deficits—overlooked cognitive marker for preclinical Alzheimer disease? *Nat. Rev. Neurol.* 14, 496–506. doi: 10.1038/s41582-018-0031-x

Daugherty, A. M., Bender, A. R., Yuan, P., and Raz, N. (2016). Changes in search path complexity and length during learning of a virtual water maze: age differences and differential associations with hippocampal subfield volumes. *Cereb. Cortex* 26, 2391–2401. doi: 10.1093/cercor/bhv061

Daugherty, A. M., and Raz, N. (2017). A virtual water maze revisited: two-year changes in navigation performance and their neural correlates in healthy adults. *NeuroImage* 146, 492–506. doi: 10.1016/j.neuroimage.2016.09.044

Dillen, K. N. H., Jacobs, H. I. L., Kukolja, J., von Reutern, B., Richter, N., Onur, Ö. A., et al. (2016). Aberrant functional connectivity differentiates retrosplenial cortex from posterior cingulate cortex in prodromal Alzheimer's disease. *Neurobiol. Aging* 44, 114–126. doi: 10.1016/j.neurobiolaging.2016.04.010

Ekstrom, A. D. (2015). Why vision is important to how we navigate. *Hippocampus* 25, 731–735. doi: 10.1002/hipo.22449

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 12, 388–396. doi: 10.1016/j.tics.2008.07.004

Epstein, R. A., and Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20120533. doi: 10.1098/rstb.2012.0533

Epstein, R. A., Patai, E. Z., Julian, J. B., and Spiers, H. J. (2017). The cognitive map in humans: spatial navigation and beyond. *Nat. Neurosci.* 20, 1504–1513. doi: 10.1038/nn.4656

Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., and Walhovd, K. B. (2014). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Prog. Neurobiol.* 117, 20–40. doi: 10.1016/j.pneurobio.2014.02.004

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Foo, P., Warren, W. H., Duchon, A., and Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 195–215. doi: 10.1037/0278-7393.31.2.195

Foster, T. C., Defazio, R. A., and Bizon, J. L. (2012). Characterizing cognitive aging of spatial and contextual memory in animal models. *Front. Aging Neurosci.* 4:12. doi: 10.3389/fnagi.2012.00012

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Gazova, I., Laczó, J., Rubinova, E., Mokrisova, I., Hyncicova, E., Andel, R., et al. (2013). Spatial navigation in young versus older adults. *Front. Aging Neurosci.* 5:94. doi: 10.3389/fnagi.2013.00094

Gazova, I., Vlcek, K., Laczó, J., Nedelska, Z., Hyncicova, E., Mokrisova, I., et al. (2012). Spatial navigation—a unique window into physiological and pathological aging. *Front. Aging Neurosci.* 4:16. doi: 10.3389/fnagi.2012.00016

Giocomo, L. M. (2016). Environmental boundaries as a mechanism for correcting and anchoring spatial maps. *J. Physiol.* 594, 6501–6511. doi: 10.1113/JP270624

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127

Goodroe, S. C., Starnes, J., and Brown, T. I. (2018). The complex nature of hippocampal-striatal interactions in spatial navigation. *Front. Hum. Neurosci.* 12:250. doi: 10.3389/fnhum.2018.00250

Greene, N. R., and Naveh-Benjamin, M. (2020). A specificity principle of memory: evidence from aging and associative memory. *Psychol. Sci.* 31, 316–331. doi: 10.1177/0956797620901760

Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., and Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nat. Neurosci.* 3, 404–408. doi: 10.1038/73980

Harris, M. A., Wiener, J. M., and Wolbers, T. (2012). Aging specifically impairs switching to an allocentric navigational strategy. *Front. Aging Neurosci.* 4:29. doi: 10.3389/fnagi.2012.00029

Harris, M. A., and Wolbers, T. (2012). Ageing effects on path integration and landmark navigation. *Hippocampus* 22, 1770–1780. doi: 10.1002/hipo.22011

Hartmeyer, S., Grzeschik, R., Wolbers, T., and Wiener, J. M. (2017). The effects of attentional engagement on route learning performance in a virtual environment: an aging study. *Front. Aging Neurosci.* 9:235. doi: 10.3389/fnagi.2017.00235

Herweg, N. A., and Kahana, M. J. (2018). Spatial representations in the human brain. *Front. Hum. Neurosci.* 12:297. doi: 10.3389/fnhum.2018.00297

Hirsch, P., Schwarzkopp, T., Declerck, M., Reese, S., and Koch, I. (2016). Age-related differences in task switching and task preparation: exploring the role of task-set competition. *Acta Psychol.* 170, 66–73. doi: 10.1016/j.actpsy.2016.06.008

Iachini, T., Poderico, C., Ruggiero, G., and Iavarone, A. (2005). Age differences in mental scanning of locomotor maps. *Disabil. Rehabil.* 27, 741–752. doi: 10.1080/09638280400014782

Iaria, G., Palermo, L., Committeri, G., and Barton, J. J. S. (2009). Age differences in the formation and use of cognitive maps. *Behav. Brain Res.* 196, 187–191. doi: 10.1016/j.bbr.2008.08.040

Iaria, G., Petrides, M., Dagher, A., Pike, B., and Bohbot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J. Neurosci.* 23, 5945–5952. doi: 10.1523/JNEUROSCI.23-13-05945.2003

Iglói, K., Doeller, C. F., Berthoz, A., Rondi-Reig, L., and Burgess, N. (2010). Lateralized human hippocampal activity predicts navigation based on sequence or place memory. *Proc. Natl. Acad. Sci. U S A* 107, 14466–14471. doi: 10.1073/pnas.1004243107

Iglói, K., Doeller, C. F., Paradis, A.-L., Benchenane, K., Berthoz, A., Burgess, N., et al. (2015). Interaction between hippocampus and cerebellum crus i in sequence-based but not place-based navigation. *Cereb. Cortex* 25, 4146–4154. doi: 10.1093/cercor/bhu132

Ito, H. T. (2018). Prefrontal-hippocampal interactions for spatial navigation. *Neurosci. Res.* 129, 2–7. doi: 10.1016/j.neures.2017.04.016

Janzen, G., and van Turennout, M. (2004). Selective neural representation of objects relevant for navigation. *Nat. Neurosci.* 7, 673–677. doi: 10.1038/nn1257

Javadi, A.-H., Emo, B., Howard, L. R., Zisch, F. E., Yu, Y., Knight, R., et al. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. *Nat. Commun.* 8:14652. doi: 10.1038/ncomms14652

Julian, J. B., Keinath, A. T., Marchette, S. A., and Epstein, R. A. (2018). The neurocognitive basis of spatial reorientation. *Curr. Biol.* 28, R1059–R1073. doi: 10.1016/j.cub.2018.04.057

Julian, J. B., Ryan, J., Hamilton, R. H., and Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Curr. Biol.* 26, 1104–1109. doi: 10.1016/j.cub.2016.02.066

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., and Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage* 132, 417–424. doi: 10.1016/j.neuroimage.2016.02.062

Kauffmann, L., Ramanoël, S., and Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Front. Integr. Neurosci.* 8:37. doi: 10.3389/fnint.2014.00037

Kimura, K., Reichert, J. F., Kelly, D. M., and Moussavi, Z. (2019). Older adults show less flexible spatial cue use when navigating in a virtual reality environment compared with younger adults. *Neurosci. Insights* 14:2633105519896803. doi: 10.1177/2633105519896803

Kirasic, K. C. (1991). Spatial cognition and behavior in young and elderly adults: implications for learning new environments. *Psychol. Aging* 6, 10–18. doi: 10.1037//0882-7974.6.1.10

Konishi, K., Etchamendy, N., Roy, S., Marighetto, A., Rajah, N., and Bohbot, V. D. (2013). Decreased functional magnetic resonance imaging activity in the hippocampus in favor of the caudate nucleus in older adults tested in a virtual navigation task. *Hippocampus* 23, 1005–1014. doi: 10.1002/hipo.22181

Korthauer, L. E., Nowak, N. T., Moffat, S. D., An, Y., Rowland, L. M., Barker, P. B., et al. (2016). Correlates of virtual navigation performance in older adults. *Neurobiol. Aging* 39, 118–127. doi: 10.1016/j.neurobiolaging.2015.12.003

Kozhevnikov, M., and Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Mem. Cognit.* 29, 745–756. doi: 10.3758/bf03200477

Kozhevnikov, M., Motes, M. A., Rasch, B., and Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Appl. Cogn. Psychol.* 20, 397–417. doi: 10.1002/acp.1192

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49. doi: 10.1016/j.tics.2012.10.011

Kuhn, S., and Gallinat, J. (2014). Segregating cognitive functions within hippocampal formation: a quantitative meta-analysis on spatial navigation and episodic memory. *Hum. Brain Mapp.* 35, 1129–1142. doi: 10.1002/hbm.22239

Laczó, J., Andel, R., Nedelska, Z., Vyhnalek, M., Vlcek, K., Crutch, S., et al. (2017). Exploring the contribution of spatial navigation to cognitive functioning in older adults. *Neurobiol. Aging* 51, 67–70. doi: 10.1016/j.neurobiolaging.2016.12.003

Laczó, J., Parizkova, M., and Moffat, S. D. (2018). Spatial navigation, aging and Alzheimer's disease. *Aging* 10, 3050–3051. doi: 10.18632/aging.101634

Lagrené, K., Bécu, M., Seiple, W. H., Raphanel Bataille, M., Combariza, S., Paques, M., et al. (2019). Healthy and pathological visual aging in a French follow-up cohort study. *Invest. Ophthalmol. Vis. Sci.* 60:5915. Available online at: https://iovs.arvojournals.org/article.aspx?articleid=2747032.

Lester, A. W., Moffat, S. D., Wiener, J. M., Barnes, C. A., and Wolbers, T. (2017). The aging navigational system. *Neuron* 95, 1019–1035. doi: 10.1016/j.neuron.2017.06.037

Li, A. W. Y., and King, J. (2019). Spatial memory and navigation in ageing: a systematic review of MRI and fMRI studies in healthy participants. *Neurosci. Biobehav. Rev.* 103, 33–49. doi: 10.1016/j.neubiorev.2019.05.005

Lithfous, S., Dufour, A., and Després, O. (2013). Spatial navigation in normal aging and the prodromal stage of Alzheimer's disease: insights from imaging and behavioral studies. *Ageing Res. Rev.* 12, 201–213. doi: 10.1016/j.arr.2012.04.007

Litman, L., Awipi, T., and Davachi, L. (2009). Category-specificity in the human medial temporal lobe cortex. *Hippocampus* 19, 308–319. doi: 10.1002/hipo.20515

Lövdén, M., Schaefer, S., Noack, H., Bodammer, N. C., Kühn, S., Heinze, H.-J., et al. (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiol. Aging* 33, 620.e9–620.e22.doi: 10.1016/j.neurobiolaging.2011.02.013

Marchette, S. A., Vass, L. K., Ryan, J., and Epstein, R. A. (2015). Outside looking in: landmark generalization in the human navigational system. *J. Neurosci.* 35, 14896–14908. doi: 10.1523/JNEUROSCI.2270-15.2015

Mazaika, P. K., Hoeft, F., Glover, G. H., and Reiss, A. L. (2009). Methods and software for fMRI analysis of clinical subjects. *NeuroImage* 47:S58. doi: 10.1016/S1053-8119(09)70238-1

Meneghetti, C., Muffato, V., Borella, E., and De Beni, R. (2018). Map learning in normal aging: the role of individual visuo-spatial abilities and implications. *Curr. Alzheimer Res.* 15, 205–218. doi: 10.2174/1567205014666171030113515

Merhav, M., Riemer, M., and Wolbers, T. (2019). Spatial updating deficits in human aging are associated with traces of former memory representations. *Neurobiol. Aging* 76, 53–61. doi: 10.1016/j.neurobiolaging.2018.12.010

Merhav, M., and Wolbers, T. (2019). Aging and spatial cues influence the updating of navigational memories. *Sci. Rep.* 9:11469. doi: 10.1038/s41598-019-47971-2

Meulenbroek, O., Petersson, K. M., Voermans, N., Weber, B., and Fernandez, G. (2004). Age differences in neural correlates of route encoding and route recognition. *NeuroImage* 22, 1503–1514. doi: 10.1016/j.neuroimage.2004.04.007

Miniaci, M. C., and De Leonibus, E. (2018). Missing the egocentric spatial reference: a blank on the map. *F1000Res.* 7:168. doi: 10.12688/f1000research.13675.1

Mitchell, A. S., Czajkowski, R., Zhang, N., Jeffery, K., and Nelson, A. J. D. (2018). Retrosplenial cortex and its role in spatial cognition. *Brain Neurosci. Adv.* 2:2398212818757098. doi: 10.1177/2398212818757098

Moffat, S. D. (2009). Aging and spatial navigation: what do we know and where do we go? *Neuropsychol. Rev.* 19, 478–489. doi: 10.1007/s11065-009-9120-3

Moffat, S. D., Elkins, W., and Resnick, S. M. (2006). Age differences in the neural systems supporting human allocentric spatial navigation. *Neurobiol. Aging* 27, 965–972. doi: 10.1016/j.neurobiolaging.2005.05.011

Muffato, V., Meneghetti, C., and De Beni, R. (2020). The role of visuo-spatial abilities in environment learning from maps and navigation over the adult lifespan. *Br. J. Psychol.* 111, 70–91. doi: 10.1111/bjop.12384

Moffat, S. D., and Resnick, S. M. (2002). Effects of age on virtual environment place navigation and allocentric cognitive mapping. *Behav. Neurosci.* 116, 851–859. doi: 10.1037//0735-7044.116.5.851

Nau, M., Julian, J. B., and Doeller, C. F. (2018). How the brain's navigation system shapes our visual experience. *Trends Cogn. Sci.* 22, 810–825. doi: 10.1016/j.tics.2018.06.008

Ohta, R. J., Walsh, D. A., and Krauss, I. K. (1981). Spatial perspective-taking ability in young and elderly adults. *Exp. Aging Res.* 7, 45–63. doi: 10.1080/03610738108259785

Packard, M. G., and Goodman, J. (2013). Factors that influence the relative use of multiple memory systems. *Hippocampus* 23, 1044–1052. doi: 10.1002/hipo.22178

Patai, E. Z., and Spiers, H. J. (2017). Human navigation: occipital place area detects potential paths in a scene. *Curr. Biol.* 27, R599–R600. doi: 10.1016/j.cub.2017.05.012

Perrochon, A., Mandigout, S., Petruzzellis, S., Soria Garcia, N., Zaoui, M., Berthoz, A., et al. (2018). The influence of age in women *in visuo*-spatial memory in reaching and navigation tasks with and without landmarks. *Neurosci. Lett.* 684, 13–17. doi: 10.1016/j.neulet.2018.06.054

Persichetti, A. S., and Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex* 77, 155–163. doi: 10.1016/j.cortex.2016.02.006

Persichetti, A. S., and Dilks, D. D. (2018). Dissociable neural systems for recognizing places and navigating through them. *J. Neurosci.* 38, 10295–10304. doi: 10.1523/JNEUROSCI.1200-18.2018

Persichetti, A. S., and Dilks, D. D. (2019). Distinct representations of spatial and categorical relationships across human scene-selective cortex. *Proc. Natl. Acad. Sci. U S A* 116, 21312–21317. doi: 10.1073/pnas.1903057116

Picucci, L., Caffò, A. O., and Bosco, A. (2009). Age and sex differences in a virtual version of the reorientation task. *Cogn. Process.* 10, S272–S275. doi: 10.1007/s10339-009-0321-8

Ramanoël, S., Durteste, M., Bécu, M., Habas, C., and Arleo, A. (2020). Differential brain activity in visuo-perceptual regions during landmark-based navigation in young and healthy older adults. *bioRiv* [Preprint]. doi: 10.1101/2020.03.13.990572

Ramanoël, S., Kauffmann, L., Cousin, E., Dojat, M., and Peyrin, C. (2015). Age-related differences in spatial frequency processing during scene categorization. *PLoS One* 10:e0134554. doi: 10.1371/journal.pone.0134554

Ramanoël, S., York, E., Le Petit, M., Lagrene, K., Habas, C., and Arleo, A. (2019). Age-related differences in functional and structural connectivity in the spatial navigation brain network. *Front. Neural Circuits* 13:69. doi: 10.3389/fncir.2019.00069

Ratliff, K. R., and Newcombe, N. S. (2008). Reorienting when cues conflict: evidence for an adaptive-combination view. *Psychol. Sci.* 19, 1301–1307. doi: 10.1111/j.1467-9280.2008.02239.x

Rochefort, C., Lefort, J. M., and Rondi-Reig, L. (2013). The cerebellum: a new key structure in the navigation system. *Front. Neural Circuits* 7:35. doi: 10.3389/fncir.2013.00035

Rodgers, M. K., Sindone, J. A. III., and Moffat, S. D. (2012). Effects of age on navigation strategy. *Neurobiol. Aging* 33, 202.e15–202.e22.doi: 10.1016/j.neurobiolaging.2010.07.021

Sack, A. T. (2009). Parietal cortex and spatial cognition. *Behav. Brain Res.* 202, 153–161. doi: 10.1016/j.bbr.2009.03.012

Schuck, N. W., Doeller, C. F., Polk, T. A., Lindenberger, U., and Li, S.-C. (2015). Human aging alters the neural computation and representation of space. *NeuroImage* 117, 141–150. doi: 10.1016/j.neuroimage.2015.05.031

Seghier, M. L. (2012). The angular gyrus. *Neurosci.* 19, 43–61. doi: 10.1177/1073858412440596

Sherrill, K. R., Chrastil, E. R., Ross, R. S., Erdem, U. M., Hasselmo, M. E., and Stern, C. E. (2015). Functional connections between optic flow areas and navigationally responsive brain regions during goal-directed navigation. *NeuroImage* 118, 386–396. doi: 10.1016/j.neuroimage.2015.06.009

Spiers, H. J., and Barry, C. (2015). Neural systems supporting navigation. *Curr. Opin. Behav. Sci.* 1, 47–55. doi: 10.1016/j.cobeha.2014.08.005

Stankiewicz, B. J., and Kalia, A. A. (2007). Acquisition of structural versus object landmark knowledge. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 378–390. doi: 10.1037/0096-1523.33.2.378

Sturz, B. R., Kilday, Z. A., and Bodily, K. D. (2013). Does constraining field of view prevent extraction of geometric cues for humans during virtual-environment reorientation? *J. Exp. Psychol. Anim. Behav. Process.* 39, 390–396. doi: 10.1037/a0032543

Sutton, J. E., Joanisse, M. F., and Newcombe, N. S. (2010). Spinning in the scanner: neural correlates of virtual reorientation. *J. Exp. Psychol. Learn. Mem. Cogn.* 36, 1097–1107. doi: 10.1037/a0019938

Techentin, C., Voyer, D., and Voyer, S. D. (2014). Spatial abilities and aging: a meta-analysis. *Exp. Aging Res.* 40, 395–425. doi: 10.1080/0361073X.2014.926773

Tommasi, L., Chiandetti, C., Pecchia, T., Sovrano, V. A., and Vallortigara, G. (2012). From natural geometry to spatial cognition. *Neurosci. Biobehav. Rev.* 36, 799–824. doi: 10.1016/j.neubiorev.2011.12.007

United Nations, Departement of Economic and Social Affairs, Population Division. (2019). *World Population Ageing 2019 Highlights (ST/ESA/SER.A/430).* New York, NY: United Nations Sales. Available online at: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf.

van der Ham, I. J. M., Baalbergen, H., van der Heijden, P. G. M., Postma, A., Braspenning, M., and van der Kuil, M. N. A. (2015). Distance comparisons in virtual reality: effects of path, context, and age. *Front. Psychol.* 6:1103. doi: 10.3389/fpsyg.2015.01103

van der Ham, I. J. M., and Claessen, M. H. G. (2020). How age relates to spatial navigation performance: functional and methodological considerations. *Ageing Res. Rev.* 58:101020. doi: 10.1016/j.arr.2020.101020

Vandenberg, S. G., and Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Mot. Skills* 47, 599–604. doi: 10.2466/pms.1978.47.2.599

Vann, S. D., Aggleton, J. P., and Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat. Rev. Neurosci.* 10, 792–802. doi: 10.1038/nrn2733

Wandell, B. A., Brewer, A. A., and Dougherty, R. F. (2005). Visual field map clusters in human cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 693–707. doi: 10.1098/rstb.2005.1628

Weiskopf, N., Hutton, C., Josephs, O., and Deichmann, R. (2006). Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. *NeuroImage* 33, 493–504. doi: 10.1016/j.neuroimage.2006.07.029

Wiener, J. M., de Condappa, O., Harris, M. A., and Wolbers, T. (2013). Maladaptive bias for extrahippocampal navigation strategies in aging humans. *J. Neurosci.* 33, 6012–6017. doi: 10.1523/JNEUROSCI.0717-12.2013

Wiener, J. M., Kmecova, H., and de Condappa, O. (2012). Route repetition and route retracing: effects of cognitive aging. *Front. Aging Neurosci.* 4:7. doi: 10.3389/fnagi.2012.00007

Wilkniss, S. M., Jones, M. G., Korol, D. L., Gold, P. E., and Manning, C. A. (1997). Age-related differences in an ecologically based study of route learning. *Psychol. Aging* 12, 372–375. doi: 10.1037//0882-7974.12.2.372

Wolbers, T., and Büchel, C. (2005). Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *J. Neurosci.* 25, 3333–3340. doi: 10.1523/JNEUROSCI.4705-04.2005

Wolbers, T., and Hegarty, M. (2010). What determines our navigational abilities? *Trends Cogn. Sci.* 14, 138–146. doi: 10.1016/j.tics.2010.01.001

Zajac, L., Burte, H., Taylor, H. A., and Killiany, R. (2019). Self-reported navigation ability is associated with optic flow-sensitive regions' functional connectivity patterns during visual path integration. *Brain Behav.* 9:e01236. doi: 10.1002/brb3.1236

Zhong, J. Y., and Moffat, S. D. (2016). Age-related differences in associative learning of landmarks and heading directions in a virtual navigation task. *Front. Aging Neurosci.* 8:122. doi: 10.3389/fnagi.2016.00122

Zhong, J. Y., and Moffat, S. D. (2018). Extrahippocampal contributions to age-related changes in spatial navigation ability. *Front. Hum. Neurosci.* 12:272. doi: 10.3389/fnhum.2018.00272

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 16 November 2020
doi: 10.3389/fnhum.2020.586963

Multidomain Cognitive Training Transfers to Attentional and Executive Functions in Healthy Older Adults

Patrick D. Gajewski[1], Sven Thönes[2], Michael Falkenstein[3], Edmund Wascher[1] and Stephan Getzmann[1]*

[1] Leibniz Research Centre for Working Environment and Human Factors (IfADo), Dortmund, Germany, [2] Department of Psychology, University of Mainz, Dortmund, Germany, [3] Institute for Working, Learning and Aging (ALA), Bochum, Germany

Healthy aging is associated with deficits in focused and sustained attention and executive functions. However, cognitive training (CT) provides a promising method to counteract these deficits. In the present randomized controlled study, we examined to what extent CT regimes can improve attention, verbal skills, and inhibition capacities. Over a period of 16 weeks, healthy older adults (65 years and older, mean: 70 years) received a trainer-guided multidomain paper-and-pencil and computerized CT. Pre- and post-training, a battery of psychometric tests was applied that measured the critical functions. This study used two control groups: a passive control and an active control group performing a relaxation training. Compared to a passive control group, the CT led to enhanced performance in the attentional endurance test and the interference list of the Stroop test, whereas no benefits in verbal and crystalized tests were found. Similar effects were found on the attentional endurance compared to the active control group. Additionally, word fluency was enhanced after CT, but the improvement in the Stroop test did not reach significance compared to the active control. The contents of CT were dissimilar to the psychometric tests showing far transfer, whereas no transfer to attentional or memory functions in the daily life assessed by the Cognitive Failures Questionnaire was found. This demonstrates specific gains of multidomain CT on cognitive functions not explicitly trained and lack of transfer to daily activities.

Keywords: cognitive training, transfer, stroop interference, executive functions, selective attention, verbal fluency

OPEN ACCESS

Edited by:
Susan Mérillat,
University of Zurich, Switzerland

Reviewed by:
Tilo Strobach,
Medical School Hamburg, Germany
Zsófia Anna Gaál,
Hungarian Academy of Sciences (MTA), Hungary

****Correspondence:***
Patrick D. Gajewski
gajewski@ifado.de

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

Received: *24 July 2020*
Accepted: *19 October 2020*
Published: *16 November 2020*

Citation:
Gajewski PD, Thönes S, Falkenstein M, Wascher E and Getzmann S (2020) Multidomain Cognitive Training Transfers to Attentional and Executive Functions in Healthy Older Adults. Front. Hum. Neurosci. 14:586963. doi: 10.3389/fnhum.2020.586963

INTRODUCTION

People in western societies are living longer and longer, and the incidence of mild cognitive impairment (MCI) and dementia has increased rapidly from decade to decade. The expected number of people who will be suffering from dementia is estimated at about 150 million in 2050 (WHO, 2015). But even healthy aging is usually associated with a decline of a number of distinct cognitive abilities which are essential for mobility and independent living. Mnemonic,

attentional, and executive functions orchestrating goal-directed behavior are especially affected, which compromises the quality of life in older age (Craik and Salthouse, 2000). Thus, in recent years, there is an increasing interest in the mechanisms and determinants of cognitive aging. On the one hand, there are genetic dispositions and other biological factors that set an individual range of cognitive abilities. On the other hand, there are several environmental factors and behavioral adjustments, like education, nutrition, physical activity, and cognitive engagement, that stimulate the cognitive system and may compensate some deficits to a certain extent due to neuronal plasticity even in older age (Stern, 2009; Greenwood and Parasuraman, 2010; Bamidis et al., 2014; Ballesteros et al., 2014, 2015; Gajewski and Falkenstein, 2016, for reviews). Thus, in addition to the influencing factors, methods and interventions ameliorating cognitive aging has also gained more and more interest.

Cognitive Training and Transfer

One method for improving cognitive fitness in older age is cognitive training (CT). CT for older adults is increasingly used to improve their performance in cognitive functions which are essential for activities in daily life and to maintain self-sufficiency but prone to age-related decline. In most CT studies, test-like tasks or serous games were trained by applying paper-and-pencil tasks or PC-based tasks for a period of several weeks or months. In particular, computer-based games were frequently used to assess the efficacy of CT. Indeed, a number of controlled studies showed effects of computer-based CT on attention (Green and Bavelier, 2003, 2006; Schubert et al., 2015; Strobach and Karbach, 2016; Bavelier and Green, 2019), executive functions (Strobach et al., 2012; Anguera et al., 2013), and working memory (Salminen et al., 2016; Strobach and Huestegge, 2017).

An important aspect of CT research is the transfer effect referring to the ability to use the knowledge and skills learned in one scenario to achieve different goals in other scenarios (Wenig et al., 2019). Transfer can be differentiated into near-transfer effects (post-training improvement in tasks similar to the training tasks) and far-transfer effects (post-training improvement in tasks that are different from the training tasks in nature or in appearance; Barnett and Ceci, 2002). Recent reviews and meta-analyses suggest that CT leads to improvements of the trained cognitive functions (near transfer) and also transfers to untrained cognitive tasks, intelligence, or even performance in everyday situations (far transfer) in healthy older adults (Baltes et al., 1989; Willis et al., 2006; Ball et al., 2007; Cassavaugh and Kramer, 2009; Karbach and Verhaeghen, 2014; Kelly et al., 2014; Strobach and Karbach, 2016; Chiu et al., 2017; Strobach and Huestegge, 2017).

Limits of Cognitive Training

The outcome of CT is usually indicated by performance changes in psychometric or neuropsychological tests. As outlined above, studies using behavioral measures showed training-related improvements of mnemonic and executive functions (Kueider et al., 2012; Lampit et al., 2014, for meta-analyses). These cognitive functions have been assumed to constitute fluid intelligence that enables planning and implementing goal-directed behavior, and the flexible coping with new and unexpected situations. In contrast, crystallized intelligence representing general knowledge and life experience as originally proposed by Horn and Cattell (1967) (see also Baltes, 1987) usually remains unchanged after CT (Baltes et al., 1989). Analyses of brain processes accompanied by CT-related gains of cognitive performance have repeatedly shown enhanced activity in EEG- and fMRI-based measures (Bamidis et al., 2014; Brehmer et al., 2014; Falkenstein and Gajewski, 2016). Nevertheless, the findings regarding training efficacy and transfer are inconsistent. Some meta-analyses showed benefits of CT on a number of cognitive functions, such as memory, attention, and executive functions, whereas other meta-analyses did not find substantial effects or found only marginal effects of CT (Lampit et al., 2014; Simons et al., 2016). It seems that the efficacy of CT depends on a large number of influencing factors, such as training content, duration and frequency of CT, adaptivity of the training to cognitive changes, amount of overlap of neural circuits affected by training contents, and performance measure but also type of measures and study design (Bavelier and Green, 2019).

Moreover, a potential shortcoming of CT regimes is a rather inconsistent transfer to other, not explicitly trained cognitive functions. On the one hand, there are several positive findings regarding near or even far transfer (Willis et al., 2006; Minear and Shah, 2008; Karbach and Kray, 2009; Karbach and Verhaeghen, 2014; Strobach and Huestegge, 2017), while some studies and meta-analyses questioned the existence of far transfer on non-trained functions (Shipstead et al., 2012; Melby-Lervåg and Hulme, 2013; Melby-Lervåg et al., 2016; Sala and Gobet, 2017). As a consequence, more randomized controlled studies are necessary to elucidate the efficacy of training and transfer effects of CT on cognitive functions in old age.

The Present Study

The aim of the present study was to evaluate effects of multidomain adaptive paper-and-pencil and computerized CT in older individuals in a randomized controlled trial. Multidomain CTs are not limited to a single cognitive process (such as memory or processing speed) but involve a number of cognitive processes that interact and increase the demands on the cognitive system (Cheng et al., 2012). Indeed, the new aspect of the present study was the complexity and diversity of the CT that aimed at maximal enhancement of general cognitive functioning in the participants by offering them an interesting and varying program with a large fun factor to enhance their motivation to train. Moreover, the training was adaptive. Adaptivity refers to an adjustment of the training difficulty to an individual performance level and a continuous feedback and has been shown to enhance the motivation of participants and to produce larger transfer effects than non-adaptive trainings (Jaeggi et al., 2008; Karbach and Kray, 2009; Anguera et al., 2013; Mishra et al., 2014). Thus, enhanced complexity, adaptivity, and increasing difficulty of the CT should improve subject's cognitive capacity and thus

increase the probability to achieve transfer effects to non-trained functions.

The data presented here are from a large training study of which some PC-based tasks using electrophysiological measures have already been reported (Gajewski and Falkenstein, 2012, 2018; Wild-Wall et al., 2012; Küper et al., 2017). For the present study, the results of standardized psychometric paper-and-pencil tests were analyzed. In particular, we focused on cognitive domains like processing speed, selective and sustained attention, word fluency, and executive functions like interference processing contributing to fluid intelligence. These tests have been widely used to assess effects of CT and to ensure comparability with other training studies.

We compared the performance of older participants before and after a complex adaptive multidomain CT with the performance of an active control (relaxation training) and no-contact control groups. The no-contact control group was implemented to evaluate test–retest effects (Bavelier and Green, 2019). The rationale to use an active control group was to extract effects that are due to the social component and regular activity during the same time with the same frequency as the CT group. Participants were trained for a period of 4 months, twice per week and 90 min per session. We expected that 4 months of multidomain CT is sufficient to improve cognitive functions constituting fluid intelligence in an older population, whereas no changes are expected with regard to crystallized intelligence. In addition, we took a closer look on the domain-specific improvements after CT and evaluated transfer effects of non-explicit trained tasks on attention-related and executive functions and a potential far transfer to everyday attentional and memory functions assessed by the Cognitive Failures Questionnaire (CFQ).

MATERIALS AND METHODS

Participants

Data of 69 participants aged 65–88 years (M = 70.3 years, SD = 4.3, 63.8% female) were analyzed, who conducted either CT, or a relaxation training (active control group), or belonged to a no-contact control group. A fourth group, which was part of the original study that received physical training, was not included here as physical intervention was outside of the focus of the present study. For details on the acquisition procedure and the characteristics of the whole sample, please refer to Gajewski and Falkenstein (2012). Shortly, the participants were included in the study after having met a number of criteria inquired by a telephone interview. They should be physically and mentally fit without any history of neurological, psychiatric, motor, cardiovascular, or oncologic diseases or any psychopharmacological medication. No participants who already engaged in CT for more than 1.5 h/week were included in the study. The participants were randomly assigned to the CT group (CT; n = 32, mean age: M = 71, SD = 4.2, 62.5% female), the active control group (ACG; n = 33, mean age: M = 71, SD = 4.5, 62.9% female), and the passive control group (CG; n = 37, mean age: M = 70, SD = 4.2, 61.5% female). The final number of participants varied between 31 and 32 in the CT group, 33 and 34 in the active control group, and 36 and 37 in the passive control group due to missing data in some of the tests (e.g., due to color blindness or formal errors). The design of the study is illustrated in **Figure 1**.

The study was carried out in accordance with the Declaration of Helsinki and was approved by the local Ethical Committee of the Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany. All participants gave written informed consent and received 100 Euro to recompense them for travel expenses.

Multidomain Cognitive Training

Participants of the CT group were trained for 16 weeks, twice per week and 90 min per session. The CT was supervised by a professional trainer and a student assistant in small groups of no more than 12 participants. Participants who had missed regular sessions had the opportunity to take part in two additional sessions after the regular training had been completed. In the first step, participants were given basic information on cognitive functions and their relevance for the daily life to enhance compliance and motivation to train. The subsequent multidomain CT consisted of paper-and-pencil tasks and PC-based games. The difficulty of the tasks was adapted and adjusted to the current individual performance of the trainees.

The basic principle of the training was to construct exercises that are as diverse as possible. Therefore, each exercise was performed two times at most, which should ensure that participants find the training interesting and entertaining. It was not intended to train a particular cognitive ability, for example, spatial awareness or reasoning, but to employ different exercises requiring different cognitive abilities. Hence, each participant got the chance to become familiar with different types of training and to find out which ones were most suitable to him or her. Furthermore, every participant was motivated to integrate the training into his or her everyday life after the study was finished. While selecting the training programs, we took care that a low-cost continuation of the training is possible.

During the first 4 weeks, a mental activation training (MAT; Lehrl et al., 1994) that consisted of short paper-and-pencil exercises was used to increase working memory capacity, attention, and speed of information processing during a short span of 10–15 min. Moreover, in the first eight sessions, participants without any computer experience were familiarized with handling a computer. During the following 12 weeks, the participants worked on selected internet-based serious games[1]. See the **Appendix** for a detailed description of the games and the training schedule. Each session consisted of different games that aimed at training most relevant cognitive functions. The exercises and games mainly involved mnemonic functions, but some of them also included multitasking, logical thinking, cognitive flexibility, spatial reasoning, and aspects of executive functions. However, the elements of these games had no similarity to the psychometric tests used for pre- and post-testing. In particular, the CT consisted of several PC-based games, whereas the

[1]www.mentaga.de; www.ahano.de; www.mental-aktiv.de; www.freshminder.de

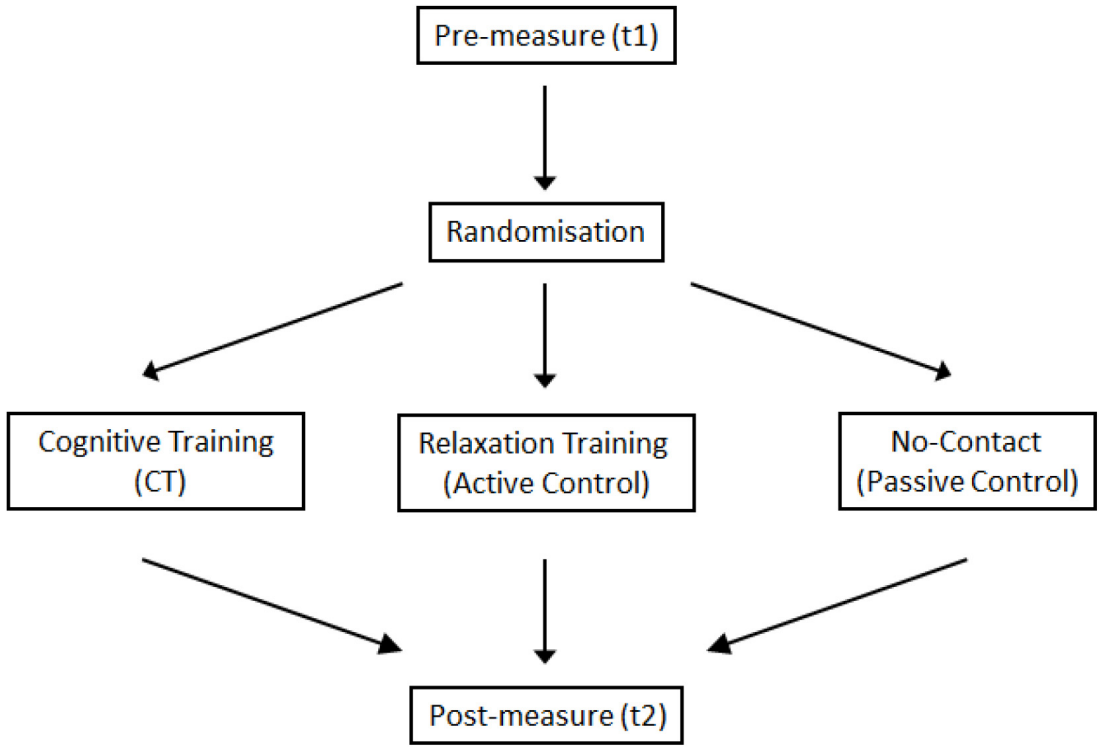


FIGURE 1 | Design of the study.

cognitive measures reflected well-structured, paper-and-pencil tests without apparent overlapping features with the games. In this case, a transfer effect should be assumed as transfer effects refer to the ability that individuals have to use the knowledge and skills learned in one scenario to achieve different goals in other scenarios. The participants were not encouraged to exercise beyond the training sessions during the study.

Active and Passive Control Groups

The participants of the active control group were trained similar to the CT group for a period of 4 months, twice per week and 90 min per session. This group received a relaxation training consisting of progressive muscle relaxation, autogenic training, breathing exercises, back training, massage, and Qigong. The aim of this training was to provide interesting and varied exercises which did not involve and hence should not train cognitive functions. The passive control group did not participate in any intervention.

Testing

Participants completed a number of sociodemographic questionnaires at home. During the test session, a number of paper-and-pencil psychometric tests that assessed a broad spectrum of cognitive abilities like perceptual speed, sustained and focused attention, short and working memory, long-term memory, interference processing, divergent thinking, and verbal abilities were used. After a 1-h break, a second test session using computer-based tasks with EEG recording was conducted, which was not the focus of analysis here (for further details, see Gajewski and Falkenstein, 2012, 2018; Wild-Wall et al., 2012; Küper et al., 2017).

Psychometric Tests

The focus of interest in the present study was fluid cognitive functions like focused and sustained attention, processing speed, cognitive flexibility, and interference control as one of the crucial executive functions. The selected tests are sensitive to subtle changes in cognitive performance of healthy adults and are mainly used in non-clinical populations. Thus, the tests are even more able to differentiate between persons with normal cognition and those with just the beginnings of MCI. The detailed description of the tests and their psychometric properties are provided in more detail in Gajewski et al. (2018) and are only shortly outlined in the following.

d2 Test

The d2 test (Brickenkamp, 1972) consists of a sheet presenting 14 lines of 47 letters (d and p), each with one to four dashes (‘), located either individually or in pairs above or below the letter. Participants were asked to go as fast as possible through each line and identify every d with two dashes by crossing it out. After 20 s of processing one line, the subjects were told to move on to the next line and to continue. The total number of correctly identified d’s with two dashes represents the test score. The d2 test is a measure of focused and sustained attention (attentional endurance) as well as processing speed.

Digit Symbol Test

The digit symbol test (DST) is an evaluation tool used to assess psychomotor functions. Initially, it was part of the Wechsler Adult Intelligence Test (WAIS; Wechsler, 1956). It consists of nine digit–symbol pairs (e.g., 7/Λ) followed by a list of digits. Under each digit, the subject should write down the corresponding symbol as fast as possible. The number of correct symbols produced within the 90 s reflects the test score. The DST measures processing speed, visuospatial processing, and selective attention.

Word Fluency Test

In the word fluency test (from LPS, Horn, 1983), participants are asked to recall as many words as possible within a given time, each word beginning with a specific letter. Three trials were conducted, in which words with the initial letters B, F, and L were asked for. For the post-measure, a parallel version was used, including the letters P, K, and S. Participants were given 60 s for each trial. The produced words were recorded by the experimenter. The total number of meaningful words represents the test result. The test measures the ability to access the verbal lexicon, semantic memory, scope of vocabulary, cognitive flexibility, and divergent thinking.

Stroop Test

The Stroop test (from NAI, Oswald and Fleischmann, 1986) consists of three parts. In the first part (Stroop 1), subjects were given a sheet of paper with names of colors printed in black ink. The participants were asked to read them out aloud as fast as possible. In the second part (Stroop 2), participants were handed another sheet of paper with colored bars on it. Participants were asked to name the colors. In the third condition (Stroop 3), subjects were given a sheet of paper with names of colors printed in various colors which did not match the names of the colors (e.g., "GREEN" was printed in red). Subjects had to name the colors the words were printed in as fast as possible and to ignore the meaning of the words. There was the same number of items in every condition. The time needed to perform each condition was analyzed. The time of the third list is considered an indicator of interference processing and inhibitory control as one of the most important executive functions. To assess the baseline-corrected interference, the time of Stroop 2 was subtracted from the time to complete Stroop 3.

Multiple-Choice Vocabulary Test (MWT-B)

The MWT-B (Lehrl, 1995) assesses crystalline intelligence. The test consists of 37 items with five words each. Only one of the five words reflects a meaningful word; the other similar words are meaningless. The subjects were required to mark the correct word. The difficulty of items increased with the increasing item number. The number of correctly identified meaningful words allows assessment of the crystalline IQ.

CFQ

To assess a possible far transfer of CT to the performance in daily activities, the CFQ (Broadbent et al., 1982) was used. CFQ is a scale including 26 questions related to attentional and memory lapses in daily life, for example, "Do you find you forget whether you've turned off a light or a fire or locked the door?" or "Do you fail to listen to people's names when you are meeting them?" Frequent lapses are scored higher than less frequent ones. The CFQ was exclusively conducted in the post-test to avoid test–retest effects (cf. Strobach and Huestegge, 2017).

Statistical Analysis

A series of two-way mixed analyses of variance with repeated measures (mixed ANOVAs) with the factors Group (CT vs. active control group and CT vs. passive control group) and Session (t1 vs. t2) were conducted. Significant interactions were further analyzed using *t*-tests. Mean values with standard errors of the mean are presented (±1 SEM). Estimators of effect size are provided by using partial eta square (η_p^2). To examine the training gains and transfer effects, we also calculated Cohen's (1977) d_z or the standardized mean difference in performance between pre- and post-test (Verhaeghen et al., 1992; see also Lakens, 2013). Therefore, the pre-test–post-test differences for the CT and the control groups were divided by the pooled standard deviation for pre- and post-measures for each test. An alternative way to calculate d_z is to divide the *t*-value of the paired-samples *t*-test by the squared number of persons (Lakens, 2013).

RESULTS

d2 Test

The repeated-measures ANOVA for the comparison between the CT and the passive control groups for the number of correctly crossed d2 symbols (**Figure 2**) indicated a main effect of Session (t1: 373 ± 9.6 vs. t2: 406 ± 10.5; $F[1, 66] = 27.0$, $p < 0.0001$, $\eta_p^2 = 0.290$) and no effect of Group ($F[1, 66] = 1.5$, $p = 0.22$, $\eta_p^2 = 0.023$). The interaction Group × Session was significant ($F[1, 66] = 7.4$, $p < 0.01$, $\eta_p^2 = 0.101$) and indicated a larger increase of correctly crossed d2s at t2 than at t1 in the CT group [426.3 ± 15.2 vs. 376.3 ± 14.0; $t(31) = 5.6$, $p < 0.0001$] relative to the passive control group [385.5 ± 14.4 vs. 369.9 ± 13.2; $t(35) = 1.7$, $p = 0.09$].

The number of correct symbols in the d2 test for the comparison between the CT and the active control group revealed a main effect of Session, indicating a larger number of crossed symbols at t2 than at t1 (t1: 370 ± 9.5 vs. t2: 403 ± 10.5; $F[1, 62] = 17.8$, $p < 0.0001$, $\eta_p^2 = 0.223$). No main effect of group was found ($F[1, 62] = 2.4$, $p = 0.12$, $\eta_p^2 = 0.038$), but there was an interaction between Group and Session ($F[1, 62] = 4.8$, $p < 0.05$, $\eta_p^2 = 0.072$), indicating a larger increase of the number of symbols at t2 compared to t1 in the CT group as outlined above [426.3 ± 15.2 vs. 376.3 ± 14.0; $t(31) = 5.6$, $p < 0.0001$] than in the active control group [377.9 ± 12.7 vs. 360.7 ± 12.7; $t(32) = 1.3$, $p = 0.18$].

DST

The total number of correctly substituted digit–symbol items for the comparison CT vs. passive control groups (**Figure 3**) increased from t1 to t2 (t1: 46.4 ± 1.1 vs. t2: 48.0 ± 1.0; $F[1, 67] = 5.2$, $p < 0.05$, $\eta_p^2 = 0.072$). There was no effect of Group ($F[1, 67] < 1$) and a weak trend for an interaction between Group and Session ($F[1, 67] = 3.2$, $p = 0.079$, $\eta_p^2 = 0.045$), indicating a slight increase of correct items from t1 to t2 [46.5 ± 1.6 vs. 49.9 ± 1.6; $t(31) = 2.7$, $p = 0.011$] in the CT group, while no substantial change was observed in the passive control group [46.4 ± 1.5 vs. 46.9 ± 1.5; $t(36) < 1$].

The total number of correctly substituted digit–symbol items for the comparison CT vs. active control increased from t1 to t2 (t1: 45.4 ± 1.0 vs. t2: 47.4 ± 1.1; $F[1, 63] = 4.4$, $p < 0.05$, $\eta_p^2 = 0.066$). There was no main effect of Group ($F[1, 63] = 2.5$, $p = 0.118$, $\eta_p^2 = 0.038$) and again a weak trend for an interaction between Group and Session ($F[1, 63] = 2.9$, $p = 0.093$, $\eta_p^2 = 0.044$), suggesting a slight increase of correct items from t1 to t2 [46.5 ± 1.6 vs. 49.9 ± 1.6; $t(31) = 2.7$, $p = 0.011$] in the CT group, while no substantial change was observed in the active control group [44.4 ± 1.5 vs. 44.7 ± 1.6; $t(34) < 1$].

Word Fluency Test

The total number of words for the comparison between CT and passive control groups (**Figure 4**) increased from t1 to t2 (t1: 43.6 ± 1.2 vs. t2: 46.8 ± 1.5; $F[1, 67] = 7.2$, $p < 0.01$, $\eta_p^2 = 0.097$). There was no effect of Group ($F[1, 67] = 1.7$, $p = 0.19$, $\eta_p^2 = 0.025$). Descriptively, the mean number of words produced by participants of the CT group increased from t1 to t2 [44.4 ± 1.7 vs. 49.4 ± 2.2; $t(31) = 2.6$, $p = 0.015$], while the control group showed only a small increase [43.0 ± 1.6 vs. 44.3 ± 2.1; $t(36) < 1$]. However, the interaction Group × Session did not reach significance ($F[1, 67] = 2.3$, $p = 0.13$, $\eta_p^2 = 0.033$).

The number of words for the comparison between CT and active control groups increased only slightly from t1 to t2 (t1: 41.9 ± 1.3 vs. t2: 44.0 ± 1.6; $F[1, 63] = 3.0$, $p = 0.085$, $\eta_p^2 = 0.046$). In contrast to the comparison above, there was an effect of Group ($F[1, 63] = 8.2$, $p < 0.01$, $\eta_p^2 = 0.115$), due to the generally larger number of words in the CT compared to the active control group (46.9 ± 1.9 vs. 39.9 ± 1.9). Moreover, there was an interaction between Group and Session ($F[1, 63] = 5.5$, $p < 0.05$, $\eta_p^2 = 0.081$), suggesting that the mean number of words produced by participants of the CT group increased from t1 to t2 [44.4 ± 1.7 vs. 49.4 ± 2.2; $t(31) = 2.6$, $p = 0.015$], while the active control group showed no effect [39.4 ± 1.8 vs. 38.6 ± 2.3; $t(34) < 1$].

Stroop Test

In the first step, a series of ANOVAs were conducted for the CT vs. passive control group and CT vs. active control group for parts 1, 2, and 3 of the Stroop test.

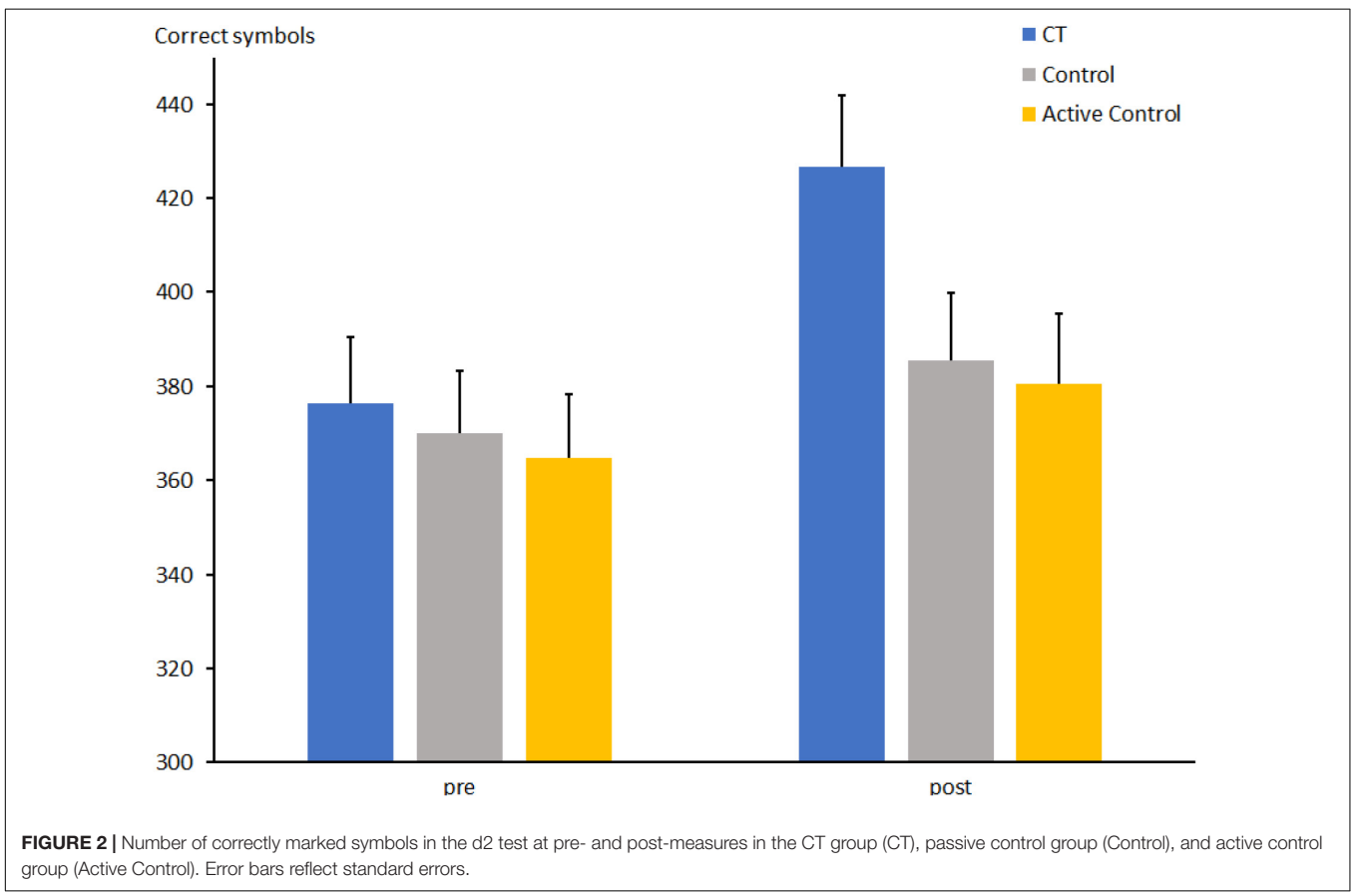


FIGURE 2 | Number of correctly marked symbols in the d2 test at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

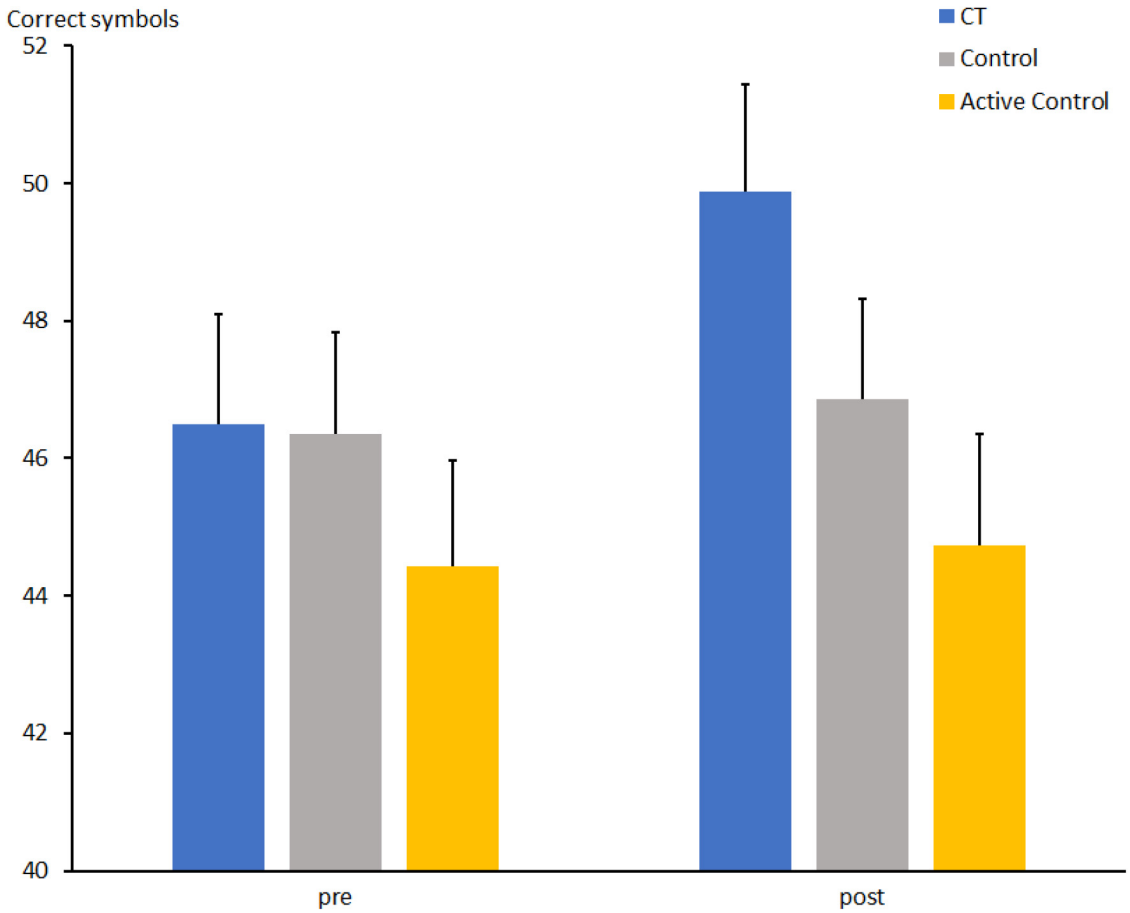


FIGURE 3 | Total number of correctly substituted symbols in the DST at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

In part 1 of the Stroop test (**Figure 5**), requiring reading of color words in black ink, the comparison of CT and passive control groups did not show effects of Session (t1: 14.1 ± 0.3 s, t2: 13.7 ± 0.3 s; $F[1, 65] = 1.8$, $p = 0.18$, $\eta_p^2 = 0.027$) or Group ($F[1, 65] < 1$) or an interaction between Group and Session ($F[1, 65] = 1.9$, $p = 0.17$, $\eta_p^2 = 0.025$).

For the comparison between the CT and active control groups in part 1, no effect of Session was observed (t1: 14.8 ± 0.4 s, t2: 14.6 ± 0.3 s; $F[1, 64] < 1$, $\eta_p^2 = 0.006$). However, there was an effect of Group due to faster responses in the CT than in the active control group (14.0 ± 0.5 s, t2: 15.2 ± 0.5 s; $F[1, 64] = 4.25$, $p < 0.05$, $\eta_p^2 = 0.063$). The interaction Group × Session ($F[1,64] = 2.3$, $p = 0.13$, $\eta_p^2 = 0.035$) did not reach significance. However, the main effect of Group ($F[1, 64] = 4.2$, $p < 0.05$, $\eta_p^2 = 0.062$) yielded slower responses in the active control than in the CT group (15.5 ± 0.5 vs. 14.0 ± 0.5 s).

The time to name color blocks in part 2 of the Stroop test for the CT and passive control groups (**Figure 6**) was significantly shorter at t2 than at t1 (20.6 ± 0.4 vs. 21.5 ± 0.4; $F[1, 65] = 6.7$, $p = 0.012$, $\eta_p^2 = 0.094$). No Group effect ($F[1, 65] < 1$) or interaction between Group and Session was found ($F[1, 65] = 1.3$, $p = 0.25$, $\eta_p^2 = 0.020$).

For the CT and active control groups, the time to name color blocks in part 2 of the Stroop test was shorter at t2 than at t1 (21.4 ± 0.4 vs. 22.3 ± 0.6; $F[1, 63] = 4.5$, $p < 0.05$, $\eta_p^2 = 0.066$). No interaction between Group and Session was found ($F[1, 63] < 1$, $\eta_p^2 = 0.008$), but again, a main effect of Group was found ($F[1, 63] = 4.8$, $p < 0.05$, $\eta_p^2 = 0.07$), indicating faster responses in CT than in the active control group (CT: 20.8 ± 0.7 vs. 22.9 ± 0.6 s).

For the CT and the passive control groups, the time for processing the interference list in part 3 (**Figure 7**) consisting of incongruent color words was similar at t2 and t1 (41.6 ± 1.2 vs. 43.1 ± 1.2 s; $F[1, 65] = 2.6$, $p = 0.13$, $\eta_p^2 = 0.038$). No Group effect was found ($F[1, 65] = 1.2$, $p = 0.28$, $\eta_p^2 = 0.018$). However, the interaction Group × Session was significant ($F[1, 65] = 5.4$, $p = 0.023$, $\eta_p^2 = 0.077$), indicating faster performance at t2 compared to t1 in the CT group [39.3 ± 1.7 vs. 43.5 ± 1.6; $t(31) = 2.4$, $p = 0.022$], but no substantial difference was detected in the passive control group [43.8 ± 1.5 vs. 43.2 ± 1.6 s; $t(35) < 1$].

The time for processing the interference list in part 3 for the CT and active control groups was shorter at t2 than at t1 (41.0 ± 0.8 vs. 44.0 ± 1.2 s; $F[1, 63] = 9.5$, $p = 0.005$, $\eta_p^2 = 0.129$). No Group effect ($F[1, 63] = 2.3$, $p = 0.13$, $\eta_p^2 = 0.035$) or

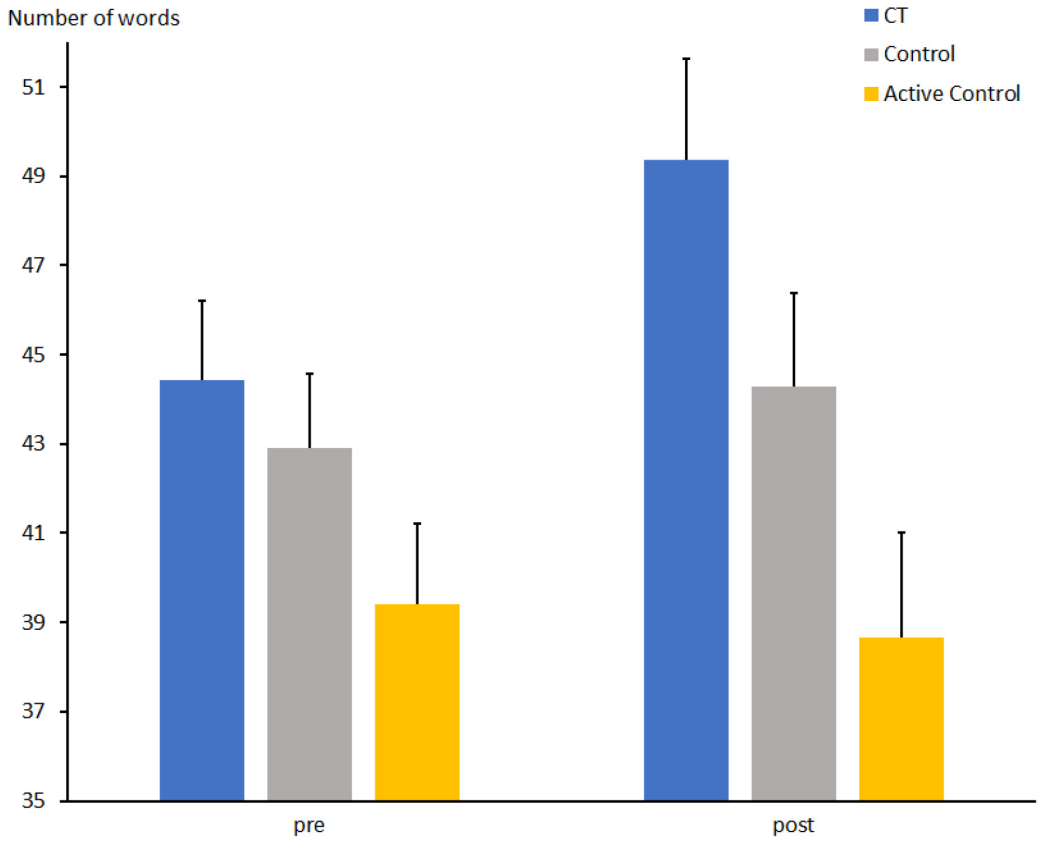


FIGURE 4 | Total number of correctly produced words in the word fluency task at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

interaction between Group and Session was found ($F[1, 63] < 1$, $\eta_p^2 = 0.006$).

Finally, we analyzed the time difference between Stroop 3 and Stroop 2 to assess the baseline-corrected interference effect (see **Figure 8**). For the comparison of the CT versus passive control groups, no effect of Session ($F[1, 65] < 1$) or Group was obtained ($F[1, 65] = 1.4$, $p = 0.24$, $\eta_p^2 = 0.021$). Yet the interaction Group × Session was corroborated ($F[1, 65] = 4.1$, $p < 0.05$, $\eta_p^2 = 0.046$), indicating a trend for reduction of interference in the CT group between pre- and post-measures [21.5 ± 0.9 vs. 19.0 ± 0.9 s, $t(31) = 1.8$, $p = 0.076$] and a slight increase between pre- and post-measures in the passive control group [21.6 ± 0.9 vs. 22.8 ± 0.9 s; $t(36) < 1$].

For the comparison between CT and active control, no Group effect or interaction between Group and Session was found (both $Fs < 1$). Nevertheless, the effect of Session was corroborated ($F[1, 63] = 6.0$, $p < 0.05$, $\eta_p^2 = 0.085$), indicating a substantial reduction of interference in both groups between pre- and post-measures (21.7 ± 0.9 vs. 19.5 ± 0.7 s).

MWT-B

The number of correctly detected words in the test measuring crystallized intelligence (**Figure 9**) for the comparison between CT and passive control groups showed no effect of Session or Group (both $Fs[1, 67] < 1$) and only a weak trend for the interaction Group × Session ($F[1, 67] = 3.0$, $p = 0.086$, $\eta_p^2 = 0.043$).

Similarly, no effect of Group or Session (both $Fs < 1$) or interaction of both factors ($F[1, 63] = 1.9$, $p = 0.16$, $\eta_p^2 = 0.029$) for the comparison between CT and the active control groups was found.

Training-Related Gains

To assess the training-related gains and transfer effects, Cohen (1977) d_z was calculated by computing pre-test minus post-test differences divided by the pooled standard deviations for pre- and post-tests (Lakens, 2013). A d_z of 1 indicates that the mean difference between pre- and post-test corresponds to one standard deviation. The results for the CT and the control groups are presented in **Table 1**. As can be seen, the training gains occurred in all of the applied tests after CT relative to the control groups.

Far Transfer to Daily Activities

No group differences in the CFQ were found at the post-test measurement (CT group: 29.2, passive control group: 28.4, active control group: 27.2; both F's < 1).

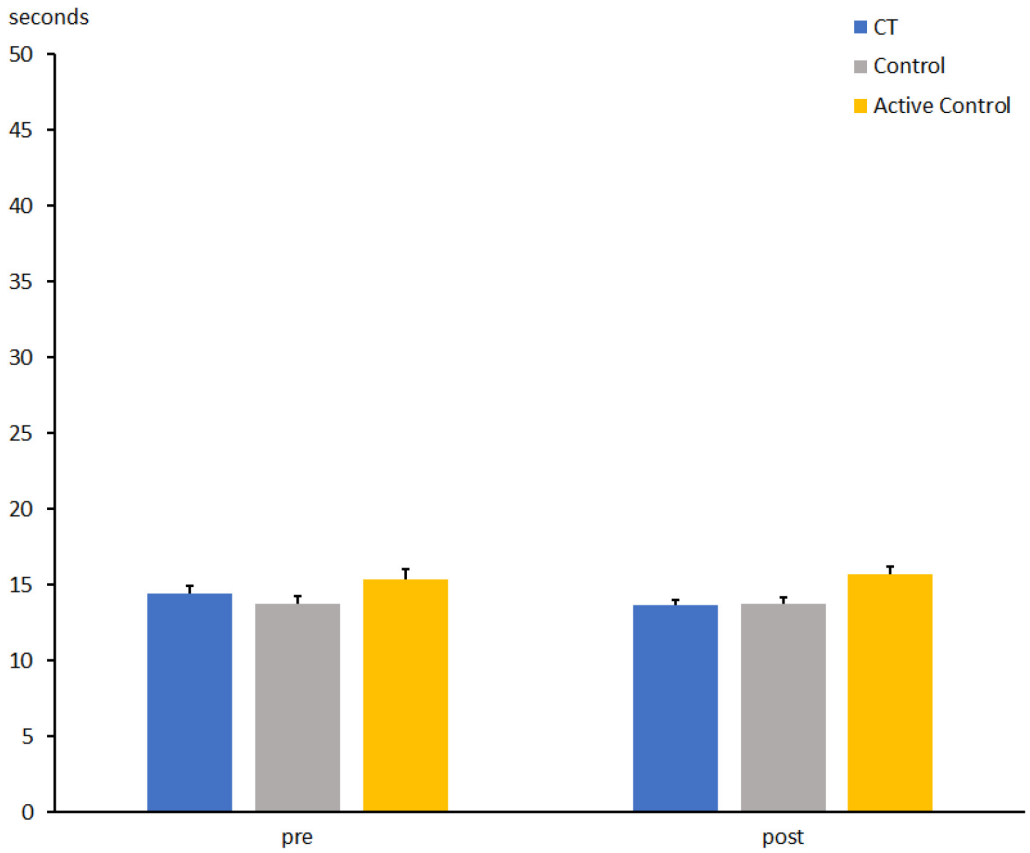


FIGURE 5 | Time to perform part 1 of the Stroop task (reading of color words) at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

DISCUSSION

The present study evaluated effects of a 16-week multidomain CT on different cognitive functions in older participants. Compared to an active control group receiving relaxation training and a passive (no-contact) control group, participants in the CT group showed enhancement of sustained and focused attention (attentional endurance) as assessed by the d2 test. The same pattern was found for the digit symbol substitution test assessing focused attention and psychomotor speed, although the interaction Group × Session showed only a trend compared to both control groups. However, whereas no significant interaction was observed in the word fluency test after CT compared to the passive control group, this difference was substantial in comparison to the active control group. In the simple tasks of the Stroop task (word reading and color naming), no interactions between Group and Session were found (though descriptively, the participants of the CT group showed faster responses than did the control groups). Interestingly, the results of the interference list (Stroop 3), in which interference processing and inhibitory control were required, revealed divergent patterns: while there was a similar improvement of performance in the CT and in the active control groups from pre- to post-measure, the participants of the CT group showed a clear performance improvement after training compared to the passive controls. No effects or interactions were found in the MWT-B that evaluates crystallized intelligence. Thus, crystallized cognitive functions were not improved after training. The findings indicate also that basic cognitive functions like sensory abilities and psychomotor speed were not affected by the CT, whereas higher-order cognitive functions were enhanced. Additionally, though some interactions between Group and Session did not reach significance, the effect sizes of the CT group were consistently larger compared to those of the control groups. For example, the improvement in the d2 task after CT corresponded to one standard deviation. Taken together, these observations indicate that a multidimensional training consisting of a number of short exercises is suitable to improve some attentional and executive functions in older adults. The findings are consistent with a number of previous training studies reporting positive effects of CT on executive functions in older age (Green and Bavelier, 2003; Brehmer et al., 2014; Karbach and Verhaeghen, 2014; Lampit et al., 2014; Ballesteros et al., 2014, 2015; Green et al., 2016).

The comparison between the CT and the two control groups provided similar effects, but there were also some interesting differences in the results. First, word fluency was improved after CT and reached significance relative to the active control group but did not reach significance compared to the passive

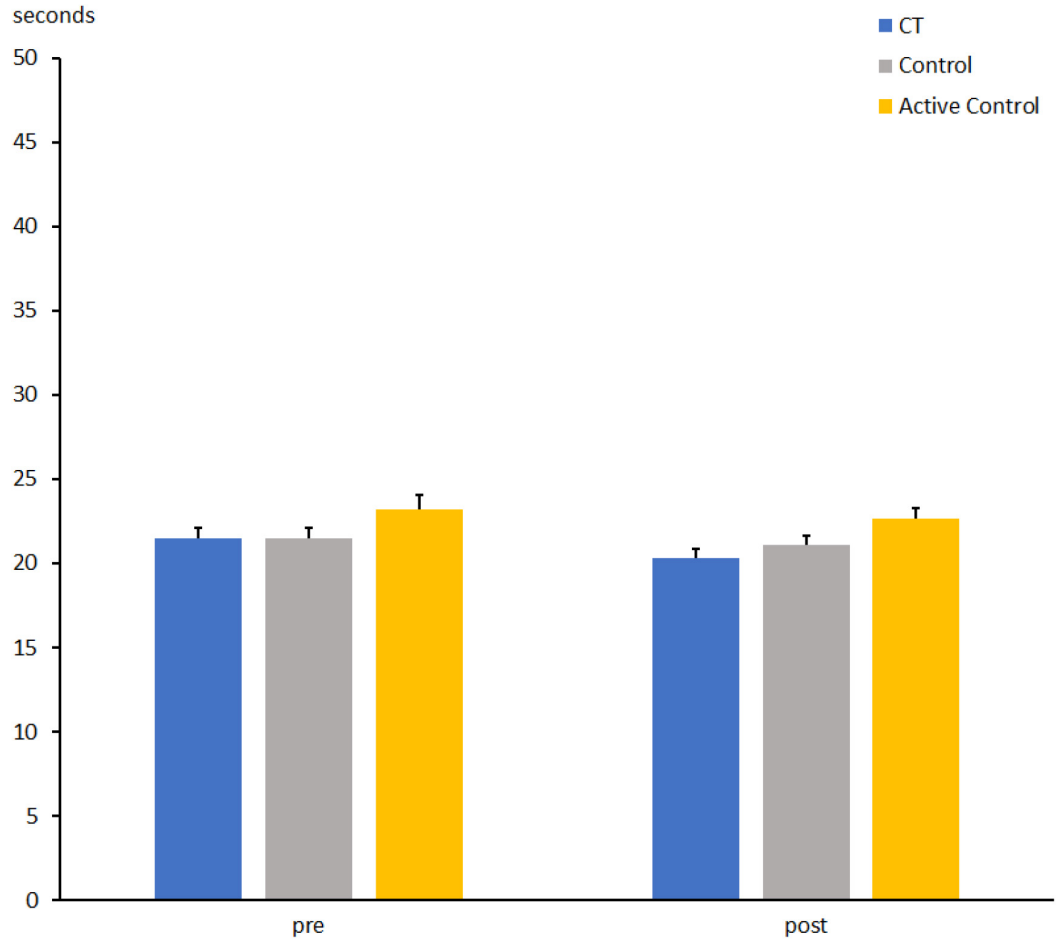


FIGURE 6 | Time to perform part 2 of the Stroop task (naming of color squares) at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

control group. In contrast, performance in the interference list of the Stroop test was significantly faster in comparison to the passive control group, but not to the active control group. While the first observation was probably due to differences in power, the unspecific enhancement in the interference list of Stroop 3 suggests that relaxation training, which was used in the present study as active control and was supposed not to involve cognitive resources, may reduce distractibility (Yesavage and Jacob, 1984). Further research may shed more light on the relationship between relaxation and interference processing.

Our study also aimed at transfer effects. Near transfer indicates performance enhancement in structurally similar tasks involving overlapping neural circuits, whereas far transfer denotes effects on structurally dissimilar tasks (Karbach and Kray, 2009; Karbach and Verhaeghen, 2014; Strobach and Huestegge, 2017). Given that the CT in the present study used games and exercises that clearly differed from the psychometric tests used to evaluate the training effects, we assume to observe some far-transfer effects. It is important to stress the point that the CT did not include any explicit elements of the psychometric tests used for the pre- and post-measures (see the **Appendix** for a detailed description of the exercises). Thus, increases in post-test relative to pre-test measures of performance would suggest far transfer of training-related gains to specific cognitive functions measured by the tests. This assumption was also supported by the medium to large effect sizes in performance differences between pre- and post-tests after CT. Differences in effect sizes between tests suggest that performance in a test assessing a particular cognitive function benefited more from the training than performance assessed by another test. Therefore, we assume that attentional functions benefited more from CT than crystallized cognitive functions.

However, it has to be noted that near and far transfers have been differently defined in the literature. Some authors proposed that far transfer is given when dissimilar tasks were used for training and evaluating CT effects. For example, Karbach and Kray (2009) used task switching training and evaluated far transfer in a Stroop task that may share the same functional components (e.g., inhibition). Other authors proposed that far transfer occurs when no functional overlaps are shared, for example, when CT improved daily functions in individuals following a training of reasoning (Willis et al., 2006). Barnett and Ceci (2002) proposed a systematic taxonomy of near and

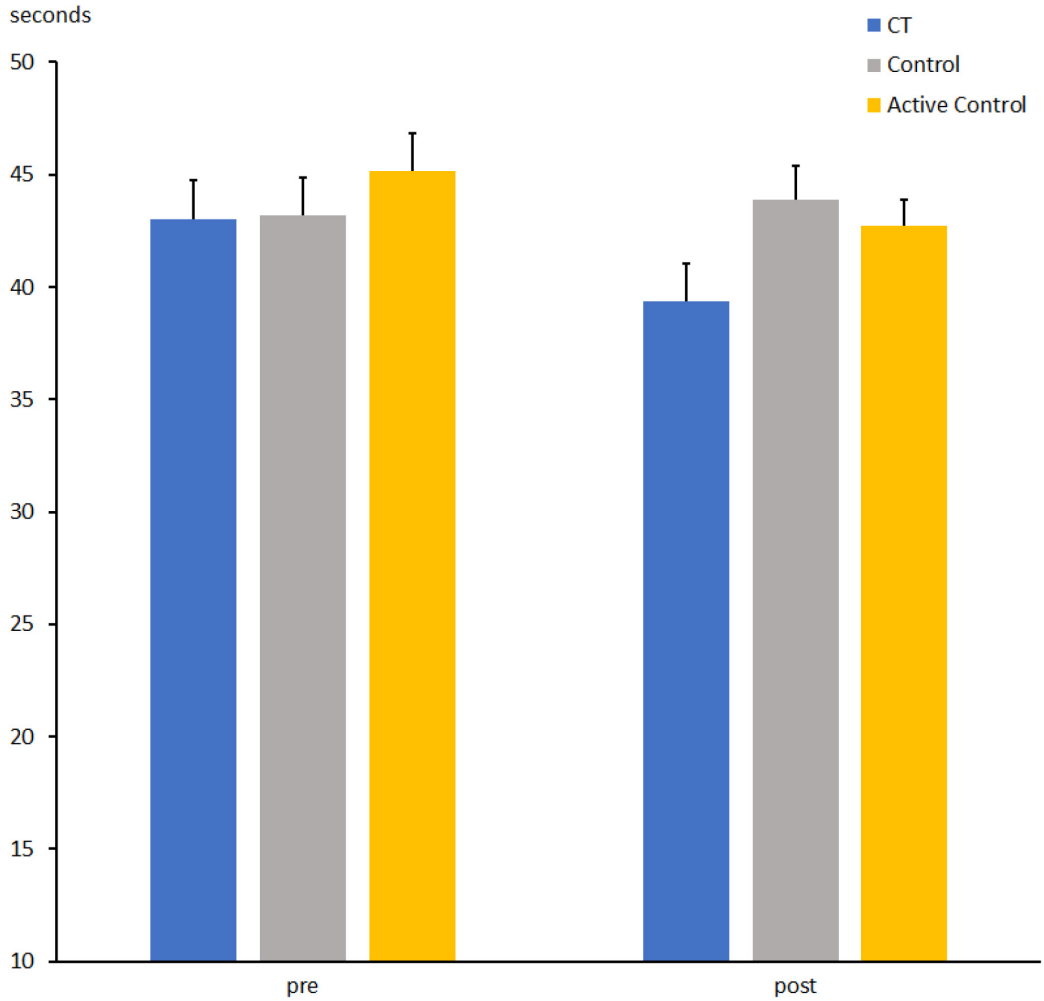


FIGURE 7 | Time to perform part 3 of the Stroop task (interference list) at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

far transfers, based on the contextual similarity and dissimilarity of training and test. A clear evidence of (very) far transfer in our study would be a positive effect of multidomain CT on performance in daily activities, as assessed by the CFQ that evaluated attentional and memory lapses in daily life. However, consistent with previous results using CFQ (Strobach and Huestegge, 2017), we did not find group differences, demonstrating similar absent-mindedness and slips of action in all groups of participants. This suggests no far transfer of our CT to daily activities at least in the cognitive functions measured by CFQ.

Nevertheless, as outlined above, far transfer to dissimilar cognitive tasks was found in the present study. This is in line with behavioral and electrophysiological findings obtained in the same study and reported previously, including computer-based versions of task switching (Gajewski and Falkenstein, 2012), Stroop task (Küper et al., 2017), *n*-back task (Gajewski and Falkenstein, 2018), and visual search task (Wild-Wall et al., 2012). All these tasks represent different executive functions but presumably share some characteristics involved in all these tasks. This could be reflected by attentional processes necessary to select particular stimuli but also by cognitive control of action required for successful response selection and execution. There are also indications that CT leads to cognitive improvements in terms of lower within-subject variability in performance (higher consistency) in specific executive processes like response selection that may be supported by attentional gating (Gajewski and Falkenstein, 2012; Gajewski et al., 2017; Küper et al., 2017). In other words, CT may improve selection of task-relevant stimuli and the corresponding activation of stimulus–response mapping, which is critical for successful task performance. This, in turn, may rely on more efficient synaptic transmission or denser neuronal network supporting these functions.

Overall, transfer effects are controversially discussed as the results are rather inconsistent (Melby-Lervåg et al., 2016). It seems that the amount of transfer as well as the efficacy of CT is determined by the type of training (e.g., its total duration and intensity), the participants' baseline performance,

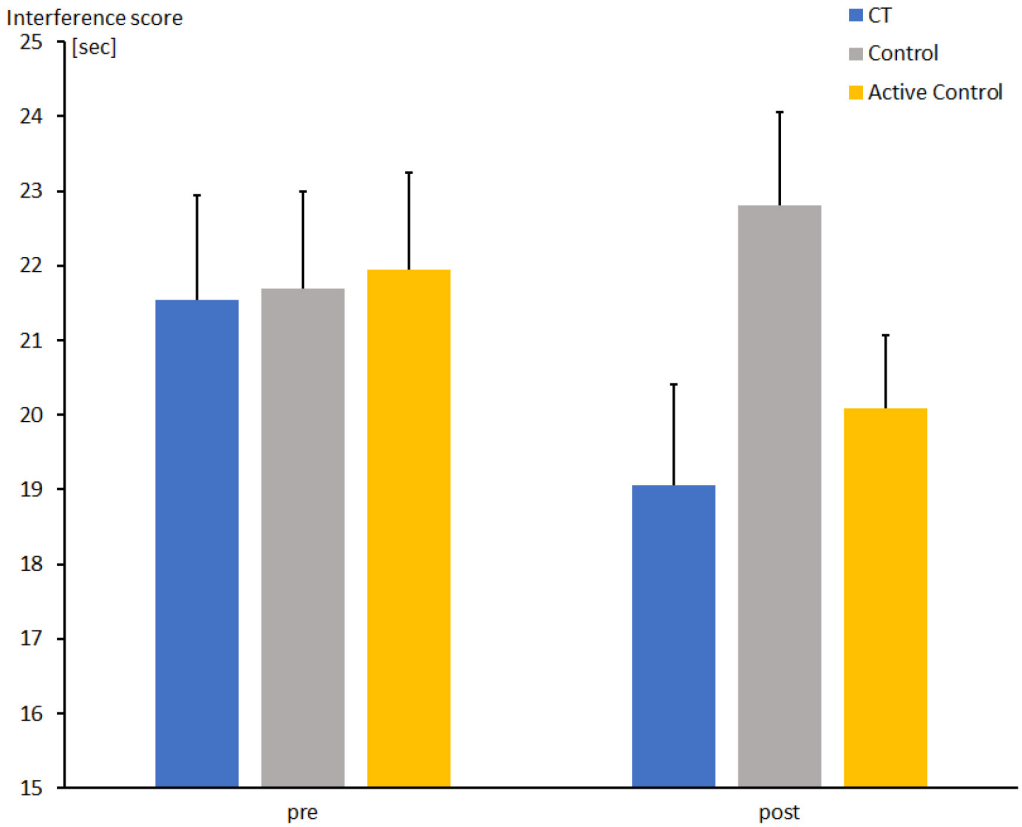


FIGURE 8 | Interference score in seconds (Stroop 3–Stroop 2) at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

whether or not the training is supervised (Lampit et al., 2014), and probably most important, the intrinsic motivation of the participants (Jaeggi et al., 2014; von Bastian and Oberauer, 2014). Additionally, differences in the efficacy and transfer of CT may be due methodological heterogeneity and structural differences between studies, depending on the trained cognitive domains. A further crucial factor for inconsistency of results is related to the measures used to evaluate training gains like standardized psychometric tests, computer-based tasks, and their specific parameters that often vary between different studies. From this perspective, it would be useful to determine characteristics and to establish guidelines for an optimal combination of factors to enhance the efficacy of CT based on a review of meta-analyses (cf. Bavelier and Green, 2019). For example, a meta-analysis by Lampit et al. (2014) suggests a group-based, multidomain training with session length between 30 and 60 min and frequency of two to three sessions per week to be most effective. The design of the present study met most of these criteria. In addition, regarding the crucial measures, it appears important to use sensitive tasks or psychometric tests with a wide range of possible outcomes to avoid ceiling effects and to ensure that possible improvements by CT can be detected reliably.

TABLE 1 | Cohens' d_z reflecting training gains in the CT group (CT) and the passive and active control groups.

Psychometric task	CT	Passive control	Active control
d2 test	1.00	0.29	0.24
Digit symbol test	0.48	0.06	0.08
Word fluency test	0.46	0.15	0.07
Stroop task (part 1)	−0.25	−0.01	−0.11
Stroop task (part 2)	−0.41	−0.19	−0.16
Stroop task (part 3)	−0.43	−0.09	−0.32
Stroop (part 3–part 2)	−0.32	−0.01	−0.27
MWT-B	0.30	0.12	0.05

Positive values reflect an increased number of correctly performed symbols or words; negative values are due to faster performance in the Stroop task at post- vs. pre-test.

Limitations

There are some limitations of the study that have to be acknowledged. Firstly, the functional overlap between the training and test tasks is not easy to evaluate because of a variety of different games and subtasks that were used. Additionally, paper-and-pencil trainings that trained working memory and speed of processing as well logical thinking were used. The goal of our multidomain training was to maximally enhance cognitive functioning in older adults by offering them an interesting and

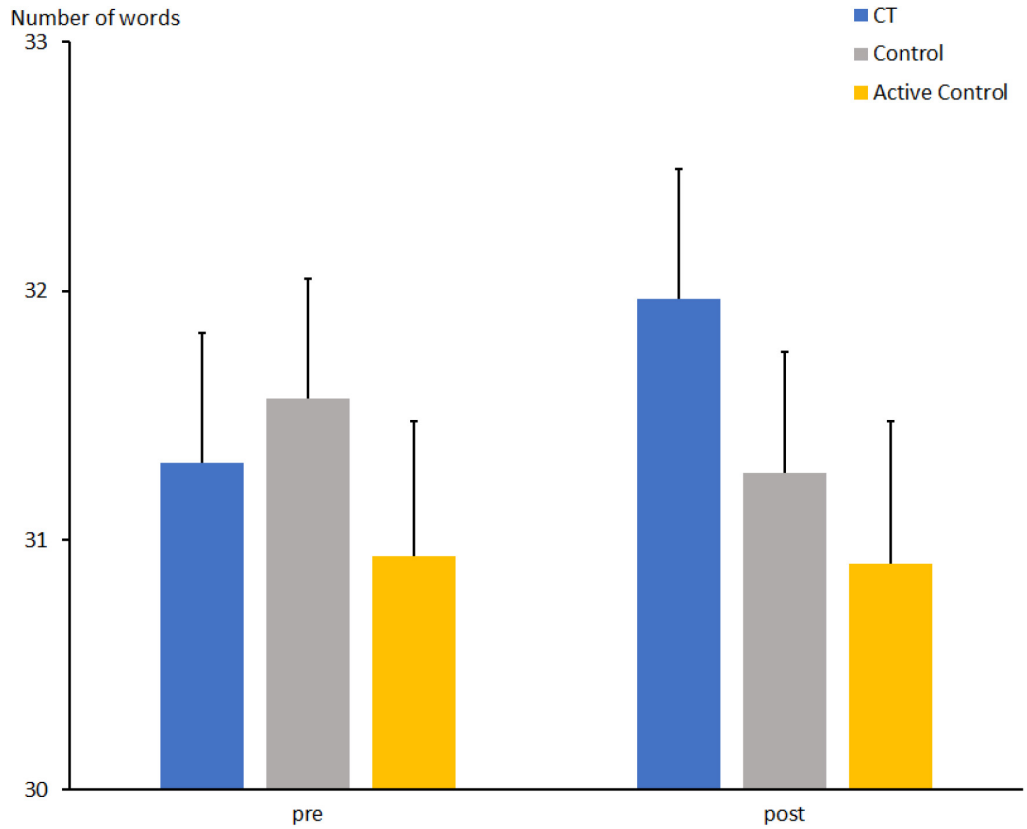


FIGURE 9 | Total number of correctly crossed words in the MWT-B at pre- and post-measures in the CT group (CT), passive control group (Control), and active control group (Active Control). Error bars reflect standard errors.

varying program with a large fun factor to fill more than 30 training sessions and to maintain their motivation to train across this relatively long time.

Secondly, it was not possible to evaluate individual training data to assess interindividual differences in the progress of training-induced gains. This was due to partly using a number of freely available internet games that did not record or permanently store the training data. Moreover, the training difficulty was individually adjusted to the training progress of the participants who absolved individually different training units, making the analysis and interpretation less meaningful.

CONCLUSION

The present study provides further evidence for positive effects of trainer-guided multidomain CT on fluid intelligence in older age. In particular, a 16-week-long CT substantially improved attentional endurance and interference processing. These effects were specifically evident for tests that require verbal fluency or selection of predefined targets and ignoring distractors, indicating enhancement of executive functioning like inhibitory control. Moreover, a transfer effect was observed to non-explicitly trained functions, but not to performance in daily life activities. The study contributes to the literature showing positive effects of CT in old age, especially on executive functions that reflect a crucial aspect of cognitive performance highly susceptible to aging.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article can be provided by the authors on request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Committee of the Leibniz Research Centre for Working Environment and Human Factors, Dortmund. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PG designed the study, analyzed the data, and wrote the manuscript. MF designed the study. ST, EW, and SG wrote and revised the manuscript. All authors approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Claudia Frieg, Ines Mombrei, and Christiane Westedt for conducting the testing and Ludger Blanke for developing the software and technical support. We thank the trainers of the CT, Rita Pfeiffer, and of the relaxation training, Antje Dreikaus. We would also like to thank the reviewers for their valuable comments on the previous version of the manuscript. The research reported in the present article was partly funded by the German Insurance Association (GDV, Gesamtverband der Deutschen Versicherungswirtschaft). The publication was supported by the open-access fund of the Leibniz Association.

REFERENCES

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101. doi: 10.1038/nature12486

Ball, K., Edwards, J. D., and Ross, L. A. (2007). The impact of speed of processing training on cognitive and everyday functions. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 62, 19–31. doi: 10.1093/geronb/62.special_issue_1.19

Ballesteros, S., Kraft, E., Santana, S., and Tziraki, C. (2015). Maintaining older brain functionality: a targeted review. *Neurosci. Biobehav. Rev.* 55, 453–477. doi: 10.1016/j.neubiorev.2015.06.008

Ballesteros, S., Prieto, A., Mayas, J., Toril, P., Pita, C., Ponce, et al. (2014). Brain training with non-action video games enhances aspects of cognition in older adults: a randomized controlled trial. *Front. Aging Neurosci.* 6:277. doi: 10.3389/fnagi.2014.00277

Baltes, P. B. (1987). Theoretical propositions of life-span developmental psychology: on the dynamics between growth and decline. *Dev. Psychol.* 23, 611–626. doi: 10.1037/0012-1649.23.5.611

Baltes, P. B., Sowarka, D., and Kliegl, R. (1989). Cognitive training research on fluid intelligence in old age: what can older adults achieve by themselves? *Psychol. Aging* 4, 217–221. doi: 10.1037/0882-7974.4.2.217

Bamidis, P. D., Vivas, A. B., Styliadis, C., Frantzidis, C., Klados, M., Schlee, W., et al. (2014). A review of physical and cognitive interventions in aging. *Neurosci. Biobehav. Rev.* 44, 206–220.

Barnett, S. M., and Ceci, S. J. (2002). When and where do we apply what we learn? a taxonomy for far transfer. *Psychol. Bull.* 128, 612–637. doi: 10.1037/0033-2909.128.4.612

Bavelier, D., and Green, C. S. (2019). Enhancing attentional control: lessons from action video games. *Neuron*, 104, 147–163. doi: 10.1016/j.neuron.2019.09.031

Brehmer, Y., Kalpouzos, G., Wenger, Y. E., and Lövden, M. (2014). Plasticity of brain and cognition in older adults. *Psychol. Res.* 78, 790–802. doi: 10.1007/s00426-014-0587-z

Brickenkamp, R. (1972). *d2-Aufmerksamkeits-Belastungs-Test*. Göttingen: Hogrefe.

Broadbent, D. E., Cooper, P. F., FitzGerald, P., and Parkes, K. R. (1982). The cognitive failures questionnaire (CFQ) and its correlates. *Br. J. Clin. Psychol.* 21, 1–16. doi: 10.1111/j.2044-8260.1982.tb01421.x

Cassavaugh, N., and Kramer, A. F. (2009). Transfer of computer-based cognitive training to simulated driving in older adults. *Appl. Ergon.* 40, 943–952. doi: 10.1016/j.apergo.2009.02.001

Cheng, Y., Wu, W., Feng, W., Wang, J., Chen, Y., Shen, Y., et al. (2012). The effects of multi-domain versus single-domain cognitive training in nondemented older people: a randomized controlled trial. *BMC Med.* 10:30. doi: 10.1186/1741-7015-10-30

Chiu, H. L., Chu, H., Tsai, J. C., Liu, D., Chen, Y. R., Yang, H. L., et al. (2017). The effect of cognitive-based training for the healthy older people: a meta-analysis of randomized controlled trials. *PLoS One* 12:e0176742. doi: 10.1371/journal.pone.0176742

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences (rev. edn.)*. New York, NY: Academic Press.

Craik, F. I. M., and Salthouse, T. A. (2000). *The Handbook of Aging and Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Falkenstein, M., and Gajewski, P. D. (2016). "Changes of electrical brain activity due to cognitive training in old adults and older industrial workers," in *Cognitive training: An overview of features and applications*, eds T. Strobach and J. Karbach (New York, NY: Springer), 177–186. doi: 10.1007/978-3-319-42662-4_17

Gajewski, P. D., and Falkenstein, M. (2012). Training-induced improvement of response selection and error detection in aging assessed by task switching: effects of cognitive, physical and relaxation training. *Front. Hum. Neurosci.* 6:130. doi: 10.3389/fnhum.2012.00130

Gajewski, P. D., and Falkenstein, M. (2016). "Lifestyle and interventions for improving cognitive performance in older adults," in *Performance psychology: Perception, Action, Cognition, and Emotion*, eds M. Raab, B. Lobinger, S. Hoffmann, A. Pizzera, and S. Laborde (Oxford: Elsevier), 189–203. doi: 10.1016/b978-0-12-803377-7.00012-0

Gajewski, P. D., and Falkenstein, M. (2018). ERP and behavioral effects of physical and cognitive training on working memory in aging: a randomized controlled study. *Neural. Plast.* 2018, 1–12. doi: 10.1155/2018/3454835

Gajewski, P. D., Freude, G., and Falkenstein, M. (2017). Cognitive training sustainably improves executive functioning in middle aged industry workers assessed by task switching: a randomized controlled ERP study. *Front. Hum. Neurosci.* 11:81. doi: 10.3389/fnhum.2017.00081

Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S., Wascher, E. (2018). What does the *n*-back task measure as we get older? Relation between working-memory measures and other executive functions as a function of age. *Front. Psychol.* 9:2208. doi: 10.3389/fpsyg.2018.02208

Green, C. S., and Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature* 423, 534–537. doi: 10.1038/nature01647

Green, C. S., and Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *J. Exp. Psychol. Human Percept. Perform*, 32, 1465–1478. doi: 10.1037/0096-1523.32.6.1465

Green, C. S., Gorman, T., and Bavelier, D. (2016). "Action video-game training and its effects on perception and attentional control," in *Cognitive Training: An Overview of Features and Applications*, eds T. Strobach and J. Karbach (New York, NY: Springer).

Greenwood, P. M., and Parasuraman, R. (2010). Neuronal and cognitive plasticity: a neurocognitive framework for ameliorating cognitive aging. *Front. Aging Neurosci.* 2:150. doi: 10.3389/fnagi.2010.00150

Horn, J. L., and Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychol.* 26, 107–129. doi: 10.1016/0001-6918(67)90011-x

Horn, W. (1983). *L-P-S Leistungsprüfsystem*. Göttingen: Hogrefe.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proc. Natl. Acad. Sci. Unit. States. Am.* 105, 6829–6833.

Jaeggi, S. M., Buschkuehl, M., Shah, P., and Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Mem. Cognit.* 42, 464–480. doi: 10.3758/s13421-013-0364-z

Karbach, J., and Kray, J. (2009). How useful is executive control training? age differences in near and far transfer of task-switching training. *Dev. Sci.* 12, 978–990. doi: 10.1111/j.1467-7687.2009.00846.x

Karbach, J., and Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive-control and working memory training in older adults. *Psychol. Sci.* 25, 2027–2037. doi: 10.1177/0956797614548725

Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., and Brennan, S. (2014). The impact of exercise on the cognitive functioning of healthy older adults: a systematic review and meta-analysis. *Ageing Res. Rev.* 16, 12–31. doi: 10.1016/j.arr.2014.05.002

Kueider, A. M., Parisi, J. M., Gross, A. L., and Rebok, G. W. (2012). Computerized cognitive training with older adults: a systematic review. *PLoS One* 7:e40588. doi: 10.1371/journal.pone.0040588

Küper, K., Gajewski, P. D., Frieg, C., and Falkenstein, M. (2017). A randomized controlled ERP study on the effects of multi-domain cognitive training and task difficulty on task switching performance in older adults. *Front. Hum. Neurosci.* 11:184. doi: 10.3389/fnhum.2017.00184

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

Lampit, A., Hallock, H., and Valenzuela, M. (2014). Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. *PLoS Med.* 11:e1001756. doi: 10.1371/journal.pmed.1001756

Lehrl, S. (1995). *Mehrfach-Wortwahl-Test (MWT)*. Erlangen: Medizinische Verlagsgesellschaft.

Lehrl, S., Lehrl, M., and Weickmann, E. (1994). *MAT Gehirnjogging, Einführung in das Mentale AktivierungsTraining*. Ebersberg: Vless Verlag.

Melby-Lervåg, M., and Hulme, C. (2013). Is working memory training effective? a meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228

Melby-Lervåg, M., Redick, T., and Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": evidence from a meta-analytic review. *Perspect. Psychol. Sci.* 11, 512–534. doi: 10.1177/1745691616635612

Minear, M., and Shah, P. (2008). Training and transfer effects in task switching. *Mem. Cognit.* 36, 1470–1483. doi: 10.3758/mc.336.8.1470

Mishra, J., de Villers-Sidani, E., Merzenich, M., and Gazzaley, A. (2014). Adaptive training diminishes distractibility in aging across species. *Neuron* 84, 1091–1103. doi: 10.1016/j.neuron.2014.10.034

Oswald, W. D., and Fleischmann, U. M. (1986). *Nürnberger-Alters-Inventar NAI . Testkasten und Kurzmanual*. Nürnberg: Universität Erlangen-Nürnberg.

Sala, G., and Gobet, F. (2017). Does far transfer exist? negative evidence from chess, music, and working memory training. *Curr. Directions Psychol. Sci.* 26, 515–520. doi: 10.1177/0963721417712760

Salminen, T., Frensch, P., Strobach, T., and Schubert, T. (2016). Age-specific differences of dual n-back training. *Aging Neuropsychol. Cogn.* 23, 18–39. doi: 10.1080/13825585.2015.1031723

Schubert, T., Finke, K., Redel, P., Kluckow, S., Müller, H., and Strobach, T. (2015). Video game experience and its influence on visual attention parameters: an investigation using the framework of the theory of visual attention (TVA). *Acta Psychol.* 157, 200–214. doi: 10.1016/j.actpsy.2015.03.005

Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). Is working memory training effective? *Psychol. Bull.* 138, 628–654. doi: 10.1037/a0027473

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., et al. (2016). Do "brain training" programs work? *Psychol. Sci. Publ. Interest.* 17, 103–186.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia* 47, 2015–2028.

Strobach, T., Frensch, P. A., and Schubert, T. (2012). Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychol.* 140, 13–24. doi: 10.1016/j.actpsy.2012.02.001

Strobach, T., and Huestegge, L. (2017). Evaluating the effectiveness of commercial brain game training with working-memory tasks. *J. Cogn. Enhance.* 1, 539–558. doi: 10.1007/s41465-017-0053-0

Strobach, T., and Karbach, J. (2016). *Cognitive Training: An Overview of Features and Applications*. New York, NY: Springer.

Verhaeghen, P., Marcoen, A., and Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: a meta-analytic study. *Psychol. Aging*, 7, 242–251. doi: 10.1037//0882-7974.7.2.242

von Bastian, C. C., and Oberauer, K. (2014). Effects and mechanisms of working memory training: a review. *Psychol. Res.* 78, 803–820. doi: 10.1007/s00426-013-0524-6

Wechsler, D. (1956). *Manual for the Wechsler Adult Intelligence Scale*. New York, NY: Psychological Corporation.

Wenig, W., Liang, J., Xue, J., Zhu, T., Jiang, Y., Wang, J., et al. (2019). The transfer effects of cognitive training on working memory among Chinese older adults with mild cognitive impairment: a randomized controlled trial. *Front Aging Neurosci.* 11:212. doi: 10.3389/fnagi.2019.00212

WHO (2015). *Governments Commit to Advancements in Dementia Research and Care*. Geneva: WHO.

Wild-Wall, N., Falkenstein, M., and Gajewski, P. D. (2012). Neural correlates of changes in a visual search task due to cognitive training. *Neural Plast.* 2012:529057. doi: 10.1155/2012/529057

Willis, S. L., Tennstedt, S. L., Marsiske, M., et al. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA* 296, 2805–2814. doi: 10.1001/jama.296.23.2805

Yesavage, J., and Jacob, R. (1984). Effects of relaxation and mnemonics on memory, attention and anxiety in the elderly. *Exp. Aging Res.* 10, 211–214. doi: 10.1080/03610738408258467

Conflict of Interest: MF was employed by the company Institut für Arbeiten Lernen Altern GmbH.

The remaining 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.

APPENDIX

The appendix contains a detailed description of the exercises included in the cognitive training and the schedule of the cognitive training procedures.

MAT (www.gfg-online.de) is a paper and pencil package with short exercises which had to be applied for 10 min daily to increase working memory capacity, visual attention and speed of information processing. In particular, the training aimed at enhancing psychomotor processing by faster perceiving and responding to objects or words, for example detection of triangles in a complex geometric figure or identification of words in a complex letter matrix, which were arranged either vertically, horizontally or diagonally. Memory training included exercises that used words, figures or digits. Participants were asked to memorize the items from each category and recall as many items as possible after several minutes. A more complex exercise consisted of association between faces and personal data like age and profession and recalling the information after a face presentation 10 min later.

The training begins with easy exercises to make quick effects possible. By creating more challenging instructions and by allowing less time for task performance, the level of difficulty gets enhanced gradually.

The training consists of the following modules:

Information processing speed: Time limited visual search. Different forms, numbers and letters are used. Identification of single words in randomly assembled sequences of letters. The hidden words are arranged forwards, backwards, vertically, horizontally or diagonally.

Memory span: Keep several numbers, words or pictures in memory and immediate recall of words or identifying missing words.

Basic learning speed: Memorization of faces with personal data and memorization of faces with distracting stimuli.

Mental-aktiv (www.mental-aktiv.de) is an internet based platform that offers a number of memory tasks using digits, letters, colors and figures and exercises to train speed of processing. The exercises were designed in cooperation with the authors of MAT and trained the same functions as listed above.

Sudoku is a logic-based number placement puzzle that consists from a 9×9 grid with digits so that each column, each row, and each of the nine 3×3 sub-grids contain all of the digits from 1 to 9.

Ahano peds (www.ahano.de) consists of units with different levels of difficulty. The free available program includes an eye-hand coordination task, money counting task, detection of word repetitions in a text, block taping task, memory for abstract figures etc.

Double:

There is a yellow ball and a red box presented on the screen. With one hand, the participant has to use the computer mouse in a certain way in order to put the ball into the box. With the other hand, the participant has to type the presented words as quickly as possible. This exercise trains peripheral visual attention as well as the coordination of multiple operations.

Euro Coins:

There are many different coins in a purse. The task is to assemble specific coins in order to reach a given amount. This should be done as often as possible within a specific interval. Visual perception, selective attention and mental arithmetic are trained.

Response:

Balloons float past the window of an aircraft. The task is to click as quickly as possible on the relevant balloon appearing on the left side of the window. This exercise trains selective attention and distractor inhibition.

Palpation:

At the time when a green light appears on the screen, one of five given forms is hidden behind a big picture. The participant's task is to touch the form by use of the computer mouse in order to decide which form is hidden in the current trial. To make a choice, the participant has to click on the corresponding picture. There is only one attempt in each trial. This exercise trains perception and spatial-visual memory.

Double Words:

A pool of words is given, which contains each single word twice. The task is to click on the currently relevant word by use of the computer mouse. There are five attempts in each trial to find the correct word. This exercise trains the participant's memory.

Chimpanzee test:

Nine fields are presented containing single digits for a short time. After the digit's disappearance, the participants are instructed to click on the fields in ascending order to reproduce the positions, where the respective figures were shown. Here, visual perception, short-term memory and spatial-visual memory can be trained.

Colors:

The participants have to memorize the colors of a presented picture. The task is to "repaint" the image by first clicking on a "paint pot" and then clicking on the image area. The participants receive one point for each correctly chosen color. Visual perception, short-term memory and spatial-visual memory can be trained by this exercise.

Mentaga (www.mentaga.com) consists of exercises enhancing vigilance, perceptual speed, spatial attention etc. like comparison of visual patterns, face learning, counting, vigilance and eye-hand coordination.

Figurative Thinking:

In each trial, two, almost identical pictures are presented. There are exactly three differences between the two pictures, which the participant has to detect as quickly as possible. This exercise is designed to support selective attention.

Capacity:

The task is to catch vertically falling balls with a basket as accurately and quickly as possible. To adjust the basket, the participant has to use the computer mouse. Simultaneously, as many numerical and alphabetical tasks as possible have to be performed. Spatial-visual attention, arithmetic, concentration and of multiple task performance should be improved by this task.

Concentration:

In each trial, an "E" surrounded by a certain number of dots is presented. The task is to identify every E which is surrounded exactly by three dots as quickly as possible. Concentration and visual attention are trained by this task.

Pattern Matching:

Four pictures are presented in each trial. There is always one original, two rotated versions of the original and one differing picture, which the participant has to identify by clicking on it. This exercise trains the abilities of mental rotation and visual search.

Person Memory:

This exercise aims at memorizing and recognizing names and faces. First, a sequence of faces and names is presented and the participants explicitly have to memorize the names. Then, faces are displayed with various names. The participant has to decide which name is related to a particular face. This exercise specifically trains object recognition.

Visual Acuity:

In each trial, two pictures are presented. As quickly as possible, the participant has to decide whether the two pictures are identical. Visual acuity and visual search are trained by this task.

Response Capacity:

Two objects are presented side by side. The participant has to decide whether the objects are identical. A response is required if the objects are identical. This exercise aims at improving visual search and decision time.

Memory for Numbers:

The participant has to memorize and reproduce numbers presented on the screen. The length of each number is adapted to the participant's capacity. The more digits a number contains, the more time is granted for memorizing and reproducing the number. Primarily, this exercise trains the memory for numbers, but also working memory in general.

TABLE A1 | A schedule of the cognitive training program.

Week	Session	Exercise
1	1	MAT
	2	MAT
2	3	MAT
	4	MAT
3	5	MAT
	6	MAT/Sudoku
4	7	MAT/Sudoku
	8	MAT/Sudoku
5	9	Mental-Aktiv/Ahano/Sudoku
	10	Mental-Aktiv/Ahano
6	11	Mental-Aktiv/Ahano
	12	Mentaga/Mental-Aktiv/Ahano
7	13	Mentaga/Mental-Aktiv/Ahano
	14	Mentaga/Ahano
8	15	Mentaga/Ahano
	16	Mentaga/Ahano/Sudoku
9	17	Mentaga/Ahano/Sudoku
	18	Mentaga/Sudoku
10	19	Mentaga/Sudoku
	20	Mentaga/Sudoku
11	21	Mentaga
	22	Mentaga
12	23	Mentaga/Ahano
	24	Mentaga/Ahano
13	25	Mentaga/Sudoku
	26	Mentaga/Ahano
14	27	Mentaga/Ahano
	28	Mentaga/Ahano
15	29	Mentaga/Ahano
	30	Mentaga
16	31	Mentaga/Ahano/Sudoku
	32	Mentaga/Sudoku

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 04 December 2020
doi: 10.3389/fnhum.2020.565927

Training and Transfer of Cue Updating in Older Adults Is Limited: Evidence From Behavioral and Neuronal Data

Jutta Kray[1*†], *Nicola K. Ferdinand*[2†] *and Katharina Stenger*[1]

[1] Department of Psychology, Saarland University, Saarbrücken, Germany, [2] Department of Psychology, Bergische Universität Wuppertal, Wuppertal, Germany

OPEN ACCESS

Edited by:
Susan Mérillat,
University of Zurich, Switzerland

Reviewed by:
Francisco Barceló,
University of the Balearic Islands, Spain
Christine Stelzel,
International Psychoanalytic University Berlin, Germany

***Correspondence:**
Jutta Kray
j.kray@mx.uni-saarland.de

[†] These authors have contributed equally to this work

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

Received: 26 May 2020
Accepted: 20 October 2020
Published: 04 December 2020

Citation:
Kray J, Ferdinand NK and Stenger K (2020) Training and Transfer of Cue Updating in Older Adults Is Limited: Evidence From Behavioral and Neuronal Data. Front. Hum. Neurosci. 14:565927. doi: 10.3389/fnhum.2020.565927

Cognitive control processes, such as updating task-relevant information while switching between multiple tasks, are substantially impaired in older adults. However, it has also been shown that these cognitive control processes can be improved by training interventions, e.g., by training in task switching. Here, we applied an event-related potential (ERP) approach to identify whether a cognitive training improves task-preparatory processes such as updating of relevant task goals. To do so, we applied a pretest-training-posttest design with eight training sessions. Two groups of older adults were either trained in task switching (treatment group) or in performing single tasks (control group) and we compared their performance to a group of untrained younger adults. To foster cue updating in the treatment group, we applied a cue-based switching task in which the two task cues were randomly selected prior to target presentation so that participants had time to prepare for the upcoming task. In contrast, the control group also received task cues but those were redundant as only one task had to be performed. We also examined whether training in cue updating during task switching can be transferred to a similar cognitive control task measuring updating of context information, namely a modified version of the AX-Continuous Performance Task (AX-CPT). The results revealed training-specific improvements in task switching, that is, a larger improvement in blocks requiring switching in comparison to single tasks at the behavioral level. In addition, training specific-effects were also found at the neuronal level. Older adults trained in cue updating while switching showed a reduction in mixing costs in the cue-related P3, indicating an improvement in preparatory updating processes. Additionally, P3 topography changed with training from a very broad to a parietally focused scalp distribution similar to the one found in younger adults. However, we did not obtain training-specific improvements in context updating in the AX-CPT neither at the behavioral level nor at the neuronal level. Results are discussed in the context of the ongoing debate on whether transfer of cognitive training improvements is possible.

Keywords: task-switching training, cue updating, Cue-P3, older adults, transfer

INTRODUCTION

It is a well-documented finding in cognitive aging research that a variety of cognitive domains, such as working memory, inhibition, and cognitive flexibility show a substantial decline with increasing age (for reviews, Nyberg et al., 2012; Hartshorne and Germine, 2015). This is of importance in light of a growing elderly population and a prolonged life expectancy. Hence, a key challenge for researchers in the field of cognitive aging is (a) to identify conditions under which cognitive decline may be reversed or cognitive functioning can at least be maintained, (b) to investigate which training interventions not only lead to performance improvements in the trained tasks but also to better performance in untrained tasks (Binder et al., 2015), and (c) how these processes are reflected in the brain (e.g., Brehmer et al., 2014; Dörrenbächer et al., 2020).

In the meantime, there is not only evidence that the ability to improve the level of cognitive functioning is preserved even in very old age (Lövdén et al., 2010; Kühn and Lindenberger, 2016). There is also evidence that training interventions aiming at improving cognitive control functioning in older adults lead to performance gains in untrained cognitive tasks (for a review, Kray and Dörrenbächer, 2019; for a meta-analysis, see Karbach and Verhaeghen, 2014). Cognitive control is required in situations in which we need to adapt our thoughts and actions according to internal task goals in the context of changing environments. A common component of cognitive control in many cognitive tasks is representing and maintaining task-relevant knowledge (of the actual task or context) in working memory (Braver and Barch, 2002; Braver, 2012). Moreover, as more specific components of cognitive control, Miyake and Friedman (2012) identified the updating of task-relevant information and the switching between task rules.

The primary goal of the present study was to train older adults in the implementation of cognitive control by means of a task-switching training. In particular, because it is known that with aging the ability to apply proactive, i.e., preparatory, control processes declines rapidly and leads to a shift toward relying on reactive control (cf. West and Schwarb, 2006; Paxton et al., 2008; Braver, 2012), we aimed at investigating whether an intensive task-switching training would improve early task-preparatory processes, such as updating task-relevant goals in advance, which may also result in better switching performance. Therefore, we applied EEG measurement that allows to assess fine-grained temporal processes during preparing the upcoming task.

Age Differences in Cognitive Control

To date there is a variety of empirical evidence for age differences in cognitive control as measured with the task-switching paradigm (for a review, see Kray and Ferdinand, 2014; for a meta-analysis, see Wasylyshyn et al., 2011). In this paradigm, participants are instructed to perform two (or more) different cognitive tasks A and B, for instance, to categorize pictures as belonging to the category of fruits or of vegetable in task A, and to categorize the size of the pictures as small or large in task B. Participants perform these tasks in two types of blocks, the so-called single-task blocks and mixed-task blocks. In single-task blocks, they only perform one task A or B in isolation, while in mixed-task blocks, they are instructed to switch between both tasks A and B either in a predictable or in a random order (for overviews, see Kiesel et al., 2010; Grange and Houghton, 2014). Two types of costs can be derived from this paradigm: (1) Mixing costs (also termed global or general shifting costs) that are defined as difference in performance between mixed-task and single-task blocks. Mixing costs are considered to reflect the ability to deal with the switching situation, that is, to maintain and select between two task-sets, such as in mixed-task blocks, as compared to maintain only one task in single-task blocks. (2) Switching costs (also termed local or specific shifting costs) are defined as the difference in performance between switch and repeat trials *within* mixed-task blocks. These costs are assumed to reflect the ability to disengage from a previous task rule and to shift to the other task rule.

From behavioral studies, there is quite some evidence that older adults show larger mixing costs than younger adults, suggesting age-related impairments in maintaining multiple task sets. Moreover, age differences in switching costs are substantially smaller than in mixing costs (for reviews, Kray and Ferdinand, 2014; Gajewski et al., 2018). Further empirical support for the differential age-related changes in these types of task-switching costs comes from a meta-analytic study on age differences in task switching (Wasylyshyn et al., 2011). These findings suggest that older adults' deficits are due to being in a switch situation and requiring the selection between tasks and their maintenance and not by executing the task switch itself. However, it has been shown that age differences in mixing costs can be reduced by increasing the time to prepare for the next task (Kramer et al., 1999; Kray and Lindenberger, 2000; Kray, 2006; Whitson et al., 2012), or by instructing participants to verbalize the next task prior to target presentation, that is, when using verbal cues that facilitate task preparation in the absence of external task cues (Kray et al., 2008). In contrast, the pattern of much larger mixing costs relative to switching costs has primarily been observed in cued task-switching paradigms in which the two tasks are randomly presented and participants need to update which task to prepare depending on an external cue (Kray, 2006). One explanation for this pattern of findings is that older adults, in contrast to younger adults, tend to update the task also in repeat trials (when it is not necessary) and not only in switch trials, therefore switch costs are low or sometimes even negative in comparison to mixing costs (Karayanidis et al., 2011; Whitson et al., 2012). This updating deficit is particularly found in situations of uncertainty, when the targets are ambiguous as they contain features associated with both relevant tasks (Mayr, 2001). Alternatively, it has been suggested that older adults prefer a reactive control mode during performing a task, meaning that they are less engaged in advance preparation according to cues presented and update the task (or cue information) only when the target is presented. If they do so in repeat and switch trials, responses are much slower in mixed as compared to single trials while difference between repeat and switch trials are relatively low. In contrast, young adults prefer an active control mode in which they prefer and update the task representation in advance and if needed (Braver and Barch, 2002). On the basis of these

theoretical consideration and previous empirical findings, we will focus on the comparison between single blocks and mixed blocks in the present study.

Because of their high temporal resolution, neurophysiological measures like event-related potentials (ERPs) can further contribute to determine the sources of age differences in mixing costs. In particular, ERPs allow to separate proactive processes required for task preparation—which are supposedly difficult for older adults—from those required for later task execution and response selection processes including reactive control processes (for reviews, Karayanidis and Jamadar, 2014; Gajewski et al., 2018). To investigate whether proactive control in older adults can be improved by a specific switching training, we compared two older training groups, a task-switching group that practiced to continuously update task-cue information, and therefore only performed mixed-task blocks in which the two tasks were randomly selected. In contrast, the single-task group only performed both tasks in separate blocks so that attending to the cue information was not necessary.

An ERP component that occurs during task preparation in cued task-switching paradigms and that is typically elicited by the presentation of a task cue is the cue-related P3. It is a parietal positivity that occurs about 400–600 ms after cue-onset (for a review, see Gajewski et al., 2018). Because it is usually more pronounced for switch as compared to non-switch trials (e.g., Kieffaber and Hetrick, 2005; Nicholson et al., 2005; Swainson et al., 2006; Lavric et al., 2008), it has also been labeled sustained posterior positivity (cf. Kopp et al., 2014) or (when measured as the difference wave between switch and repeat trials) switch positivity (Karayanidis et al., 2003, 2011). This parietal positivity has been linked to the idea of the classic P300/ P3b account (Donchin, 1981; Donchin and Coles, 1988; Polich, 2007), that has been assumed to reflect updating of task relevant knowledge, an idea that is corroborated by the finding that larger cue-P3s are related to smaller switch costs (e.g., Karayanidis et al., 2009, 2011; Elchlepp et al., 2012). However, it recently has been shown that the cue-related P3 and the target-related P3 are functionally distinct mechanisms, the one associated with the updating of higher-order rules (or task sets), and the other with updating of lower-level S-R rules (Barceló and Cooper, 2018a,b). Here we will focus on the updating of task rules instead of S-R rules. There is also evidence that cue-related P3 does not reflect a unitary updating process, but different subcomponents of the cue-P3 represent different aspects of updating. In particular, in cued switching paradigms an early and late positivity can be differentiated that are assumed to be associated with the intention to switch and updating the now relevant task set, respectively (cf. Karayanidis and Jamadar, 2014).

Older adults usually show less efficient task-set updating: When comparing mixed with single task blocks, older adults show a longer cue-P3 latency than younger adults (e.g., Kray et al., 2005; Eppinger et al., 2007; West and Travers, 2008). Additionally, when comparing switch and repeat trials, younger adults display larger cue-P3s on switch than repeat trials. In contrast, older adults show a reduced amplitude difference between these two trial types. This finding is interpreted as older adults needing to update task sets on each trial when they are in a switching context, no matter whether it is actually necessary or not (Eppinger et al., 2007; Friedman et al., 2008; Whitson et al., 2014). At the same time, ERPs show evidence for an additional recruitment of frontal brain regions in older adults, as can be inferred from a frontal shift in the topography of the cue-P3 (Kray et al., 2005; Eppinger et al., 2007; Karayanidis et al., 2011). This has been interpreted as a compensatory mechanism which helps older adults to keep their performance up (Friedman, 2008; Reuter-Lorenz and Cappell, 2008).

Another paradigm that can be used to measure preparatory cognitive control processes, such as updating task and response rules, is the AX-continuous performance task (AX-CPT). In this task, stimuli are presented in cue-target pairs and performance is compared across four types of cue-probe combinations. Participants are instructed to respond (e.g., with a right button press) to a specific target pair "AX" (when the target "X" is following the cue "A") that is presented in 70% of the all trials to induce a strong response bias. In AY trials, another target (e.g., M, L, K) is following the cue "A," while in BX trials the target "X" is preceded by other cues (e.g., P, T, S). Finally, in BY trials neither the cue A nor the target X is presented. Each of the three combinations (AY, BX, BY) is presented only in 10% of the trials and participants are instructed to respond with a left button press. Two types of control processes have been identified with this type of paradigm. If participants produce more AY than BX errors, they are strongly engaged in advance preparation of the response associated with the A cue (proactive control). In contrast, BX errors occur when participants are less engaged in advance preparation and press the wrong response button because they fail to correctly reactivate the preceding cue information (reactive control mode). In a number of studies, Braver and colleagues found that whereas younger adults showed strong engagement in task preparation processes, like updating and maintaining the cue (context) information (proactive control), older adults had deficits in advance preparation and instead needed to reactivate the cue information when confronted with the target (reactive control; Braver et al., 2005; Rush et al., 2006; Paxton et al., 2008; for a review, see Braver and Barch, 2002).

This general pattern has also been confirmed in ERP studies with a modified version of the AX-CPT, as the former paradigm is less suitable for ERP research given the distribution and low number of trials in critical conditions. In this version, two types of trials can be separated, context-dependent and independent trials. Similar to the cued task-switching paradigm, two cues (the context-dependent ones) induce the preparation of two response alternatives, but only one of them is actually selected after target presentation. In the other half of the cases, two different cues are redundant (context-independent) as the response selection is clearly associated with the target only so that preparation and cue updating is not needed. In younger adults, larger P3 amplitudes after cue presentation have been found in trials where the response is dependent on the preceding cue (context-dependent trials) than in trials where the response is independent on the cue (context-independent trials). This has been interpreted as younger adults ability to flexibly adapt to the more difficult conditions that need more preparatory updating of task-relevant information (Lenartowicz et al., 2010; Schmitt

et al., 2014a,b). In contrast, older adults' P3 amplitudes did not differ between conditions, suggesting that older adults needed to update context information on every trial, even when it was not required (Schmitt et al., 2014a,b), similar to the results obtained in cued-task switching. Of importance for the present study, it has been found that these preparatory updating processes can be strengthened in the elderly, e.g., by extended practice and directed strategy training (Paxton et al., 2006; Braver et al., 2009) or by the prospect of a reward (Schmitt et al., 2015, 2017).

In sum, preparing an upcoming task depending on cues and the respective neuronal correlates have been investigated with cued task-switching paradigms in which a cue either indicated a task switch or task repetition, and in the modified AX-CPT in which cues either were informative for response selection or not. Hence, in both paradigms younger adults recruit more cognitive control (as indexed by larger cue-P3 amplitudes) when needed (after a cue switch or in context-dependent trials), while older adults also invest in control when it is not needed.

Training of Task Switching

There is an ongoing debate in cognitive training research mainly about whether cognitive training gains can be transferred and generalized to untrained cognitive tasks (e.g., Novick et al., 2019). While a lot of studies have examined age differences in near and far transfer effects in the domain of working memory and multitasking (Anguera et al., 2013; Strobach et al., 2016), there is also some evidence on the effectiveness of task-switching training (for recent reviews, Karbach and Kray, 2016; Kray and Dörrenbächer, 2019). It has been shown that training switching by performing mixed-task blocks as compared with training only single tasks leads to a larger reduction of mixing and switching costs, and this reduction was even more pronounced in older than in younger adults (Karbach and Kray, 2009; Kray and Fehér, 2017). For both, younger and older adults, larger transfer effects, i.e., a larger reduction of mixing costs under a variable training in which participants received a new set of stimuli and tasks in each of the training sessions were found (cf. Karbach and Kray, 2009). In addition, older adults also showed larger transfer effects of task-switching training when the ambiguity and by this the interference between two tasks was high (Kray and Fehér, 2017). Given these previous findings, we decided to apply a variable training with ambiguous stimuli in the present study.

There are only a few studies examining different training conditions (e.g., cognitive, physical, relaxation) on changes in neuronal correlates of task switching in older adults. One study by Gajewski and Falkenstein (2012) found that the target P3 was enhanced after the cognitive training intervention compared to the control group and the physical and relaxation group but unspecifically on single, repeat and switch trials, which was interpreted as higher cognitive resources to perform the actual task. In a recent study, the effects of multi-domain cognitive training on task preparation and task execution was examined in a switching task in which participants were asked to switch between responding to the word measuring or colors of Stroop stimuli (Küper et al., 2017). Interestingly, the cognitive training intervention, as compared to the active and passive (social) control group, influenced processing in the cue processing stage. Here, they found that in the cognitive training group, the cue-locked P3 increased from pretest to posttest for repeat trials at all electrodes (Fz, Cz, and Pz), and in the active control group only at the central electrode while such an increase was not obtained in the passive control group. This finding was interpreted by the authors as better maintenance of task rules from one trial to the next under switching conditions. The effect was unspecific to the training intervention as it was also obtained in the active control group. However, the cognitive training was a multi-domain training including memory, speed, and reasoning tasks that may not be similar enough to induce transfer of training, in particular, a boost in the updating of task rules.

The Present Study

We already have some evidence from previous training studies that training in task switching is useful and effective and can lead to performance improvements in untrained similar switching tasks and other cognitive tasks in older adults. The primary question of this study was whether potential training gains mainly result from improved task preparation and whether such improvements can be transferred to other cognitive control tasks. Therefore, we applied an ERP approach to determine changes in cue processing during the anticipation of the next task with a cued-based task-switching paradigm at pre- and post-test. To examine the effects of training on cue updating, older participants were assigned to two different training groups. One group received eight practice sessions of switching by performing only mixed-task blocks in which the cue was relevant to correctly perform the next task (task-switching group). The other group, as active control group, only received single-task blocks (single-task group) during the eight practice sessions in which the cue was non-informative for the task. Note that the training was variable, meaning, that all participants performed a new set of stimuli and tasks in each of the eight training sessions and all stimuli were ambiguous to induce interference between both tasks, as we found larger transfer effects under these training conditions in two previous studies (Karbach and Kray, 2009; Kray and Fehér, 2017). To determine transfer effects, our participants also performed a modified AX-CPT task that, similar to the cued task-switching paradigm, required the updating of cue-relevant information (see Methods section). Here, ERPs were also measured to the onset of the cue. We also assessed a group of younger adults at pre- and post-test in both tasks to determine age differences in cognitive control processes.

On the behavioral level, we expected that both training groups would improve during the training sessions, that is, we should find faster reaction times and less errors with increasing practice. We also expected transfer effects, that is, a larger reduction of task-switching costs and a larger reduction in the context effect in the task-switching training than in the single-task training group. On the neuronal level, we focused on the cue-P3 in the preparation interval and expected changes in cue-P3 amplitudes, indexing engagement in updating task rules. In particular, we expected updating to become more specific to the trials where it is needed (switch trials) and by this result in a decrease of mixing and an increase or emergence of switching costs. If practice effects in updating cue-relevant

TABLE 1 | Characteristics of the sample and baseline measurement of cognitive tasks at pretest.

	Group		
	Younger adults ($n = 31$)	Single-task training ($n = 34$)	Task-switching training ($n = 30$) adults ($n = 31$)
	M (SD)	*M (SD)*	*M (SD)*
Mean age (years)	22.9 (2.74)	70.35 (4.68)	68.27 (3.97)
DSST test score	61.97 (11.2)	42.29 (8.94)	46.30 (9.70)
Spot-a-word test score	22.29 (3.5)	27.68 (3.80)	27.43 (2.78)
RT mixing costs (ms)	67 (61.0)	123 (93.0)	124 (96.1)

DSST, Digit Symbol Substitution Test; RT, Reaction Time.

information can be transferred to another task, we would also expect changes in the cue-P3 of the AX-CPT. More specifically, a difference in P3 amplitudes between context-dependent and context-independent trials should emerge after the training resulting from a more flexible investment of cognitive control in those trials where it is actually needed.

METHODS

Participants

Overall, 64 older adults (33 males) were willing to participate and complete the intensive training, while 31 younger adults (18 males) only participated in the pretest and the posttest sessions. Participants were recruited through a newspaper article and from a subject pool of Saarland University. All participants signed informed consent in advance and received a monetary compensation of €8 per hour for their attendance. The older adults additionally received a reimbursement of €20 for travel expenses. The study procedure and the written informed consent were approved by the local ethics committee at Saarland University.

According to self-reports, all participants were native German speakers, reported normal or corrected to normal vision and hearing, and none of the participants reported neurological or psychological disorders. Moreover, all participants were right-handed as measured with the Edinburgh Inventory (Oldfield, 1971). Characteristics of the final sample are shown in **Table 1**. In line with previous studies on aging and the two-component model of intelligence (Baltes et al., 1999), we found significant age differences in processing speed as measured with the Digit Symbol Substitution Test (DSST, adapted from Wechsler, 1981), that is, a typical slowing in perceptual speed of processing in older adults as compared to younger adults, $F_{(1,92)} = 64.61$, $p < 0.001$, $\eta_p^2 = 0.41$. In contrast, in a semantic knowledge test, the Spot-a-Word Test (adapted from Lehrl, 1977), older adults achieved a higher score than younger adults, $F_{(1,92)} = 51.96$, $p < 0.001$, $\eta_p^2 = 0.36$.

Importantly, in order to avoid baseline differences in cognitive measures between the two training groups, these two groups were matched according to their performance in the perceptual speed task and the magnitude of mixing costs at pretest. As can be seen in **Table 1**, the two trainings groups did not significantly differ in both measures ($p = 0.09$ and $p = 0.98$, respectively). Moreover, they also did not differ in mean age ($p = 0.06$).

Study Design and Procedure

To measure training and transfer effects, we used a pretest-training-posttest design. The pretest sessions included the assessment of cognitive functioning by means of a cognitive test battery as well as the baseline measurement of neuronal correlates of cognitive control functioning by means of EEG recordings (of about 180 min) and functional imaging (fMRI) that was measured in a separate session (of about 150 min). The pretest EEG and fMRI sessions were identical to the posttest session. Here, we will report only the results from the EEG sessions (for the fMRI results, see Dörrenbächer et al., 2019, 2020). Only the older adults performed the eight training sessions between pre- and post-test that were spaced over 4 weeks. Thus, the participants received the training intervention twice a week for about 45 min. We tested participants individually in the training sessions by one experimenter and at pretest and posttest by two experimenters. Each of the sessions will be described in detail below.

Pre- and Post-test Sessions

In the pre- and post-test sessions, each participant performed three cognitive control tasks: a switching task, a context-updating task (modified AX-CPT), and a working-memory filtering task, while the EEG was recorded. As task-preparatory processes cannot be observed in the memory-filtering task, we will focus on the results of the switching task and the context-updating task. The experimental tasks for the pre- and post-test as well as for the training sessions were programmed using E-Prime® 2.0 Professional (Psychology Software Tools, Inc, 2012).

Measurement of Task Switching

To measure task-switching performance, we applied a modified version of the task-switching paradigm as used in the training study by Karbach and Kray (2009). In this paradigm, participants are instructed to perform two categorization tasks, either in isolation (i.e., single-task blocks), or they have to switch between them (i.e., mixed-task blocks). Targets were pictures that had to be categorized according to semantic meaning either as fruit or vegetable (task A), or according to a perceptual feature as small or large in size (task B) by pressing one of two buttons on a response pad. Hence, the stimulus-response mappings of both tasks were overlapping because one feature of each of the two tasks was mapped onto the same response key. Participants were instructed to use their left and right index fingers for responding as well as to respond as quickly and accurately as possible. In contrast to the previous study, participants received cues indicating the next task A or B that were presented in a random order. Cues consisted of two letters indicating the food task (ES = "Essensaufgabe") or the format task (FO = "Formataufgabe") in advance of the target stimulus.

Target stimuli were 32 colored pictures of food items (16 fruits, 16 vegetables) adapted from the Snodgrass and Vanderwarts'

pictorial set (Rossion and Pourtois, 2004). All targets were presented in a pseudorandom order in either small size (90 × 90 pixels) or large size (220 × 220 pixels) at the center of the computer screen.

In the practice phase, participants first performed two single-task blocks of 12 trials and two mixed-task blocks of 12 trials. Thereafter, they performed eight experimental blocks (four single-task blocks and four mixed-task blocks). Each block consisted of 41 trials while the first trial as a re-start trial was always excluded from data analyses. The mixed-task blocks consisted of 20 repeat and 20 switch trials presented in a random sequence. Each block consisted of an equal number of stimulus and response types. Performance feedback of mean reaction times and error rates was given at the end of each block. Stimulus-response assignments were counterbalanced across participants as well as the order of single-task blocks. Testing time lasted about 25 min.

The trial procedure was identical for single and mixed trials. Each trial started with a 300 ms fixation cross, followed by the cue (i.e., ES or FO), which was presented for 800 ms. After the cue, a second fixation cross was presented for 1,000 ms, followed by the target that was displayed for 1,800 ms. Responses had to be executed within this given time window. Otherwise, the trial was excluded from further analyses. The inter-trial interval (ITI) lasted for 500 ms.

Measurement of Context Updating

To measure context updating, we applied a modified AX-CPT (Lenartowicz et al., 2010), that was further adapted to examine age differences in neuronal correlates of context updating in studies from our lab (e.g., Schmitt et al., 2014a).

In this version of the AX-CPT, participants were instructed to respond to four different cue-target combinations by pressing either the left or right response button (i.e., if-then rules). The task consisted of two trial types. In context-dependent trials (c-dep), correct responses to targets were depended on the preceding cue. For instance, if the female person was followed by a bird, then participants should press the left key, and if the male person was followed by a cat, then should press the right key. In context-independent (c-indep) trials, responses to the targets did not depend on the preceding cue. Nevertheless, the same type of instructions was given. For instance, if the female person was followed by a fish, then participants should press the left key, and if the male person was followed by the fish, then participants should press the left key. In context-independent trials, the cue can be neglected in order to select the correct response (the fish is always assigned to the left response). Participants were further instructed to use the left and right index fingers for responding and to respond as quickly and accurately as possible.

As target stimuli, we used four colored animal pictures adapted from the Snodgrass and Vanderwarts' pictorial set (Rossion and Pourtois, 2004). Cues were four photographs of human faces (young/old woman, young/old man) from the lifespan database of adult facial stimuli (Minear and Park, 2004) as this task has already been used in a previous aging study (Schmitt et al., 2014a).

In the practice phase, participants performed three blocks of 16 trials each in the following fixed order. The first practice block contained only c-indep trials, followed by a second block of only c-dep trials, which was followed by a third block containing intermixed trials. In the experimental phase, participants performed four intermixed blocks, each containing 41 trials: one start trial, 20 c-indep and 20 c-dep trials. Each block was equal regarding stimulus and response type. Stimulus-response assignments were counterbalanced across participants. The same tasks were used at pre- and post-test, but with different stimulus sets. The testing time was ∼15 min.

The trial procedure started with a 250 ms fixation cross, followed by the informative cue, which stayed on the display for 750 ms. After a second 750 ms fixation cross, the target was presented for 3,600 ms and responses had to be executed within the given time window. If participants did not respond within the given time window, the trial was excluded from further analyses. The 500 ms ITI separated two consecutive trials.

Training Sessions

The training intervention consisted of eight sessions. Older adults practiced over a time period of 4 weeks with two 45-min sessions per week at Saarland University with the restriction to not train on two consecutive days per week. According to their baseline performance (see above), older participants were assigned to either the single-task training group or the task-switching training group. Both groups received and performed identical tasks (same cues and stimuli) during the training sessions, with the difference that the single-task training group practiced the two tasks (e.g., A and B) in separate blocks, while the task-switching training group only received mixed-task blocks, and by this, practiced the updating of cue information while switching between two tasks.

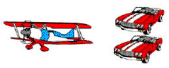
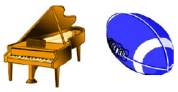
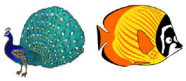
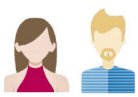
Each training session consisted of 10 blocks of 41 trials with the block design and trial procedure being identical between the pre- and post-test sessions. In each session, the single-task training group performed five blocks of the one task and five blocks of the other task, whereas the task-switching training group performed 10 mixed-task blocks.

Across the eight training sessions, we used different stimulus material and different tasks given that a previous training study has shown that a variable training lead to greater transfer effects in older adults (Karbach and Kray, 2009). The eight different training tasks were constructed in a way that participants were required to classify pictures according to a semantic or perceptual task (see **Table 2**). The first four training tasks were adopted from Karbach and Kray (2009) and the other four training tasks were newly created. Stimuli were taken from the databases of Snodgrass and Vanderwart (1980) and Rossion and Pourtois (2004). The order of the eight training tasks was kept constant across the sessions.

EEG Recording and Pre-processing

During EEG recording, participants were seated in an electrically shielded and noise-protected room. EEG was recorded using the Brain Vision Recorder Software (Brain Products, Munich, Germany). Fifty-nine Ag/AgCl active Electrodes were attached to

TABLE 2 | Semantic and perceptual tasks used in the eight training sessions.

Training session	Semantic task (task A)	Categories (task A)	Perceptual task (task B)	Categories (task B)	Example stimuli
Session 1	Transportation	Planes/cars	Number	One/two	
Session 2	Hobby	Music/sports	Color	Blue/orange	
Session 3	Animal	Bird/fish	Direction	Left/right	
Session 4	Plant	Leaf/flower	Chromaticity	Colored/black-and-white	
Session 5	Clothing	Hat/shoe	Texture	Dotted/squared	
Session 6	Landscape	Building/tree	Orientation	Upright/rotated	
Session 7	Gadget	Toy/tool	Luminance	Bright/dark	
Session 8	Gender	Male/female	Hair-color	Blond/brown	

As in the pretest and posttest sessions, the cues for each the task A and task B referred to the first two letters of the German translation of each task: transformation task (TR = "Transportmittel"; number task (ZA = "Zahl"); hobby task (HO = "Hobby"); color task (FA = "Farbe"); animal task (TI = "Tier"); direction task (RI = "Richtung"); plant task (PF = "Pflanze"); chromaticity task (SÄ = "Sättigung"); clothing task (KL = "Kleidung"); texture task (MU = "Muster"); landscape task (TE = "Terrain"); orientation task (OR = "Orientierung"); gadget task (OB = "Object"); luminance task (LU = "Luminanz"); gender task (GE = "Geschlecht"); hair-color task (HA = "Haarfarbe").

elastic caps (Acticap) and arranged in the extended international 10–20 system (Jasper, 1958). Additionally, electro-ocular activity (EOG) was recorded from four electrodes placed at the outer canthi of both eyes and above and below the right eye. The ground electrode was placed at AFz, the online reference at the right mastoid, and impedances were kept below 20 kΩ. EEG data were filtered online with a low-pass filter of 250 Hz and converted analog-to-digital with a sampling rate of 500 Hz. Offline EEG processing was done using EEProbe (ANT). Recordings were band-pass filtered offline from 0.01 to 30 Hz, re-referenced to linked mastoids, and averaged according to the respective experimental conditions. Whenever the standard deviation in a moving 200 ms time interval exceeded 30 μV in ocular electrodes, data were marked as artifacts. These eye movements were corrected by using a linear regression approach (Gratton et al., 1983). All other artifacts in the EEG data were rejected prior to averaging if the standard deviation of the amplitude in a 200 ms interval was above 20 μV in the representative electrode Cz. For the task-switching task, this procedure resulted in the rejection of 43.2 and 46.2% of trials for younger adults, 58.7 and 59.5% for older adults from the single-task training group, and 62.2 and 55.8% for older adults from the task-switching training group, for pre- and post-test, respectively. For the context-updating task, this procedure resulted in the rejection of 33.2 and 31.9% of trials for younger adults, 55.3 and 55.0% for older adults from the single-task training group, and 55.5 and 49.6% for older adults from the task-switching training group, for pre- and post-test, respectively. The rejection rates were similarly distributed for single and mixed trials at pretest and posttest for younger and older adults, and therefore did probably not selectively influence the EEG results. Overall, we made sure that for each condition in each subject, we at least obtained 16 artifact-free trials. Participants that did not fulfill this criterion were excluded from data analyses (see next section).

Data Analysis

Before analyzing, behavioral performance was screened for extreme values in the baseline and the training data. Mean reaction times (RT) or error rates more than three standard deviations from the corresponding group mean were considered as extreme values which resulted in the exclusion of one older participant from the switching group from all behavioral and ERP

analyses. Four participants (two younger adults, one older adult from the single-task training group, and one from the switching group) producing <16 artifact-free trials per task condition were excluded from statistical analyses of the task-switching EEG data. Three participants (two older adults from the single-task training group and one from the switching group) were excluded from the analyses of the AX-CPT EEG data. Data were analyzed with the software package SPSS 24.

For behavioral data analysis of the switching task, the first trial in each block and trials including reaction times (RT) below 100 ms were excluded from the subsequent analyses. Data exclusion was only very minor for the training data (0.16% of the trials in the task-switching group; 0.11% of the trials for the single task group). At pretest, data exclusion was 2.75% for the older group and 0.28% for the younger group, and at posttest, 0.68% for the older group and 0.19% for the younger group). Behavioral results were based on mean RT for correct responses and on error rates.

ERPs were averaged stimulus-locked to the cue and separately for each trial type in time windows from −200 to 1,000 ms (in the switching task) and from −200 to 800 ms (in the AX-CPT) using a 100 ms prestimulus baseline. For both tasks, the cue-locked P3 was analyzed using a mean amplitude measure in a respective time window. The selection of the time windows for statistical analyses was based on the literature and our previous analyses of the cue-locked P3 in these two tasks (e.g., Karayanidis et al., 2003; Kray et al., 2005; Schmitt et al., 2014a), together with the visual inspection of the peaks obtained in the present data. Visual inspection of the cue-locked P3 in the switching task indicated a prolonged P3 component without a clear peak in older adults which is consistent with the literature (for a review, see Gajewski et al., 2018). For this reason, the cue-locked P3 amplitudes were measured in an early 300–500 ms window and a late 500–700 ms window after cue onset and we refrained from analyzing P3 latencies. For the AX-CPT, ERP amplitudes were averaged in different time windows for pretest and posttest due to clear temporal shifts of the peak amplitudes from pretest to posttest in the present data: Cue-locked P3 amplitudes were averaged in a 470–670 ms window at pretest and in a 400–600 ms window at posttest. In line with previous ERP studies, analyses were restricted to the midline electrodes Fz, Cz, and Pz (e.g., Karayanidis et al., 2011). Please note that the trials included in the ERP analyses did not fully correspond to the trials that entered the behavioral analyses.

For all analyses, the alpha level was set to 0.05. Greenhouse–Geisser corrections for non-sphericity (Keselman and Rogan, 1980) were applied when necessary. In this case, epsilon corrected p-values are reported together with epsilon values and uncorrected degrees of freedom.

RESULTS

In the following, we will first report the results of the training data to make sure that participants showed improved task performance during the eight training sessions. Then, we will first analyze age-related differences in the two cognitive control tasks at the behavioral and neuronal level in order to prove whether previous findings on reported age differences can be replicated. Thereafter, we will present the results on the transfer of training to a structural similar switching task at the behavioral and neuronal level. Finally, we will report the results on the transfer to an untrained cognitive control task, also requiring the updating of cue information, as the trained task.

TABLE 3 | Means (M) and standard deviations (SD) for RTs (ms) and error rates (%) as a function of training group (single task training, task-switching training) and time (bin 1, bin 2, bin 3, bin 4).

	Time							
	Bin 1		Bin 2		Bin 3		Bin 4	
Training group	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
RTs (in ms)								
Single task training	651	78	623	72	611	70	601	69
Task-switching training	735	137	697	134	661	123	647	111
Error rates (in %)								
Single task training	2.11	1.04	1.65	1.10	1.60	1.05	1.64	1.07
Task-switching training	6.14	4.10	4.25	2.79	3.18	2.25	2.82	2.24

Training Data

Given that we applied a variable training intervention in which participants performed a new set of tasks in each of the training session, training data will not be analyzed as a function of training session because the eight tasks also differed in task difficulty. Information about training gains in each of the eight sessions is provided in the **Supplementary Material**. Instead, we determined the training gains within each of the eight training sessions by dividing them into four bins (one bin contains 35 trials) that were then aggregated across the eight sessions (see also Pereg et al., 2013). The corresponding data are shown in **Table 3**. Mean RTs and error rates were separately analyzed with an ANOVA design, including the between-subjects factor Training Group (task-switching training, single-task training) and the within-subjects factor Bin (1, 2, 3, 4).

The results on mean RTs revealed main effects for Bin, $F_{(3,183)} = 121.82$, $p < 0.001$, $\eta_p^2 = 0.67$, and Training Group, $F_{(1,61)} = 6.37$, $p < 0.05$, $\eta_p^2 = 0.10$, as well as a significant interaction between Bin and Training Group, $F_{(3,183)} = 11.54$, $p < 0.001$, $\eta_p^2 = 0.16$. Separate analyses for each training group revealed a linear reduction of mean RTs within training sessions for the task-switching training group, $F_{(1,28)} = 74.59$, $p < 0.001$, $\eta_p^2 = 0.73$, as well as for the single-task training group, $F_{(1,33)} = 78.54$, $p < 0.001$, $\eta_p^2 = 0.70$. That means, both training groups became continuously faster with increasing practice, and this decrease was larger in the task-switching group than in the single-task training group, while the single-task training group was generally faster in responding (see **Table 3**).

The results on error rates showed the same pattern. The ANOVA results revealed main effects for Bin, $F_{(3,183)} = 38.77$, $p < 0.001$, $\eta_p^2 = 0.39$, and Training Group, $F_{(1,61)} = 23.27$, $p < 0.05$, $\eta_p^2 = 0.28$, as well as an interaction between both

factors, $F_{(3,183)} = 21.19$, $p < 0.001$ $\eta_p^2 = 0.26$. Again, separate analyses for each group indicated a linear reduction of error rates with increasing practice for the task-switching training group, $F_{(1,28)} = 38.48$, $p < 0.001$, $\eta_p^2 = 0.58$, that was only marginally significant in the single-task training group, $F_{(1,33)} = 3.84$, $p = 0.059$, $\eta_p^2 = 0.10$. In line with the results on mean RTs, both groups became more accurate in responding, while practice effects were larger in the task-switching group than the single-task group, whereas the latter group also produced generally less errors (see **Table 3**).

Analysis of Pretest Data

Before analyzing the transfer effects, we looked at the pretest data in order to make sure that we were able to replicate previous results on age differences in the two cognitive control measures, and to examine potential baseline differences between the two training groups. We will report the behavioral and neuronal results first for the switching task, and then for the context-updating task (AX-CPT).

Results of the Switching Task at Pretest

To examine age and possible a priori training group differences in task switching, we performed an ANOVA with the between-subjects factor Group (older/single task training, older/task-switching training, younger adults) and the within-subjects factor Trial Type (single, repeat, switch) for the behavioral data that are shown in **Table 4**. As we were mainly interested in whether younger adults differed from older ones, and whether the older adults training group differed from each other, we conducted an orthogonal group contrasts in the ANOVA design (−2 1 1; 0 −1 1) and focused on interactions with two trial-type contrasts: The first contrast reflects mixing costs (Trial Type Contrast 1: −2 1 1) and the second contrast reflects switching costs (Trial Type Contrast 2: 0 −1 1).

Results for the mean RTs showed reliable mixing costs, $F_{(1,93)} = 153.76$, $p < 0.001$, $\eta_p^2 = 0.63$, and switching costs, $F_{(1,93)} = 27.07$, $p < 0.001$, $\eta_p^2 = 0.23$. Older adults showed larger mixing costs compared to younger adults, $F_{(1,91)} = 4.88$, $p < 0.05$, $\eta_p^2 = 0.05$, but no larger switching costs, $p = 0.64$. Importantly, the two training groups did not significantly differ in the magnitude of mixing and switching costs ($p = 0.91$, $p = 0.59$, respectively).

The analysis of error rates also revealed reliable mixing costs, $F_{(1,93)} = 48.80$, $p < 0.001$, $\eta_p^2 = 0.34$, and switching costs, $F_{(1,93)} = 14.89$, $p < 0.001$, $\eta_p^2 = 0.14$. For the error rates, older adults showed larger mixing and switching costs than younger adults, $F_{(1,91)} = 7.47$, $p < 0.01$, $\eta_p^2 = 0.08$, $F_{(1,91)} = 9.75$, $p < 0.01$, $\eta_p^2 = 0.10$, respectively. Again, the two training groups did not differ in mixing and switching costs at pretest ($p = 0.41$, $p = 0.82$, respectively).

To examine group differences in early and late cue-locked P3 mean amplitudes in task switching (see **Figure 1**), we performed ANOVAs with the between-subjects factor Group (older/single, older/switching, younger adults) and the within-subjects factors Trial Type (single, repeat, switch) and Electrode Location (Fz, Cz, Pz), using the same contrasts as before. In addition, to reduce unnecessary comparisons to those of interest, the factor Electrode Location was tested in a repeated contrast (−1 1 0, 0 −1 1).

In early cue-locked P3 amplitudes, this ANOVA yielded a significant Trial Type Contrast 1, $F_{(1,87)} = 79.65$, $p < 0.01$, $\eta_p^2 = 0.48$, indicating reliable mixing costs, as well as an interaction between Group and Trial Type Contrast 1, $F_{(2,87)} = 4.20$, $p < 0.05$, $\eta_p^2 = 0.09$. *Post-hoc* tests showed that mixing costs were significant in all three groups [younger adults: $F_{(1,28)} = 60.38$, $p < 0.01$, $\eta_p^2 = 0.68$; older/single: $F_{(1,32)} = 10.51$, $p < 0.01$, $\eta_p^2 = 0.25$; older/switching: $F_{(1,27)} = 27.41$, $p < 0.01$, $\eta_p^2 = 0.50$]. However, they were larger for younger adults than for the two older training groups, *mean difference* = 1.19, $SE = 0.35$, $p < 0.01$, while the two older training groups did not differ, $p = 0.83$. Additionally, the three-way interaction between Group, Trial Type Contrast 1, and Electrode Location (Cz vs. Pz) was marginally significant, $F_{(1,87)} = 2.84$, $p = 0.06$, $\eta_p^2 = 0.06$. Dissolving this interaction showed that for younger adults, mixing costs in early cue-locked P3 amplitudes differed between electrode locations, $F_{(1,28)} = 5.38$, $p < 0.05$, $\eta_p^2 = 0.16$. They were more pronounced at Pz, $F_{(1,28)} = 95.30$, $p < 0.01$, $\eta_p^2 = 0.77$, than at Cz, $F_{(1,28)} = 29.78$, $p < 0.01$, $\eta_p^2 = 0.52$, which can be inferred from the effect sizes. This interaction was not observed for older adults (all $p > 0.31$).

The same ANOVA was calculated for the late cue-locked P3 amplitudes. This analysis again yielded a significant Trial Type Contrast 1, $F_{(1,87)} = 38.92$, $p < 0.01$, $\eta_p^2 = 0.31$, suggesting reliable mixing costs. Mixing costs differed over electrode locations (Cz vs. Pz), $F_{(1,87)} = 10.46$, $p < 0.01$, $\eta_p^2 = 0.27$, being larger at parietal, $F_{(1,87)} = 50.95$, $p < 0.01$, $\eta_p^2 = 0.37$, than central electrodes, $F_{(1,87)} = 20.43$, $p < 0.01$, $\eta_p^2 = 0.19$, which can be inferred from the effect sizes. There was also an interaction between Trial Type Contrast 2 and Electrode Location (Cz vs. Pz), $F_{(1,87)} = 10.30$, $p < 0.05$, $\eta_p^2 = 0.11$, which was due to significant switching costs at Pz, $F_{(1,87)} = 50.95$, $p < 0.01$, $\eta_p^2 = 0.37$, but not Cz, $p = 0.92$. Although mixing and switching costs did not interact with group, we had hypotheses concerning group differences and thus additionally analyzed each group separately. These analyses demonstrated that for younger adults, the P3 was parietally distributed, $F_{(1,28)} = 9.83$, $p < 0.05$, $\eta_p^2 = 0.26$, and mixing, $F_{(1,28)} = 16.14$, $p < 0.01$, $\eta_p^2 = 0.37$, as well as switching costs, $F_{(1,28)} = 5.38$, $p < 0.05$, $\eta_p^2 = 0.16$, were present at Pz only. In contrast, for older adults mixing costs were still present in this later time interval [older/single: $F_{(1,32)} = 14.79$, $p < 0.01$, $\eta_p^2 = 0.32$; older/switching: $F_{(1,27)} = 33.19$, $p < 0.01$, $\eta_p^2 = 0.55$] and did not differ between the two groups, $p = 0.35$. Moreover, switching costs were absent (all $p > 0.49$) and there were no effects involving Electrode Location for the older age groups, all $p > 0.16$.

In sum, the behavioral data showed no baseline difference in mixing and switching costs between the two training groups. Age differences were found for mixing costs but not for switching costs at the level of reaction times as well as for mixing and switching costs at the level of errors. P3 amplitudes demonstrated the typical age-related frontal shift in P3 distribution with a clear parietal focus in younger and more evenly distributed P3 in older adults. In the early cue-locked P3, mixing costs were found in

TABLE 4 | Means (M) and standard deviations (SD) for RTs (ms) and error rates (%) as a function of group and trial type separately for the pretest and posttest.

	Younger adults/no training		Older adults/single-task training		Older adults/task-switching training	
Trial type	***M***	***SD***	***M***	***SD***	***M***	***SD***
RTs (in ms) at pretest						
Single	543	90	776	92	743	108
Repeat	611	115	888	132	854	157
Switch	641	122	909	153	882	170
RTs (in ms) at posttest						
Single	506	81	693	95	659	94
Repeat	551	117	782	140	714	138
Switch	557	129	817	159	738	166
Error rates (in %) at pretest						
Single	2.60	1.61	6.47	4.48	6.96	5.76
Repeat	4.74	3.68	11.38	9.56	10.46	10.78
Switch	4.54	3.51	14.17	7.90	13.51	7.90
Error rates (in %) at posttest						
Single	2.92	2.33	4.46	2.57	4.49	2.43
Repeat	6.31	1.78	11.47	6.17	9.33	5.35
Switch	4.04	3.01	10.12	5.75	7.29	4.51

all three groups. They were larger for younger adults and the two older groups did not differ from each other. In the later cue-locked interval, switching costs emerged for younger adults, while for older adults mixing costs extended into this interval and switching costs were not present.

Results of the Context-Updating Task at Pretest

To examine age and group differences in context updating, we performed an ANOVA with the between-subjects factor Group (older/single task training, older/task-switching training, younger adults) and the within-subjects factor Trial Type (c-dep, c-indep), using the same group contrasts as before. The behavioral data are shown in **Table 5**.

Results indicated a reliable context effect for reaction times, $F_{(1,93)} = 129.94$, $p < 0.001$, $\eta_p^2 = 0.58$, and for error rates, $F_{(1,93)} = 45.73$, $p < 0.001$, $\eta_p^2 = 0.33$, that is, better performance in c-indep than c-dep trials. For both measures, older adults showed a greater context effect compared to younger adults, $F_{(1,91)} = 20.85$, $p < 0.001$, $\eta_p^2 = 0.19$ and $F_{(1,91)} = 11.13$, $p = 0.001$, $\eta_p^2 = 0.11$, respectively, but again the two training groups did not significantly differ from each other ($p = 0.60$, $p = 0.92$, respectively).

For the analysis of the cue-locked P3 amplitudes (see **Figure 2**) in the context updating task, we applied the same ANOVA design as for the behavioral data with the additional factor Electrode Location (Fz, Cz, Pz). Again, to reduce unnecessary comparisons to those of interest, the factor Electrode Location was tested in a repeated contrast ($-1\ 1\ 0, 0\ -1\ 1$).

Cue-locked P3 amplitudes displayed an interaction between Group and Electrode [Fz/Cz: $F_{(2,88)} = 54.70$, $p < 0.01$, $\eta_p^2 = 0.55$; Cz/Pz: $F_{(2,88)} = 5.73$, $p < 0.01$, $\eta_p^2 = 0.12$], which was due to a clear parietal topography of the P3 for younger adults [Fz < Cz: $F_{(1,30)} = 84.95$, $p < 0.01$, $\eta_p^2 = 0.74$; Cz < Pz: $F_{(1,30)} = 17.84$, $p < 0.01$, $\eta_p^2 = 0.37$], but a frontal topography for both groups of older adults [Fz > Cz for older/single, $F_{(1,31)} = 11.59$, $p < 0.01$, $\eta_p^2 = 0.27$, and older/switching, $F_{(1,27)} = 41.12$, $p < 0.01$, $\eta_p^2 = 0.60$, respectively]. Additionally, the factor Trial Type showed interactions with Group, $F_{(2,88)} = 3.22$, $p < 0.05$, $\eta_p^2 = 0.07$, and with Electrode (Cz/Pz), $F_{(1,88)} = 5.23$, $p < 0.05$, $\eta_p^2 = 0.06$. In younger adults, a clear context effect emerged with P3 amplitudes for c-indep trials being smaller than those for dependent trials, $F_{(1,30)} = 8.80$, $p < 0.01$, $\eta_p^2 = 0.23$. Context effects were not present for older adults from the single-task group, $p = 0.10$. Older adults from the switching group displayed a marginally significant interaction between Trial Type and Electrode Location (Cz/Pz), $F_{(1,27)} = 3.34$, $p = 0.09$, $\eta_p^2 = 0.10$, but follow-up test at Cz and Pz did not result in any significant context effects, all $p > 0.37$.

In sum, we found a larger context effect in older than in younger adults for reaction times and error rates while both training groups did not differ in the magnitude of the context effect at pretest. The cue-locked P3 displayed a clear parietal topography in younger adults and was shifted toward frontal electrodes in both groups of older adults. Moreover, there was a P3 context effect present in younger adults, but absent for both groups of older adults.

Transfer of Training to an Untrained Switching Task

Transfer at the Behavioral Level

To examine the transfer of training in task switching to a similar switching task, we focused the analysis on reaction times, as previous studies usually did not find transfer at the level of error rates, mostly because error rates are relatively small and there is less room for improvement. To determine relative improvements from pretest to posttest (taken into account age-group differences

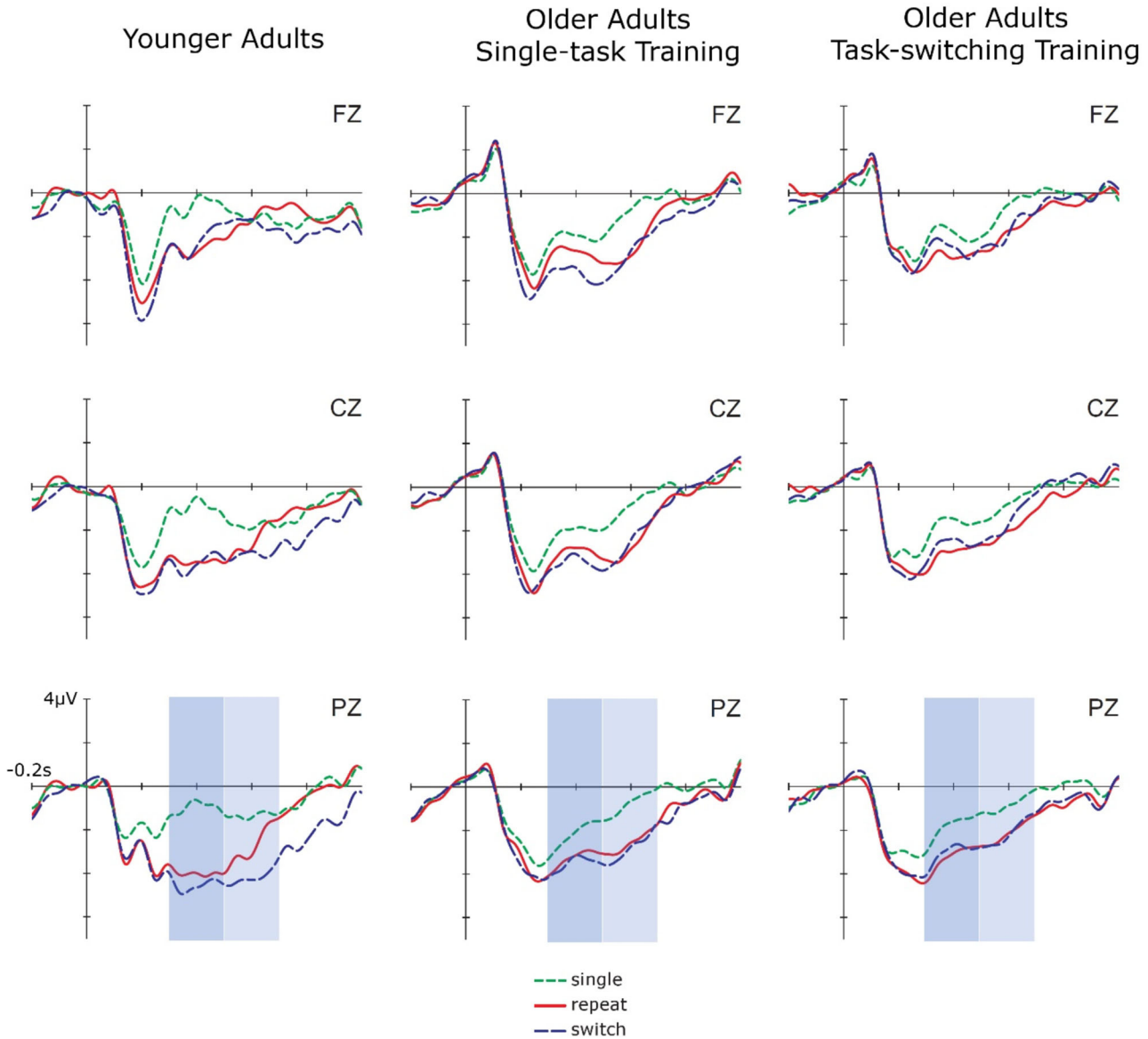


FIGURE 1 | Cue-locked ERPs for single, repeat, and switch trials at pretest for younger adults, older adults in the single-task training, and older adults in the task-switching training group displayed at electrodes Fz, Cz, and Pz.

at pretest), we also analyzed log-transformed RTs. Note that the difference between log-transformed variables corresponds to a proportional score.

We calculated an ANOVA with the within-subjects factors Session (pretest, posttest) and Trial Type (single, repeat, switch) and the between-subjects factor Group (older/single task training, older/task-switching training, younger adults). As can be seen in **Figure 3**, there was a relative reduction in mixing costs from pretest to posttest, $F_{(1,93)} = 11.34$, $p < 0.001$, $\eta_p^2 = 0.11$, with the task-switching group showing a larger improvement than the single task training group from pretest to posttest, $F_{(1,91)} = 4.04$, $p < 0.05$, $\eta_p^2 = 0.04$. Moreover, the difference in mixing costs between both groups was significant at posttest, $F_{(1,91)} = 4.51$, $p < 0.05$, $\eta_p^2 = 0.05$, while the magnitude of mixing costs between the older task-switching training group and the younger control group was not significant ($p = 0.39$). We also obtained group differences in relative improvements of switching costs namely a larger reduction of switching costs in the younger control group than in the two training groups, $F_{(1,91)} = 9.30$, $p < 0.01$, $\eta_p^2 = 0.09$.

Transfer at the Neuronal Level

To examine near transfer to an untrained task-switching task in early and late cue-locked P3 mean amplitudes (see **Figure 4**), we performed ANOVAs with the between-subjects factor Group (older/single task training, older/task-switching

TABLE 5 | Means (M) and standard deviations (SD) for RTs (ms) and error rates (%) as a function of group and trial type separately for the pretest and posttest in the AX-CPT.

	Younger adults/no training		Older adults/single task training		Older adults/task-switching training	
Trial type	***M***	***SD***	***M***	***SD***	***M***	***SD***
RTs (in ms) at pretest						
c-indep	477	66	717	173	655	125
c-dep	572	124	952	159	911	222
RTs (in ms) at posttest						
c-indep	429	66	614	118	555	106
c-dep	498	144	801	229	702	188
Error rates (in %) at pretest						
c-indep	1.23	1.50	4.92	11.74	1.33	2.25
c-dep	3.77	3.00	15.06	12.36	11.73	12.40
Error rates (in %) at posttest						
c-indep	0.76	1.23	2.72	5.98	0.84	1.75
c-dep	3.64	2.25	11.77	14.08	6.36	7.09

training, younger adults) and the within-subjects factors Session (pretest, posttest), Trial Type (single, repeat, switch), and Electrode Location (Fz, Cz, Pz), using the same contrasts as before (see Results of the Switching Task at Pretest) and focusing on effects including the factor Session. For the early cue-locked P3 amplitude, this ANOVA yielded a significant four-way interaction between Group, Session, Trial Type Contrast 1, and Electrode Location (Cz/Pz), $F_{(1,87)} = 3.99$, $p < 0.05$, $\eta_p^2 = 0.08$. For younger adults, there was a tendency for mixing costs to change over sessions, $F_{(1,28)} = 3.86$, $p = 0.06$, $\eta_p^2 = 0.12$, with mixing costs being larger at pretest, $F_{(1,28)} = 60.38$, $p < 0.01$, $\eta_p^2 = 0.68$, than at posttest, $F_{(1,28)} = 16.21$, $p < 0.01$, $\eta_p^2 = 0.37$, as can be inferred from the effect sizes. Moreover, for younger adults the topography (Fz/Cz) of the P3 changed over sessions, $F_{(1,28)} = 7.76$, $p < 0.01$, $\eta_p^2 = 0.22$. While there was no difference between frontal and central electrodes at pretest, $p = 0.21$, there was a clear difference at posttest, $F_{(1,28)} = 12.49$, $p < 0.01$, $\eta_p^2 = 0.31$, indicating less frontal involvement. For older adults from the single-task group, there was also a tendency for mixing costs to change from pretest to posttest, $F_{(1,38)} = 3.07$, $p = 0.09$, $\eta_p^2 = 0.09$, but with mixing costs being smaller at pretest, $F_{(1,32)} = 10.51$, $p < 0.01$, $\eta_p^2 = 0.25$, than at posttest, $F_{(1,32)} = 23.27$, $p < 0.01$, $\eta_p^2 = 0.42$ (see effect sizes). For older adults from the switching group, mixing costs marginally changed with Session and Electrode (Cz/Pz), $F_{(1,27)} = 3.78$, $p = 0.06$, $\eta_p^2 = 0.12$. While mixing costs were present at Cz, $F_{(1,27)} = 20.61$, $p < 0.01$, $\eta_p^2 = 0.43$, and Pz, $F_{(1,27)} = 23.25$, $p < 0.01$, $\eta_p^2 = 0.46$, at pretest, they were only present at Pz at posttest, $F_{(1,27)} = 9.79$, $p < 0.01$, $\eta_p^2 = 0.27$(Cz: $p = 0.66$). Moreover, from the effect sizes it can be inferred that mixing costs were smaller at posttest as compared to pretest (see **Figure 5**).

For the late cue-locked P3 amplitude, this ANOVA resulted in a marginally significant four-way interaction between Group, Session, Trial Type Contrast 1, and Electrode Location (Cz/Pz), $F_{(2,87)} = 3.01$, $p = 0.05$, $\eta_p^2 = 0.07$, and a significant interaction between Session and Trial Type Contrast 2, $F_{(1,87)} = 4.32$, $p < 0.05$, $\eta_p^2 = 0.05$. For younger adults and for older adults from the single-task group, there was no significant interaction including Session and Trial Type Contrast 1, all $p > 0.23$. In contrast, for older adults from the switching group, there was an interaction between Session and Trial Type Contrast 1, $F_{(1,27)} = 5.15$, $p < 0.05$, $\eta_p^2 = 0.16$, and a marginally significant interaction between Session, Trial Type Contrast 1, and Electrode (Cz/Pz), $F_{(1,27)} = 3.16$, $p = 0.09$, $\eta_p^2 = 0.11$. In this group, mixing costs were present at Cz, $F_{(1,27)} = 21.47$, $p < 0.01$, $\eta_p^2 = 0.44$, and Pz, $F_{(1,27)} = 17.59$, $p < 0.01$, $\eta_p^2 = 0.39$, at pretest, but only at Pz, $F_{(1,27)} = 7.65$, $p < 0.05$, $\eta_p^2 = 0.22$, at posttest (Cz: $p = 0.89$). Again, as for the early cue-locked P3, effect sizes indicate that mixing costs were smaller at posttest as compared to pretest. As for switching costs, they were not significant at pretest ($p = 0.55$), but reliable at posttest, $F_{(1,87)} = 14.90$, $p < 0.01$, $\eta_p^2 = 0.15$.

In sum, older adults from the switching group showed a reduction in behavioral mixing costs to the level of (untrained) younger adults. Improved behavioral switching costs were also found for younger adults at posttest. In cue-locked P3 amplitudes, a reduction in mixing costs was found for younger adults and for older adults from the switching group. Additionally, the switching group also showed a parietally focused P3 topography after training, i.e., it became similar to the younger adults' P3. For all groups, switching costs emerged post-training.

Transfer of Training to an Untrained Context Updating Task (AX-CPT)

Transfer at the Behavioral Level

To examine the transfer of training in task switching to the untrained AX-CPT, we also focused the analysis on reaction times and we controlled for relative improvements from pretest to posttest by also analyzing log-transformed RTs.

The ANOVA included the within-subjects factors Session (pretest, posttest) and Trial Type (c-indep, c-dep) and the between-subjects factor Group (older/single task training,

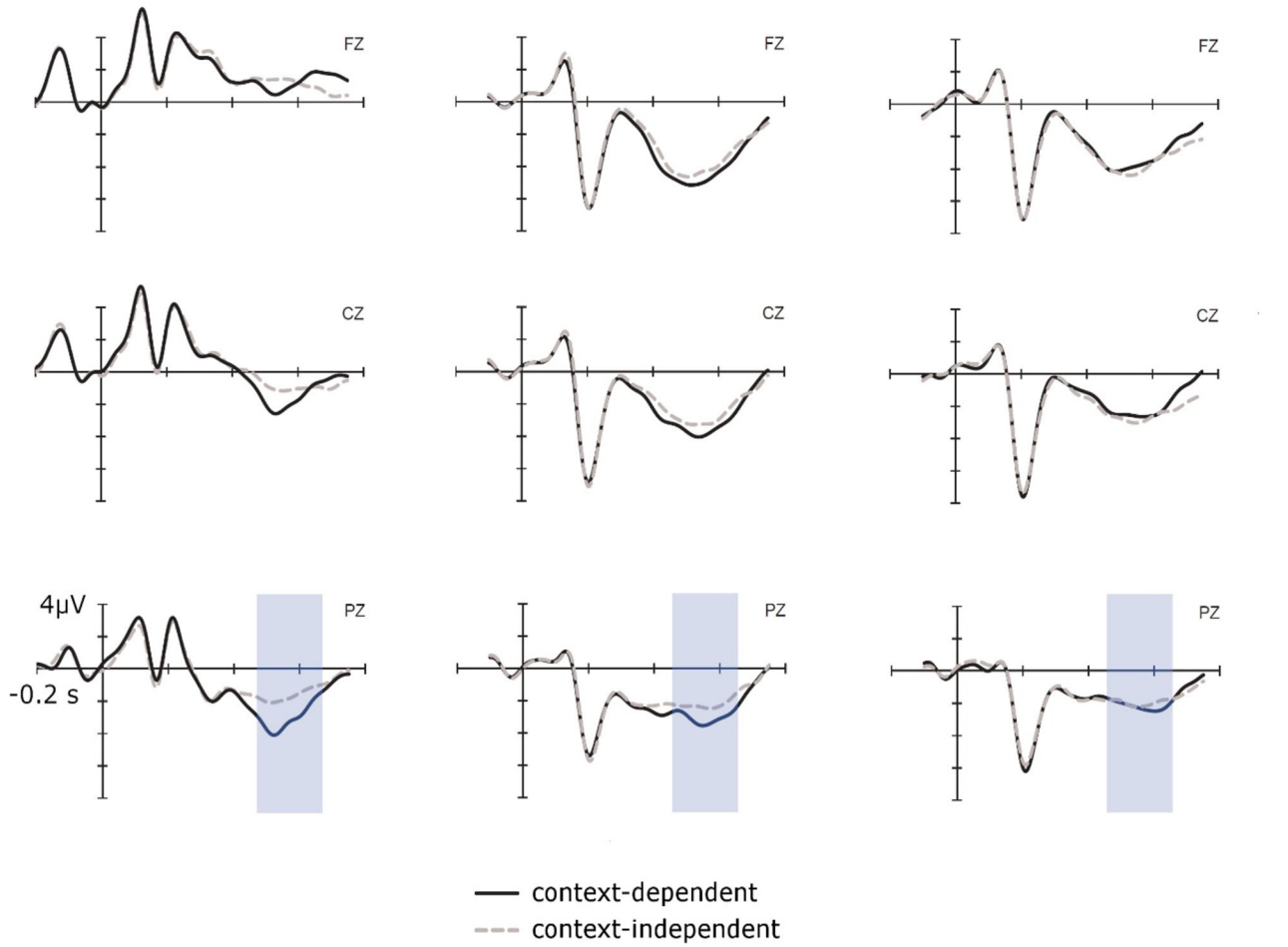


FIGURE 2 | Cue-locked ERPs for context-independent and context-dependent trials at pretest for younger adults, older adults in the single-task training group, and older adults in the task-switching training group displayed at electrodes Fz, Cz, and Pz.

older/task-switching training, younger adults). The results are shown in **Figure 6**. All groups showed a relative reduction in the context effect from pretest to posttest, $F_{(1,93)} = 10.35$, $p < 0.01$, $\eta_p^2 = 0.10$. There was only a tendency that the task-switching group showed a larger improvement than the single-task training group, that is, a greater reduction in the context effect, $F_{(1,91)} = 2.94$, $p = 0.09$, $\eta_p^2 = 0.03$. However, the difference in the context effect between the task-training group and the single-task training group was not significant at posttest, $p < 0.21$, $\eta_p^2 = 0.05$.

Transfer at the Neuronal Level

To examine near transfer to an untrained context-updating task in cue-locked P3 mean amplitudes (see **Figure 7**), we performed an ANOVA with the between-subjects factor Group (older/single task training, older/task-switching training, younger adults) and the within-subjects factors Session (pretest, posttest), Trial Type (c-dep, c-indep), and Electrode Location (Fz, Cz, Pz), using the same contrasts as before (see Results of the Context-Updating Task at Pretest) and focusing on effects including the factor Session. This ANOVA did not yield any effect including the factor Session (all $p > 0.23$), indicating no training-induced changes and no changes in age-related differences.

In sum, we obtained no training-specific changes in the context effect or the cue-P3. Thus, neither the behavioral data nor the neuronal data provided convincing evidence for transfer of training in updating in a switching task to updating of cue information in the modified AX-CPT.

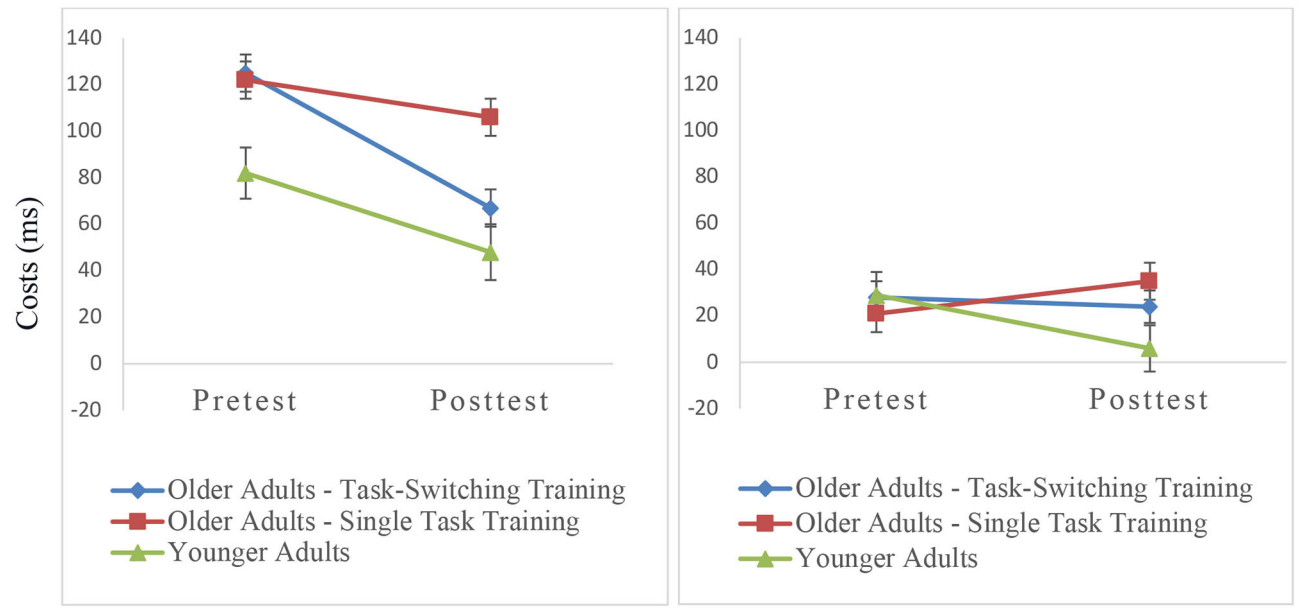


FIGURE 3 | Mixing costs **(left)** and switching costs **(right)** on the level of mean RT as a function of group and session. Error bars refer to standard errors of the mean.

DISCUSSION

The primary aim of this study was to examine whether an intensive task-switching training would improve older adults' early task-preparatory processes, such as updating higher-order task rules in advance in an untraining switching task, and whether it would transfer to another control task also requiring the updating of cue and rule information. To this end, we applied a cued-based switching task and a context updating task at pretest and posttest and measured changes in task performance and neuronal correlates of cue updating, the cue-P3, in older as well as in younger adults. Only older adults received an intensive variable training in which new tasks and stimuli were presented in each of the eight training sessions. Training-specific effects were determined by comparing a task-switching training group and a single-task training group. The task-switching training group worked through a task-switching training with two tasks changing at random in mixed blocks, thus supposedly training cue updating. The single-task training group also performed both tasks but in separate blocks, so that the task cue was redundant and updating unnecessary.

First, the results of our study revealed that both groups of older adults showed improving task performance over the eight training sessions and that these improvements were larger for the task-switching than for the single-task training group. At pretest, we also found age differences in cognitive control measures that were generally in line with those reported in the literature, namely larger mixing costs (task switching task) and context effects (context updating task) in older than in younger adults on the behavioral level (in reaction times and error rates; Kray and Ferdinand, 2014; Schmitt et al., 2014a) as well as on the neuronal level (in cue-P3; Eppinger et al., 2007; Friedman et al., 2008; Whitson et al., 2014). In addition, we found the commonly reported age-related frontal shift in cue-P3 amplitudes in the task-switching (Kray et al., 2005; Eppinger et al., 2007; Karayanidis et al., 2011; Gajewski et al., 2018) and the context updating task (Schmitt et al., 2014a,b), which is often interpreted in terms of compensation: Older adults need to recruit frontal brain areas to a larger extent in order to keep their performance up (Friedman, 2008; Reuter-Lorenz and Cappell, 2008). In the behavioral data of the task-switching task, we additionally obtained switching costs which were larger for younger than for older adults. Switching costs also emerged in the late cue-P3, but only for younger adults (Eppinger et al., 2007; Friedman et al., 2008; Karayanidis et al., 2011; Whitson et al., 2014). Importantly, both training groups of older adults did not differ in performance nor in the amplitude or topography of the cue-P3 at pretest, thus there were no baseline differences between the two older training groups.

Second, we examined potential transfer of training in task switching to an untrained switching task. On the behavioral level, we found that the task-switching group showed a larger reduction in mixing costs compared to the single task group. Moreover, at posttest, the magnitude of mixing costs did not differ between the task-switching group and the young control group but was smaller in comparison to the single-task group. Thus, only the intensive task-switching training of the older task-switching group resulted in the disappearance of age effects in behavioral mixing costs. In line with the behavioral results, in the cue-P3, younger adults showed a tendency for decreased mixing costs (cf. Küper et al., 2017) and younger as well as older adults from the switching group showed less involvement of frontal brain areas after

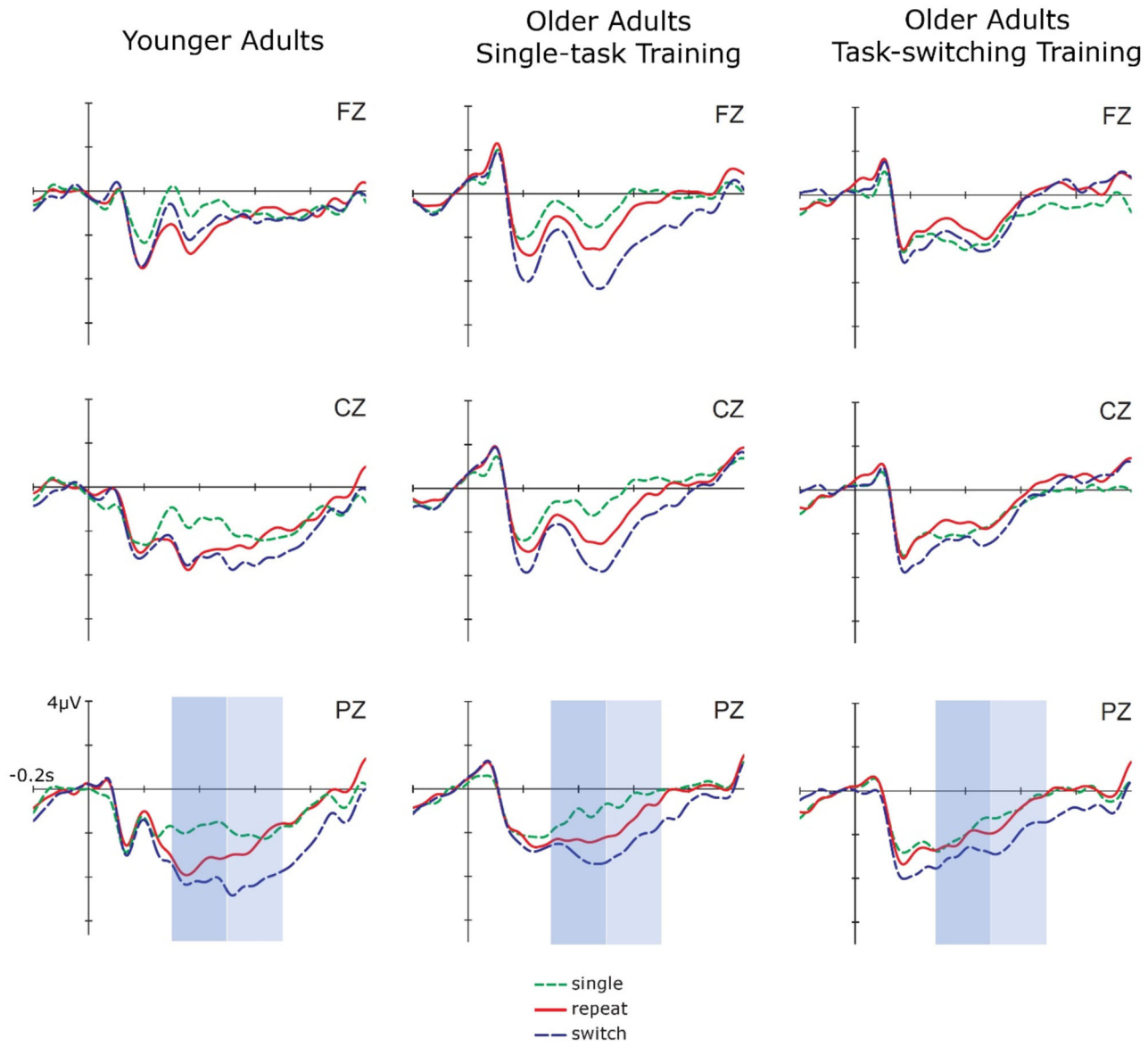


FIGURE 4 | Cue-locked ERPs in single, repeat, and switch trials at posttest for younger adults, older adults in the single task training group, and older adults in the task-switching training group displayed at electrodes Fz, Cz, and Pz.

the training. This means that a) the training resulted in more focused P3 topographies for younger and older adults, and b) that after the task-switching training, older adults' cue-P3 topography became more similar to younger adults' P3 topography, i.e., the usually found frontal P3 shift which probably indicates compensation mechanisms in older adults was diminished. In contrast, for the older adults from the single-task training group, mixing costs were even larger at posttest. By this, our results suggest that older adults can be trained in applying proactive control (here updating of task rules after cue presentation).

Together, these findings support the notion of a training-specific transfer to a similar but untrained switching task. According to the classic view of the P3b (Donchin, 1981; Donchin and Coles, 1988; Polich, 2007), one could argue that the process reflecting updating of the upcoming task rules has become more efficient after training in older adults because improvements in task switching seem to be due to updating being applied in those trials where it is actually needed. This idea also matches the result of emerging switching costs in posttest (although this effect was not specific to the switching group). These findings may also explain why a training that is less specific as for the processes trained (e.g., the cognitive training used by Küper et al., 2017) does only result in a general and unspecific improvement in updating processes, but not in more efficient updating in the sense of being able to apply it specifically in those situations

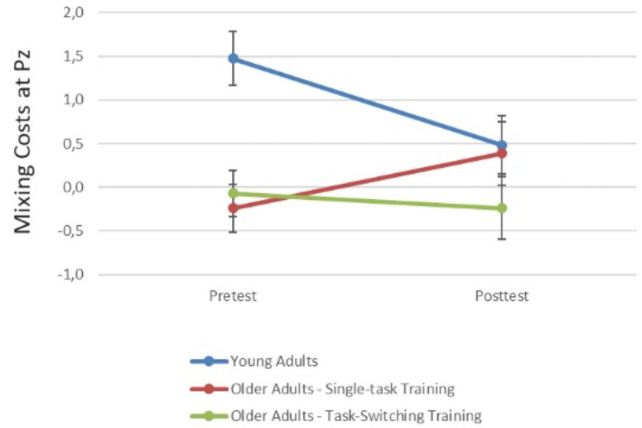


FIGURE 5 | Mixing costs (calculated as the difference between single trials and the mean of repeat and switch trials) in early cue-locked P3 amplitudes at electrode Pz for pre- and post-test. Smaller values reflect smaller mixing costs. Error bars refer to standard errors of the mean. Mixing costs seem to decrease from pre- to post-test for younger adults and older adults from the task-switching training group, while they show a tendency to increase for older adults from the single-task training group.

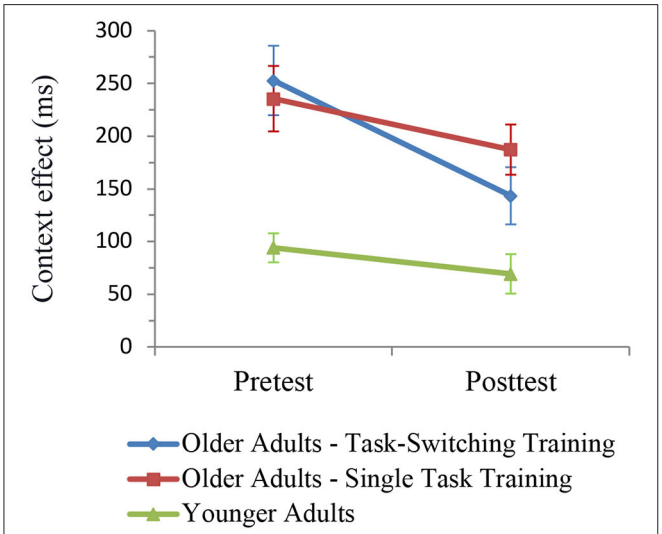


FIGURE 6 | Context effects (c-dep–c-indep) on the level of mean RT as a function of group and session. Error bars refer to standard errors of the mean.

where it is needed. Additionally, our results demonstrate that training in an easier task (the single-task training) that does not support the necessary cognitive processes (here, the preparatory updating processes that are trained in mixed blocks) may also hinder performance in a more difficult (switching) task.

Third, there was no transfer to a different cognitive control task, that is, we found no training-specific improvements in updating cue-relevant information in the AX-CPT, neither at the behavioral nor at the neuronal level. Although there was a tendency that the behavioral context effect was more strongly reduced in the task-switching group than in the single-task group, the magnitude of the context effect at posttest did not differ across groups. Moreover, the P3 amplitudes of context-dependent and context-independent trials of the older groups were nearly fully overlapping (in contrast to younger adults) with no changes from pretest to posttest. The lack of transfer in the efficiency of updating task-relevant information is surprising, given the similarity across both tasks and the assumed underlying processes involved. A possible explanation, however, could be the differences in task cues between the two tasks. While symbolic cues, namely the first two letters of the semantic category of a task rule, were used in the switching task, the cues in the AX-CPT were more arbitrary, namely pictures of young and old men and women that cued several very specific if-then rules. Thus, the symbolic cues are most likely much easier translated into the actual task goal as compared to the pictorial cues. This also means that the training of cue updating in the switching task was rather specific (letters) and did not generalize to other kinds of cues (pictures) indicating the need for updating. Another difference across the two tasks is that in the switching task cues (letters) and targets (pictures) are clearly separable which is not the case in the AX-CPT in which cues (pictures of persons) and targets (pictures of animals) are more similar. Furthermore, the response conflict was higher in the AX-CPT as four different types of S-R rules had to be performed repeatedly and in context-dependent trials the S-R rule was reversed. In contrast, in the switching task, participants performed two task sets, meaning that a class of stimuli was linked to a response depending on goals and not a specific kind of stimulus. Hence, although at first glance both tasks and paradigms have some aspects in common, such as requiring updating and switching between rules, they also differ in crucial aspects, such as cue translation, type of stimuli, and the amount of response conflict, which might all hamper transfer. The present findings also correspond to a previous finding from one of our training studies in which older adults were trained in using a verbal self-instruction, that is, they were told to name aloud the next task during the preparation interval (while no task cues were present). Interestingly, such verbal self-instructions are very useful in improving switching performance (i.e., reducing mixing costs as well as age differences therein; Kray et al., 2008). However, they were not easily transferred to a new switching situation requiring updating of new task goals and by this new verbal self-instruction rules (Karbach et al., 2010). Hence, transfer of training in updating processes seems rather narrow and limited to a specific type of switching situation.

Interestingly, there is also evidence suggesting that the updating process used to prepare for the upcoming task is not a unitary process (for a review, see Karayanidis and Jamadar, 2014), but varies with different task characteristics. Important for the present study is a finding by Nicholson et al. (2006), who found distinct early and late aspects of the cue-P3 that were related to specific aspects of switch preparation in a task-switching paradigm. Particularly, they found that cues indicating a switch to a new task rule (called switch-to cues) elicited an early and a late cue-P3, while cues signaling switch without informing about what task participants will have to conduct (termed switch-away cues) did only elicit the early aspect of the

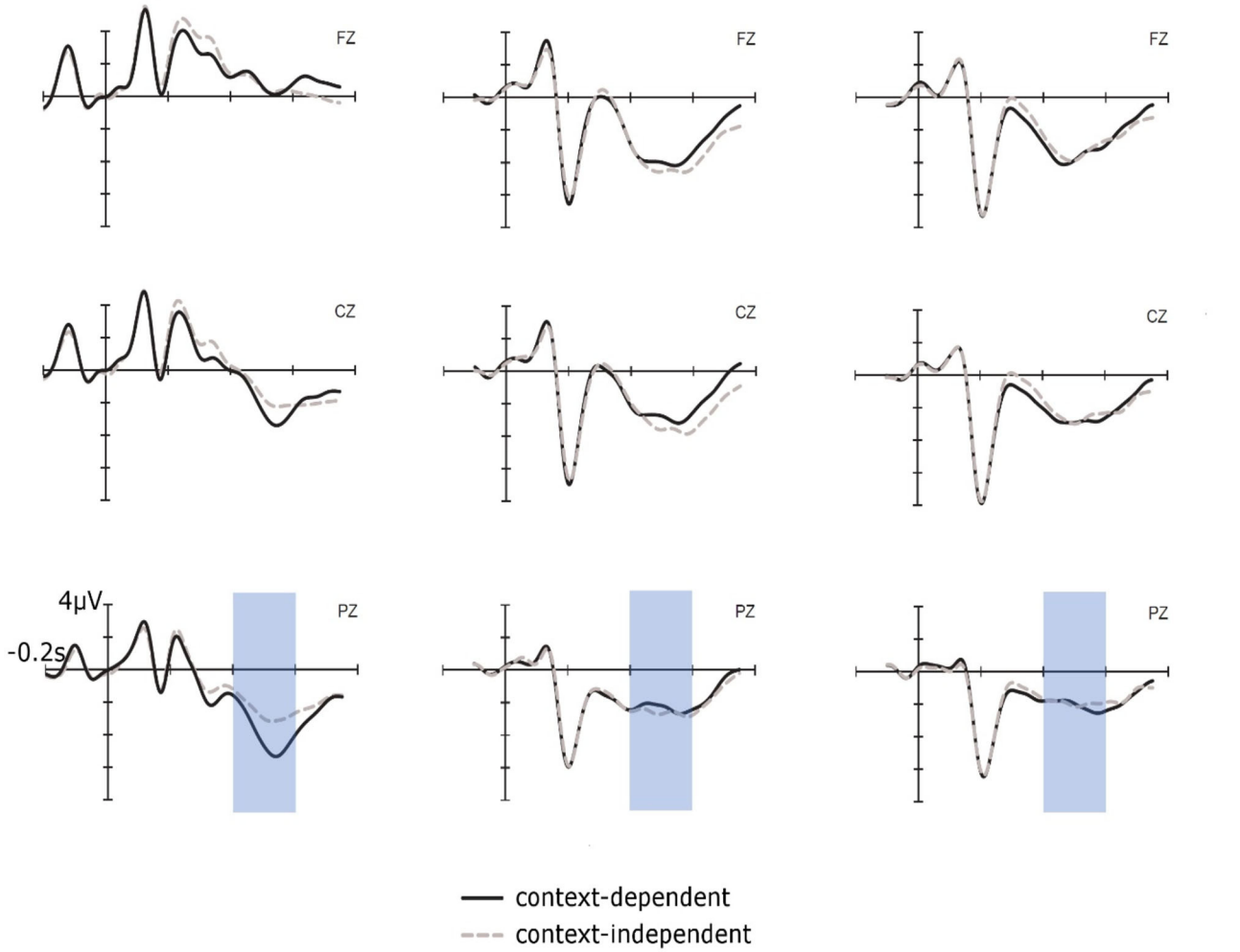


FIGURE 7 | Cue-locked ERPs for context-dependent and context-independent trials at posttest for younger adults, older adults in the single-task training group, and older adults in the task-switching training group displayed at electrodes Fz, Cz, and Pz.

cue-P3. Nicholson et al. (2006) therefore argued that the early aspect of the cue-P3 relates to disengagement from the irrelevant task or an activation of the intention to shift, while the later cue-P3 indexes the actual reloading of the relevant task. This distinction might be of great importance for the present study because they might contribute to explain the lack of transfer from our task-switching task to the context updating task. In the task-switching task used here, a cue always signaled the need to switch as well as the task rule participants would have to switch to. This was not so easy to differentiate in the context updating task where a cue could be non-informative (context-independent trials) or inform about possible tasks to execute after target presentation. Thus, the findings of Nicholson et al. (2006) further corroborate our *post-hoc* speculation that the updating processes that were trained in the task-switching training are very specific to the task at hand and not necessarily exactly the same updating processes that are needed in the context updating task.

A first limitation of the present study is the selectivity of the older sample. In general, the older participants who took part in this study were very healthy and motivated (they were able to come to the university for 10 sessions, mostly with their own car, and also took part in the fMRI part of the study). This means, that their performance probably was in the upper range of their age group and they had less room for training-related improvement. However, this was true for both groups of older adults, but still both groups showed very different

patterns in the near transfer to the untrained switching task. Therefore, it is very unlikely that it is the reason for the lack of transfer to the context-updating task. A second limitation is that the older single-task training group also had at least some experience in switching, because both types of tasks had to be performed in one training session. Nevertheless, it is unlikely that this influenced the present pattern of results, as the single-task group showed clear differences from the switching group and even a slightly negative transfer to the untrained switching task. A third limitation is that the study also included an additional fMRI session at pretest for a smaller subsample of older adults ($n = 25$ for the task-switching group and $n = 25$ for the single-task group) that might already have induced fast learning effects in performing the switching tasks. Indeed, the behavioral mixing costs were already quite low at the fMRI pretest session (71 ms for the single task group and 61 ms for the single task group) which could either be due to the smaller and selective subsample or due to differences in the trial procedure of the switching task that needed to be adapted for the event-related fMRI design. Although task-cue and target presentation times were identical, the time intervals after responding were partly longer in the fMRI design (for details, see Dörrenbächer et al., 2020). Nevertheless, although improvements during training under variable training conditions are sometimes absent, transfer effects can occur and can be larger compared to identical training conditions (practice the same switching task across sessions) as older adults are trained in adapting to new updating situations (see Karbach and Kray, 2009). This also means that the amount or the presence of training improvements is not always a precondition for the occurrence of transfer effects as it depends on the type and complexity of the training situations. More important probably is whether the training situation is demanding, and by this induces a mismatch between an individual's actual performance and the demands of the training task. Here the mismatch is induced by a variable set of tasks participants had to perform in each of the training session but can be also induced by adaptive training procedures (see Lövdén et al., 2010), which needs to be considered in the planning of training studies. Finally, although behavioral task-specific transfer effects were small in the fMRI study, we found evidence for neuronal transfer (Dörrenbächer et al., 2020). In this study, we applied a hybrid fMRI design that allowed to examine training-related changes in spatial-temporal brain activation changes. In line with the reduction of mixing costs in the cue-P3 in the present study, we found training-specific changes in brain activations for the cue-related time interval, namely a selective reduction in brain activation in the bilateral mid ventro-lateral prefrontal cortex and the left inferior frontal junction (IFJ) that are known to be involved in maintaining and top-down biasing of task-set representations, and by this support proactive task preparation. Future studies need to clarify whether the observed training-related changes in the cue-P3 in the present study are related to these brain activation changes.

To conclude, our results revealed that older adults who were trained in cue updating show training-specific improvements in preparatory processes during task switching. These improvements were mainly visible in a reduction of behavioral mixing costs and a reduction of mixing costs in the cue-related P3, indicating an improvement specifically in preparatory updating processes. Additionally, the topography of the cue-P3 changed with training from a very broad to a parietally focused scalp distribution closely resembling those in younger adults. However, transfer of the training to context-updating processes in the untrained AX-CPT were not obtained, neither at the behavioral nor at the neuronal level. These results demonstrate that transfer of training updating processes is rather narrow and limited to a specific type of switching situation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethikkommission der Fakultät HW, Universität des Saarlandes, Campus A1.3, 66123 Saarbrücken. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KS did the pre-processing of the behavioral and EEG data, run the analyses and wrote the first draft of the methods and the results part. JK re-run the behavioral analyses. NF re-run the statistical analyses of the ERP data and wrote the corresponding parts in the introduction, methods, results, and discussion. All authors contributed to the article and approved the submitted version.

FUNDING

This work was part of a dissertation project and was funded by the German Research Foundation (DFG) through a research grant to JK within the International Research Training Group (IRTG) 1457 Adaptive Minds.

ACKNOWLEDGMENTS

We acknowledge support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and Saarland University within the funding programme Open Access Publishing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.565927/full#supplementary-material

REFERENCES

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101. doi: 10.1038/nature12486

Baltes, P. B., Staudinger, U. M., and Lindenberger, U. (1999). Lifespan psychology: theory and application to intellectual functioning. *Annu. Rev. Psychol.* 50, 471–507. doi: 10.1146/annurev.psych.50.1.471

Barceló, F., and Cooper, P. S. (2018a). An information theory account of late frontoparietal ERP positivities in cognitive control. *Psychophysiology* 55:e12814. doi: 10.1111/psyp.12814

Barceló, F., and Cooper, P. S. (2018b). Quantifying contextual information for cognitive control. *Front. Psychol.* 9:1693. doi: 10.3389/fpsyg.2018.01693

Binder, J. C., Zöllig, J., Eschen, A., Mérillat, S., Röcke, C., Schoch, S., et al. (2015). Multi-domain training in healthy old age: hotel Plastisse as an iPad-based serious game to systematically compare multi-domain and single-domain training. *Front. Aging Neurosci.* 7:137. doi: 10.3389/fnagi.2015.00137

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn. Sci.* 16, 106–113. doi: 10.1016/j.tics.2011.12.010

Braver, T. S., and Barch, D. (2002). A theory of cognitive control, aging cognition, and neuromodulation. *Neurosci. Biobehav. Rev.* 26, 809–817. doi: 10.1016/S0149-7634(02)00067-2

Braver, T. S., Paxton, J. L., Locke, H. S., and Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 106, 7351–7356. doi: 10.1073/pnas.0808187106

Braver, T. S., Satpute, A. B., Rush, B. K., Racine, C. A., and Barch, D. M. (2005). Context processing and context maintenance in healthy aging and early stage dementia of the Alzheimer's type. *Psychol. Aging* 20, 33–46. doi: 10.1037/0882-7974.20.1.33

Brehmer, Y., Kalpouzos, G., Wenger, E., and Lövdén, M. (2014). Plasticity of brain and cognition in older adults. *Psychol. Res.* 78, 790–802. doi: 10.1007/s00426-014-0587-z

Donchin, E. (1981). Presidential address, 1980. Surprise!...surprise? *Psychophysiology* 18, 493–513. doi: 10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., and Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* 11, 357–427. doi: 10.1017/S0140525X00058027

Dörrenbächer, S., Schütz, C., Woirgardt, M., Wu, C., Zimmer, H. D., and Kray, J. (2019). Spatio-temporal neural changes after task-switching training in old age. *Front. Aging Neurosci.* 11:267. doi: 10.3389/fnagi.2019.00267

Dörrenbächer, S., Wu, C., Zimmer, H., and Kray, J. (2020). Plasticity in brain activity dynamics after task-shifting training in older adults. *Neuropsychologia* 136:107285. doi: 10.1016/j.neuropsychologia.2019.107285

Elchlepp, H., Lavric, A., Mizon, G. A., and Monsell, W. (2012). A brain-potential study of preparation for and execution of a task-switch with stimuli that afford only the relevant task. *Hum. Brain Mapp.* 33, 1137–1154. doi: 10.1002/hbm.21277

Eppinger, B., Kray, J., Mecklinger, A., and John, O. (2007). Age differences in response monitoring: Evidence from ERPs. *Biol. Psychol.* 75, 52–67. doi: 10.1016/j.biopsycho.2006.12.001

Friedman, D. (2008). "The components of aging," in *Oxford Handbook of Event-Related Potential Components*, eds S. J. Luck and E. S. Kappenman (New York, NY: Oxford University Press), 513–536.

Friedman, D., Nessler, D., Johnson, R. Jr., Ritter, W., and Bersick, M. (2008). Age-related changes in executive function: an event-related potential ERP investigation of task switching. *Aging Neuropsychol. Cogn.* 15, 95–128. doi: 10.1080/13825580701533769

Gajewski, P. D., and Falkenstein, M. (2012). Training-induced improvement of response selection and error detection in aging assessed by task switching: effects of cognitive, physical, and relaxation training. *Front. Hum. Neurosci.* 6:130. doi: 10.3389/fnhum.2012.00130

Gajewski, P. D., Ferdinand, N. K., Falkenstein, M., and Kray, J. (2018). Understanding sources of adult age differences in task switching: evidence from behavioral and ERP studies. *Neurosci. Biobehav. Rev.* 19, 255–275. doi: 10.1016/j.neubiorev.2018.05.029

Grange, J. A., and Houghton, G. (2014). *Task Switching and Cognitive Control*. New York, NY: Oxford University Press.

Gratton, G., Coles, M. G., and Donchin, E. (1983). A new method for offline removal of ocular artifact. *Electroencephalogr. Neurophysiol.* 55, 468–484. doi: 10.1016/0013-4694(83)90135-9

Hartshorne, J. K., and Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychol. Sci.* 26, 433–443. doi: 10.1177/0956797614567339

Jasper, H. H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 370–375. doi: 10.1016/0013-4694(58)90053-1

Karayanidis, F., Coltheart, M., Michie, P. T., and Murphy, K. (2003). Electrophysiological correlates of anticipatory and poststimulus components of task switching. *Psychophysiology* 40, 329–348. doi: 10.1111/1469-8986.00037

Karayanidis, F., and Jamadar, S. D. (2014). "Event-related potentials reveal multiple components of proactive and reactive control in task switching," in *Task Switching and Cognitive Control*, eds J. A. Grange and G. Houghton (New York, NY: Oxford University Press), 200–236.

Karayanidis, F., Mansfield, E. L., Galloway, K. L., Smith, J., Provost, A., and Heathcote, A. (2009). Anticipatory reconfiguration elicited by fully and partially informative cues that validly predict a switch in task. *Cogn. Affect. Behav. Neurosci.* 9, 202–215. doi: 10.3758/CABN.9.2.202

Karayanidis, F., Whitson, L., Heathcote, A., and Michie, P. (2011). Variability in proactive and reactive cognitive control processes across the adult lifespan. *Front. Psychol.* 2:318. doi: 10.3389/fpsyg.2011.00318

Karbach, J., and Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Dev. Sci.* 12, 978–990. doi: 10.1111/j.1467-7687.2009.00846.x

Karbach, J., and Kray, J. (2016). "Executive functions," in *Cognitive Training,* eds T. Strobach and J. Karbach (New York, NY: Springer), 93–103.

Karbach, J., Mang, S., and Kray, J. (2010). Transfer of task-switching training in older age: the role of verbal processes. *Psychol. Aging* 25:677. doi: 10.1037/a0019845

Karbach, J., and Verhaeghen, P. (2014). Making working memory work a meta-analysis of executive-control and working memory training in older adults. *Psychol. Sci.* 25, 2027–2037. doi: 10.1177/0956797614548725

Keselman, H. J., and Rogan, J. C. (1980). Repeated measures F tests and psychophysiological research: controlling the number of false positives. *Psychophysiology* 17, 499–503. doi: 10.1111/j.1469-8986.1980.tb00190.x

Kieffaber, P. D., and Hetrick, W. P. (2005). Event-related potential correlates of task switching and switch costs. *Psychophysiology* 42, 56–71. doi: 10.1111/j.1469-8986.2005.00262.x

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching-a review. *Psychol. Bull.* 136, 849–874. doi: 10.1037/a0019842

Kopp, B., Lange, F., Howe, J., and Wessel, K. (2014). Age-related changes in neural recruitment for cognitive control. *Brain Cogn.* 85, 209–219. doi: 10.1016/j.bandc.2013.12.008

Kramer, A. F., Hahn, S., and Gopher, D. (1999). Task coordination and aging: explorations of executive control processes in the task switching paradigm. *Acta Psychol.* 101, 339–378. doi: 10.1016/S0001-6918(99)00011-6

Kray, J. (2006). Task-set switching under cue-based and memory-based switching conditions in younger and older adults. *Brain Res.* 1105, 83–92. doi: 10.1016/j.brainres.2005.11.016

Kray, J., and Dörrenbächer, S. (2019). "The effectiveness of training in task switching: new insights and open issues from a lifespan view," in *Cognitive and Working Memory Training: Perspectives From Psychology, Neuroscience, and Human Development,* eds J. M. Novick, M. F. Bunting, R. W. Engle, and M. R. Dougherty (New York, NY: Oxford University Press), 508–538.

Kray, J., Eber, J., and Karbach, J. (2008). Verbal self-instructions in task switching: a compensatory tool for action-control deficits in childhood and old age? *Dev. Sci.* 11, 223–236. doi: 10.1111/j.1467-7687.2008.00673.x

Kray, J., Eppinger, B., and Mecklinger, A. (2005). Age differences in attentional control. An event-related potential approach. *Psychophysiology* 42, 407–416. doi: 10.1111/j.1469-8986.2005.00298.x

Kray, J., and Fehér, B. (2017). Age differences in the transfer and maintenance of practice-induced improvements in task switching: the impact of working-memory and inhibition demands. *Front. Psychol.* 8:410. doi: 10.3389/fpsyg.2017.00410

Kray, J., and Ferdinand, N. (2014). "Task switching and aging," in *Task Switching and Cognitive Control*, eds J. Grange and G. Houghton (New York, NY: Oxford University Press), 350–373.

Kray, J., and Lindenberger, U. (2000). Adult age differences in task switching. *Psychol. Aging* 15, 126–147. doi: 10.1037/0882-7974.15.1.126

Kühn, S., and Lindenberger, U. (2016). "Research on human plasticity in adulthood: a lifespan agenda," in *Handbook of the Psychology of Aging*, (Cambridge, MA: Academic Press), 105–123.

Küper, K., Gajewski, P. D., Frieg, C., and Falkenstein, M. (2017). A randomized controlled erp study on the effects of multi-domain cognitive training and task difficulty on task switching performance in older adults. *Front. Hum. Neurosci.* 11:184. doi: 10.3389/fnhum.2017.00184

Lavric, A., Mizon, G. A., and Monsell, S. (2008). Neurophysiological signature of effective anticipatory task-set control: a task-switching investigation. *Eur. J. Neurosci.* 28, 1016–1029. doi: 10.1111/j.1460-9568.2008.06372.x

Lehrl, S. (1977). *MWT-B: Mehrfach-Wahl-Wortschatz-Test B*. Erlangen: Straube.

Lenartowicz, A., Escobedo-Quiroz, R., and Cohen, J. D. (2010). Updating of context in working memory: an event-related potential study. *Cogn. Affect. Behav. Neurosci.* 10, 298–315. doi: 10.3758/CABN.10.2.298

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., and Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychol. Bull.* 136, 659–676. doi: 10.1037/a0020080

Mayr, U. (2001). Age differences in the selection of mental sets: The role of inhibition, stimulus ambiguity, and response-set overlap. *Psychol. Aging* 16, 96–109. doi: 10.1037/0882-7974.16.1.96

Minear, M., and Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behav. Res. Methods Instruments Comput.* 36, 630–633. doi: 10.3758/BF03206543

Miyake, A., and Friedman, N. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Curr. Direct. Psychol. Sci.* 21, 8–14. doi: 10.1177/0963721411429458

Nicholson, R., Karayanidis, F., Davies, A., and Michie, P. T. (2006). Components of task-set reconfiguration: differential effects of "switch-to" and "switch-away" cues. *Brain Res.* 1121, 160–176. doi: 10.1016/j.brainres.2006.08.101

Nicholson, R., Karayanidis, F., Poboka, D., Heathcote, A., and Michie, P. T. (2005). Electrophysiological correlates of anticipatory task-switching processes. *Psychophysiology* 42, 540–554. doi: 10.1111/j.1469-8986.2005.00350.x

Novick, J. M., Bunting, M. F., Engle, R. W., and Dougherty, M. R. (eds.). (2019). *Cognitive and Working Memory Training: Perspectives from Psychology, Neuroscience, and Human Development*. Oxford: Oxford University Press.

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

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Paxton, J., Barch, D., Racine, C., and Braver, T. S. (2008). Cognitive control, goal maintenance, and prefrontal function in healthy aging. *Cereb. Cortex* 18, 1010–1028. doi: 10.1093/cercor/bhm135

Paxton, J. L., Barch, D. M., Storandt, M., and Braver, T. S. (2006). Effects of environmental support and strategy training on older adults' use of context. *Psychol. Aging* 21, 499–509. doi: 10.1037/0882-7974.21.3.499

Pereg, M., Shahar, N., and Meiran, N. (2013). Task switching training effects are mediated by working-memory management. *Intelligence* 41, 467–478. doi: 10.1016/j.intell.2013.06.009

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

Psychology Software Tools, Inc (2012). *E-Prime 2.0*. Pittsburgh, PA.

Reuter-Lorenz, P. A., and Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Curr. Direct. Psychol. Sci.* 17, 177–182. doi: 10.1111/j.1467-8721.2008.00570.x

Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33, 217–236. doi: 10.1068/p5117

Rush, B. K., Barch, D. M., and Braver, T. S. (2006). Accounting for cognitive aging: context processing, inhibition or processing speed? *Aging Neuropsychol. Cogn.* 13, 588–610. doi: 10.1080/13825580600680703

Schmitt, H., Ferdinand, N. K., and Kray, J. (2014a). Age-differential effects on updating cue information: evidence from event-related potentials. *Cogn. Affect. Behav. Neurosci.* 14, 1115–1131. doi: 10.3758/s13415-014-0268-9

Schmitt, H., Ferdinand, N. K., and Kray, J. (2015). The influence of monetary incentives on context processing in younger and older adults: an event-related potential study. *Cogn. Affect. Behav. Neurosci.* 15, 416–434. doi: 10.3758/s13415-015-0335-x

Schmitt, H., Kray, J., and Ferdinand, N. K. (2017). Does the effort of processing potential incentives influence the adaption of context upating in older adults? *Front. Psychol.* 8:1969. doi: 10.3389/fpsyg.2017.01969

Schmitt, H., Wolff, M. C., Ferdinand, N. K., and Kray, J. (2014b). Age differences in the processing of context information. *J. Psychophysiol.* 28, 202–214. doi: 10.1027/0269-8803/a000126

Snodgrass, J., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Memory* 6:174. doi: 10.1037/0278-7393.6.2.174

Strobach, T., and Karbach, J., (eds.). (2016). *Cognitive Training: An Overview of Features and Applications*. New York, NY: Springer.

Swainson, R., Jackson, S. R., and Jackson, G. M. (2006). Using advance information in dynamic cognitive control: an ERP study of task-switching. *Brain Res.* 1105, 61–72. doi: 10.1016/j.brainres.2006.02.027

Wasylyshyn, C., Verhaeghen, P., and Sliwinski, M. J. (2011). Aging and task switching: a meta-analysis. *Psychol. Aging* 26, 15–20. doi: 10.1037/a0020912

Wechsler, W. (1981). *The Manual for the Wechsler Adult Intelligence Scale Revised*. New York, NY: The Psychological Cooperation.

West, R., and Travers, S. (2008). Differential effects of aging on processes underlying task switching. *Brain Cogn.* 68, 67–80. doi: 10.1016/j.bandc.2008.03.001

West, T., and Schwarb, H. (2006). The influence of aging and frontal function on the neural correlates of regulative and evaluative aspects of cognitive control. *Neuropsychology* 20, 469–481. doi: 10.1037/0894-4105.20.4.468

Whitson, L., Karayanidis, F., Fulham, R., Provost, A., Michie, P., Heathcote, A., et al. (2014). Reactive control processes contributing to residual switch cost and mixing cost across the adult lifespan. *Front. Psychol.* 5:383. doi: 10.3389/fpsyg.2014.00383

Whitson, L. R., Karayanidis, F., and Michie, P. T. (2012). Task practice differentially modulates task-switching performance across the adult lifespan. *Acta Psychol.* 139, 124–136. doi: 10.1016/j.actpsy.2011.09.004

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 18 December 2020
doi: 10.3389/fnhum.2020.610548

The Effect of Bilingualism on Cue-Based vs. Memory-Based Task Switching in Older Adults

Jennifer A. Rieker[1,2], *José Manuel Reales*[1,3] *and Soledad Ballesteros*[1,2]*

[1]*Studies on Aging and Neurodegenerative Diseases Research Group, Madrid, Spain,* [2]*Department of Basic Psychology II, Facultad de Psicología, Universidad Nacional, Madrid, Spain,* [3]*Department Methodology of Behavioral Sciences, Facultad de Psicología, Universidad Nacional de Educación a Distancia, Madrid, Spain*

Findings suggest a positive impact of bilingualism on cognition, including the later onset of dementia. However, it is not clear to what extent these effects are influenced by variations in attentional control demands in response to specific task requirements. In this study, 20 bilingual and 20 monolingual older adults performed a task-switching task under explicit task-cuing vs. memory-based switching conditions. In the cued condition, task switches occurred in random order and a visual cue signaled the next task to be performed. In the memory-based condition, the task alternated after every second trial in a predictable sequence without presenting a cue. The performance of bilinguals did not vary across experimental conditions, whereas monolinguals experienced a pronounced increase in response latencies and error rates in the cued condition. Both groups produced similar switch costs (difference in performance on switch trials as opposed to repeating trials within the mixed-task block) and mixing costs (difference in performance on repeat trials of a mixed-task block as opposed to trials of a single-task block), but bilinguals produced them with lower response latencies. The cognitive benefits of bilingualism seem not to apply to executive functions *per se* but to affect specific cognitive processes that involve task-relevant context processing. The present results suggest that lifelong bilingualism could promote in older adults a flexible adjustment to environmental cues, but only with increased task demands. However, due to the small sample size, the results should be interpreted with caution.

Keywords: aging, bilingualism, cued task switching, memory-based task switching, executive function

OPEN ACCESS

Edited by:
Susan Mérillat,
University of Zurich, Switzerland

Reviewed by:
Kenneth R. Paap,
San Francisco State University, United States
Ashley Chung Fat Yim,
Northwestern University, United States

***Correspondence:**
Soledad Ballesteros
mballesteros@psi.uned.es

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

Received: *26 September 2020*
Accepted: *27 November 2020*
Published: *18 December 2020*

Citation:
Rieker JA, Reales JM and Ballesteros S (2020) The Effect of Bilingualism on Cue-Based vs. Memory-Based Task Switching in Older Adults. Front. Hum. Neurosci. 14:610548. doi: 10.3389/fnhum.2020.610548

INTRODUCTION

Modern societies are characterized by population aging due to increased life expectancy and falling birth rates, with older adults making up a growing proportion of the population (Gavrilov and Heuveline, 2003). This demographic aging implies exponential growth in the number of people who will experience age-related declines in cognition, and in the incidence and prevalence of dementia, and entails an important economic impact for caregivers and public health systems (World Health Organization, 2012; Hurd et al., 2013). However, not all people respond similarly to a neuropathological burden. While cerebral changes result in significant cognitive declines in some older adults, others can compensate for these changes and maintain their normal cognitive functioning up to advanced age (Riley et al., 2002). This phenomenon is referred to as cognitive reserve (Barulli and Stern, 2013).

Cognitive reserve is defined as the interindividual variability in how tasks are processed, allowing some people to cope better than others with brain pathology and age-related brain changes (Stern, 2009). Several activities and other environmental factors have been identified as fostering cognitive reserve, such as higher educational and occupational achievements (Bennett et al., 2003), or engaging in cognitively stimulating leisure activities (Ferreira et al., 2015; Ballesteros et al., 2018). It has been suggested that bilingualism contributes to this reserve as well, as it has been shown that, on average, bilinguals are diagnosed with Alzheimer's Disease approximately 4 years later than monolinguals (Bialystok et al., 2007; Craik et al., 2010; Woumans et al., 2015), although some large prospective studies could not replicate this effect (for a recent review see Van den Noort et al., 2019). The benefits of the cognitive reserve can also be observed in healthy aging. Normal aging is associated with neurobiological changes that produce progressive declines in different cognitive domains (Park and Reuter-Lorenz, 2009; Reuter-Lorenz and Park, 2014), and most older adults manage to compensate for these cerebral changes by recruiting additional brain areas, or by overrecruiting frontal areas (Davis et al., 2008; Osorio et al., 2010). It appears that healthy older bilinguals perform non-verbal executive tasks without having to over-activate frontal areas (Gold et al., 2013; Ansaldo et al., 2015; for a recent review see Zhang et al., 2020) suggesting that the simultaneous management of two languages might lead to better maintenance of cerebral functionality in advanced age.

Bilinguals constantly need to monitor and control two different language codes that share the same neural substrate (Crinion et al., 2006), and one language is produced by inhibiting the other (Runnqvist et al., 2012). This increased demand for cognitive control seems to lead on some occasions to superior performance in tasks that involve executive functions (EF; see Adesope et al., 2010; Bialystok et al., 2012). Studies with children (Carlson and Meltzoff, 2008; Kapa and Colombo, 2013; for a review see Barac et al., 2014) and older adults (Bialystok et al., 2004; Salvatierra and Rosselli, 2010; Goral et al., 2015) have reported a bilingual advantage in executive control. With younger adults, results are more mixed (for reviews of results in young adults vs. results with children and older adults, see Bialystok, 2017; Antoniou, 2019), and bilingual brain mechanisms might compensate for lower-level executive functioning, for example, in childhood when executive functions are still developing (Casey et al., 2000), or in late adulthood when age-related decline appears (Zelazo et al., 2004). Several studies have shown that the bilingual advantage increases with task difficulty (Bialystok, 2006; Costa et al., 2009; Hernández et al., 2013; Qu et al., 2015). However, other studies have failed to find evidence for a cognitive benefit of bilingualism (Paap and Greenberg, 2013; Antón et al., 2016; Scaltritti et al., 2017). Different factors have been proposed as contributing to the inconsistencies found in the literature, such as task impurities when assessing EF (Hartanto and Yang, 2020), as well as differences in study designs, assessment tasks, and insufficient assessment of other variables known to modulate cognition such as physical exercise and cognitive stimulation (Calvo et al., 2016). Recent meta-analyses (Lehtonen et al., 2018; Donnelly et al., 2019) conclude that the average effect size for a bilingual advantage is small and that it disappears when controlling for publication bias (Paap et al., 2020). However, growing evidence suggests that attentional advantages might be related to long-term dual-language management (Stocco et al., 2014). The amount of the second language (L2) immersion (time spent in the country where L2 is spoken) and the frequency of language switching are important modulating factors of the effects of bilingualism on cognition (Prior and Gollan, 2011; Pliatsikas et al., 2016; Pot et al., 2018).

Most of the studies that have investigated EF in bilinguals have focused on inhibitory control (Bialystok et al., 2004; Costa et al., 2009) and task switching (Costa et al., 2008; Prior and Gollan, 2011; for a review see Bialystok, 2017). The assumption that inhibition is part of the mechanism for bilingual effects on cognition is based on the inhibitory control model (Green, 1998). According to this model, a supervisory attention system is guided by top-down cues, leading to the inhibition of the non-target language so that language processing can adapt to the contextual requirements. Extensions of this model (Green and Abutalebi, 2013; Green and Wei, 2014) include the differential influences of cognitive control processes as a function of the type of interactional context for language use and distinguish between three different contexts: (1) single-language; (2) dual-language; and (3) dense code-switching. In a single-language context, bilinguals use only one language in the same situation. In dual-language and code-switching contexts, bilinguals switch between the two languages in the same situation, but in the case of code-switching, languages are freely mixed in single utterances. Hartanto and Yang (2020) found that bilinguals with greater exposure to a dual-language context displayed significantly better task-switching abilities, replicating their findings of a previous study (Hartanto and Yang, 2016). They also found that dense code-switching was related to better inhibitory control and goal maintenance (Hartanto and Yang, 2020), a result that contrasts with a nonsignificant result regarding the relationship between dense code-switching and inhibitory control in another recent study (Kaamaa et al., 2020). It seems that within dual-language contexts, situations that require constant goal reconfiguration and top-down control in response to outside constraints are more likely to translate into a cognitive advantage than free and unrestrained language switches (Blanco-Elorrieta and Pylkkänen, 2018).

On the other hand, the interest in the relationship between bilingualism and task-switching stems from behavioral data that show similar dynamics when shifting between dominant and less dominant templates (Meuter and Allport, 1999; Runnqvist et al., 2012). Further support for the commonalities between attentional set-shifting and dual-language management comes from neuroimaging evidence that shows an overlap in brain networks involved in language selection and nonverbal task switching (Meuter and Allport, 1999; Abutalebi and Green, 2007; Luk et al., 2011; Runnqvist et al., 2012; Baene et al., 2015; Coderre et al., 2016).

Cognitive processing of mental set-shifting might also vary as a function of task requirements. The conditional routing model (Stocco et al., 2010, 2014) proposes that bilingualism improves the ability to flexibly reallocate attention in complex and non-habitual task requirements, whereas the management of more direct stimulus-response mappings is not influenced by bilingual language processing. An example could be the reorientation in response to unpredictable external cues vs. reorientation in response to rule changes that occur in a sequenced order. In both cases, working memory (WM) plays an important role. WM allows for simultaneously maintaining and processing information to guide goal-directed behavior (Baddeley and Hitch, 1994). In memory-based, as well as in cued task switches, task sets need to be monitored and retrieved from memory and assembled with the correct stimulus-response mapping. However, the activation process is different for memory-based and randomly cued task switches. In memory-based set-shifting, the activation is triggered endogenously by a goal-directed monitorization in WM. When cued task switches occur randomly, the demand for a set shift is unpredictable and cannot be controlled by internal monitoring. In this case, the task-set activation is stimulus-driven; that is, triggered by a task-relevant cue (Corbetta et al., 2008).

Task-switching paradigms typically consist of blocks of switch and repeat trials and blocks of non-switch trials where only single-task sets are performed. The difference in performance between switch and repeat trials is called "switch cost" and reflects task-set reconfiguration processes associated with changing task sets across trials (Monsell, 2003). The difference in performance between repeat trials in the switch block and trials in the single-task block is called "mixing cost." This difference is thought to reflect the active maintenance of multiple task configurations in working memory and is more sensitive to age-related cognitive changes (Kray and Lindenberger, 2000).

Task-switching paradigms comprise different variants of switch tasks. In the cued-switching version, shifts are generally random, and a cue signals the task to be performed next. In alternating-run versions, shifts occur in a predictable sequence after every N-trial, with or without the appearance of a cue. If no cue accompanies the sequence, then set-shifting is "memory-based," as switches are triggered endogenously by working memory. To our knowledge, to date, only four studies have investigated task-switching abilities in older adults and three of them found significant group differences. Gold et al. (2013) analyzed performance in memory-based switching with predictable task sequences and found that bilinguals showed lower switch costs than their monolingual counterparts, with overall better levels of behavioral performance. Using a cued task-switching paradigm, Houtzager et al. (2015) found that switch and mixing costs were lower in the bilingual group. de Bruin et al. (2015) compared active and nonactive older bilinguals and monolinguals. They found a significant difference in raw switch costs between active bilinguals and monolinguals, which disappeared when controlling for baseline performance. Soveri et al. (2011) also used a cued task-switching paradigm, but their within-group design did not include a monolingual control group. Although the participants were slightly younger than in the other two studies, a positive relation was found between lower mixing costs and frequent language switching.

The present study had two main goals. The first was to investigate the influence of explicitly cued vs. memory-based switching conditions on the set-shifting abilities of bilingual and monolingual older adults. Specifically, we were interested to find out whether bilingualism would influence mental flexibility *per se*, or if differences between monolinguals and bilinguals would be more prominent when task switches were externally triggered (aleatory rule changes in response to a cued) in comparison to task switches that were endogenously triggered (memory-based sequential changes).

Therefore, our experimental design included two conditions requiring different types of attentional control: first, a memory-based switching condition based on the alternating-runs paradigm in which the task alternates every N-trial; second, a cued switching condition based on an explicit task-cuing paradigm with randomly alternating tasks, each preceded by an instructive cue (Monsell et al., 2003). Memory-based task switching is predictable and controlled endogenously by working memory processes (Monchi et al., 2001), whereas cued task-switching requires a context-dependent reorientation of attention (Monchi et al., 2001; Baene et al., 2015). Given the similarity of explicitly cued task switching and context-related dual-language management, we expected bilinguals to produce lower switch costs than monolinguals when task-set reconfiguration had to be adjusted in response to unpredictable external cues, whereas there would be no difference between groups when set-shifting was memory-based and triggered endogenously.

The second goal of our study was to investigate whether bilingualism influences age-related decline in WM. A large body of research has provided evidence of a positive relationship between cognitive aging and mixing costs (i.e., the difference between repeat trials of a mixed task block and non-switch trials of a single-task block; Kray and Lindenberger, 2000; Reimers and Maylor, 2005; Wasylyshyn et al., 2011; Huff et al., 2015). Mixing costs reflect the active maintenance of multiple task configurations in working memory and could be expected to increase when task switches are memory-based. However, the aging effect on mixing costs seems to increase with increasing task complexity (Kray, 2006; Terry and Sliwinski, 2012). Task complexity increases when rule changes are unpredictable and dependent on external cues, as the reconfiguration process additionally requires the correct interpretation and implementation of the informative cue (Tornay and Milán, 2001). For this reason, we expected to find larger mixing costs in the cued-switching condition than in the memory-based condition and that mixing costs would be larger in monolingual older adults than in bilingual older adults.

MATERIALS AND METHODS

Participants

Forty-two older adults were recruited through flyers and media postings, informative talks at strategic locations, and snowball sampling (referrals from participants). The inclusion

criteria were a score of 26 or above on the Mini-Mental State Examination (MMSE; Folstein et al., 1975), a score of below 5 on the Yesavage Geriatric Depression Scale (Yesavage et al., 1983; Spanish adaptation by Martínez de la Iglesia et al., 2002), no current history of psychiatric or neurological pathology, and for the monolingual participants, no mastery of a foreign language above the A1 level of the *Common European Framework of Reference for Languages* (CEFR). One bilingual participant did not meet the inclusion criteria (score above 5 on the depression scale) and was excluded from further analysis. Data of one monolingual participant was not recorded due to technical problems. Thus, the final sample was composed of 20 monolingual native Spanish older adults (eight males, M_{age} = 72.65, SD = 6.38, range = 60–83 years) and 20 German-Spanish bilingual older adults (four males, M_{age} = 72.25, SD = 9.12, range = 60–95 years). **Table 1** summarizes the demographics and screening test scores for monolinguals and bilinguals. *T*-tests showed no significant differences between the two groups (all *p*s > 0.05) for all of these measures. Growing evidence suggests that the amount of the second language (L2) immersion (time spent in the country where L2 is spoken) and the frequency of language switching are important modulating factors of the effects of bilingualism on cognition (Prior and Gollan, 2011; Pliatsikas et al., 2016; Pot et al., 2018; Hartanto and Yang, 2020). Our bilingual sample was composed of highly balanced, late bilinguals who had been exposed to their L2-environment for more than 40 years on average. Fourteen bilinguals reported German as their first language (L1) and Spanish as their second language (L2), and six reported Spanish as their L1 and German as their L2. All participants were right-handed, had normal or corrected-to-normal vision and none reported color blindness.

Bilingualism was assessed with the validated Bilingual Language Profile questionnaire (BLP; Birdsong et al., 2012; see Table S1 for detailed information on the BLP). It has four components with a mean Cronbach's alpha of 0.787 (Gertken et al., 2014): language history (e.g., "At what age did you start learning the following languages?" "How many years have you spent in a country/region where the following languages are spoken?"), language use (e.g., "In an average week, what percentage of the time do you use the following languages with friends?" "When you count, how often do you count in the following languages?"), language proficiency (e.g., "How well do you speak Spanish?" "How well do you read Spanish?") and language attitudes (e.g., "I feel like myself when I speak Spanish," "I identify with a Spanish-speaking culture"). For each component, two scores are computed (one for each language) and the difference between the two scores indicates the relative dominance of each language in that specific area. The scores for each component vary as follows: −120 to +120 for language history, −50 to +50 for usage, −24 to +24 for proficiency, and −24 to +24 for attitudes. The score of each component is multiplied by a weighting factor so that each component receives equal weighting (54.5) in the global language score. The difference between the total scores of the two languages constitutes the language dominance index, which ranges from −218 to +218. In the present study, we subtracted the German score from the Spanish score. A positive score indicated dominance in Spanish, and a negative score indicated dominance in German. A score of zero represents a balanced bilingualism. The linguistic background information for bilinguals is shown in **Table 2**. No statistically significant differences were found between monolinguals and bilinguals regarding the demographic background information.

TABLE 1 | Mean values of socio-demographic background variables for monolinguals and bilinguals.

	Monolinguals (*n* = 20)	Bilinguals (*n* = 20)	*t* (*df*)	*p*
Men/women	8/12	4/16	$t_{(38)} = -1.378$	0.176
Age	72.25 (6.38)	72.65 (9.12)	$t_{(38)} = -0.161$	0.873
Education[1]	4.55 (2.06)	4.65 (1.42)	$t_{(38)} = -0.178$	0.859
MMSE[2]	28.85 (1.04)	29.3 (0.8)	$t_{(38)} = -1.533$	0.134
Depression[3]	1.2 (1.2)	0.7 (0.92)	$t_{(38)} = -1.480$	0.147

[1]Level of educational attainment was defined as follows: 1 = Primary education, 2 = Lower secondary education, 3 = Post-secondary non-tertiary education, 4 = Upper secondary education, 5 = Short-cycle tertiary education, 6 = Bachelor's or equivalent, 7 = Doctoral or equivalent. [2]Mini-Mental State Examination (Folstein et al., 1975). [3]Short Form of the Geriatric Depression Scale (GDS; Yesavage et al., 1983). SDs are shown in parentheses.

TABLE 2 | Mean values of socio-demographic background variables[1] for monolinguals and bilinguals.

Spanish use (% week)	40 (21.82)
German use (% week)	60 (21.95)
Age of acquisition	19.9 (7.41)
BLP global score	−28.66 (61.93)
Language history	−12.3 (23.22)
Language use	−11.35 (23.93)
Language proficiency	−0.67 (9.49)
Language attitudes	−5.09 (15.77)

[1]Bilingual Language Profile (BLP; Birdsong et al., 2012). Negative values indicate dominance in German. SDs are shown in parentheses.

All participants gave their written informed consent. The study protocol was approved by the Institutional Review Board of the Universidad Nacional de Educación a Distancia (UNED) and the study was conducted following the ethical guidelines of the 1975 Declaration of Helsinki.

Assessing Task Switching

The experimental task was adapted from Rubin and Meiran (2005) and contained three conditions: (1) in the single-task condition only one task had to be performed at a time; (2) in the cued-switching condition two tasks alternated in random order and a cue signaled the task to be performed next; and (3) in the memory-based switching condition two tasks alternated after every second trial without the appearance of a cue. It involved two bivalent target stimuli with two possible shapes (circle or square, both 60 × 60 mm) and in one of two possible colors (yellow or blue), presented in the center of the screen on a black background. In the cued-switching condition, a visual cue signaled the next task to be performed: a white splotch (18.8 mm) indicated that participants would have to identify the color of the target stimulus, and the white outline of a star (18.8 mm) that they would have to identify its shape. Although the cue was irrelevant in single-task blocks, it was

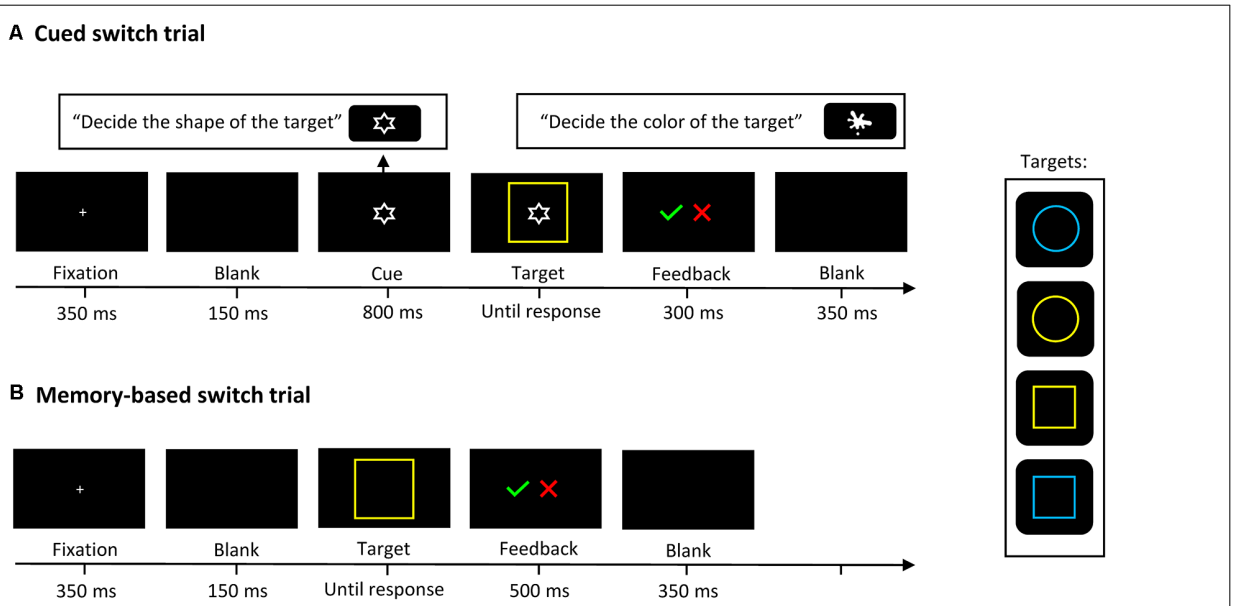


FIGURE 1 | A schematic representation of the task-switching paradigm. In Cued switch trials **(A)**: an instructional cue indicated the next task to be performed. In Memory-based switch trials **(B)**: the task changed after every second trial without the appearance of a cue; that is, participants had to identify the shape of two consecutive stimuli and the color of the next two stimuli, and so forth. In single-task trials (not figured), participants only had to identify the color or the shape of the target.

presented in both single-task and cued-switching blocks to minimize differences between the conditions. In the memory-based switching condition, to help participants keep track of the correct trial sequence in the event of an error, two cues appeared on the screen (the same pictorial cues as in the cued-switching block), one indicating the correct condition of the just-completed trial, and one signaling the following trial condition. For a schematic representation of the task-switching paradigm, see **Figure 1**.

Each experimental run comprised eight blocks of trials. The first two blocks (23 trials each) were single-task blocks, one for shape and one for color. The third block was a cued-switching block with 46 trials (23 switch trials and 23 repeat trials), presented in a semi-random order with a maximum of three consecutive trials of the same condition and a maximum of two trials in which the condition and response mapping were identical to the preceding trial. The fourth block was a memory-based switching block, composed of 23 switch trials and 23 repeat trials. The following four blocks were a repetition of the previous trial blocks but in reverse order, starting with the memory-based switching block, followed by the cued-switching block, and ending with the single-task blocks. Altogether, the experiment contained 46 switch trials and 46 repeat trials in the cued condition, 46 switch trials and 46 repeat trials in the memory condition, and 92 non-switch trials (46 for color and 46 for shape) in the single-task condition, yielding a total of 276 trials per run.

Procedure

Participants were tested individually in a single session. The experimental session lasted about 90 min. Stimuli were displayed on a laptop computer with a 15.6-inch monitor and a refresh rate of 60 Hz. Experimental scripts were designed, and data collection was managed with E-Prime 2.0 (Psychology Software Tools Inc., Pittsburg, PA, USA) experimental software. Participants were comfortably seated approximately 60 cm from the monitor. Non-switch trials and cued switch trials started with the presentation of the fixation point in the center of the screen for 350 ms, followed by a 150 ms blank screen. Then the instructional task cue appeared, and after 800 ms the target stimulus surrounded the cue and both stimuli remained on the screen until a response was given, or for a maximum of 10 s. Auditory feedback was presented for 300 ms (an incorrect response was followed by a low-frequency beep and a correct response by a high-frequency beep). The trial ended with a 350 ms blank screen. Memory-based switch trials also started with a 350 ms fixation point, followed by a 150 ms blank screen. Then the target stimulus appeared in the middle of the screen and remained until an answer was given or for 10 s. The auditory feedback was presented for 500 ms, and in the event of an incorrect response, two informative cues appeared on the screen simultaneously with the tone, indicating the correct response for the present task and the one that would follow. The trial ended with a 150 ms blank screen. At the beginning of each experimental block, written instructions for the upcoming task were displayed on the screen and remained until the space key was pressed. The response mapping was as follows: the *blue* response was assigned to the left index finger and the *yellow* response to the left middle finger. Similarly, the *square* response was assigned to the right index finger and the *circle* response to the right middle finger. The response keys for the color task were labeled with the appropriate colors, and the response keys for

TABLE 3 | Mean reaction time (RT) in milliseconds and error rates in the switch, repetition, and non-switch trials, and switch and mixing costs by experimental condition for monolinguals (n = 20) and bilinguals (n = 20).

Trial type	Task block	Monolinguals	Bilinguals
		Response latencies in ms	
Switch	Cued	1,475 (330)	1,288 (305)
	Memory	1,353 (322)	1,321 (328)
	Cued-Memory	123 (153)	−33 (158)
Repeat	Cued	1,327 (291)	1,158 (299)
	Memory	1,066 (234)	1,018 (250)
	Cued-Memory	260 (220)	140 (155)
Non-switch	Single task	921 (181)	787 (239)
		Error rates in %	
Switch	Cued	8.35 (5.5)	5.55 (4)
	Memory	6.25 (4)	4.05 (3)
Repeat	Cued	7.3 (5.5)	2.6 (2)
	Memory	5.05 (2)	3.5 (2)
Non-switch	Single task	1.15 (0)	1.45 (1)
		Switch and mixing costs	
Switch costs	Cued	148 (189)	130 (119)
	Memory	286 (179)	303 (155)
Mixing costs	Cued	406 (181)	371 (187)
	Memory	145 (173)	231 (151)

SDs for RTs and Medians for error rates are shown in parentheses.

the shape task were labeled with the appropriate shape. Before beginning the actual task, participants performed 16 practice trials of each condition. Data from these practice trials were not included in the analyses.

Data Analysis

RTs in color vs. shape judgments in single-task blocks did not differ significantly across participants ($t_{(39)} = -0.072$, $p = 0.943$, so we collapsed the data across the two conditions. For all reaction time (RT) analyses, only correct trials were included. Trials with response latencies below 200 ms and above 3,000 ms were excluded from the analysis. The RT-trimming procedure eliminated 2.28% and 2.93% of non-switch trials, 10.11% and 7.01% of repeat trials, and 12.55% and 8.26% of switch trials for monolinguals and bilinguals, respectively. In total, 7.19% of the trials were eliminated and were not included in the analysis. After data trimming, all distributions of response latencies showed acceptable levels of normality, homoscedasticity, and independence.

There were no negative associations between error rates and reaction times (RT) in any experimental condition, thus ruling out the possibility of a speed-accuracy trade-off. Error rates were analyzed using Mann–Whitney U tests. A significance level of $p < 0.05$ was adopted for all contrasts. Significance levels of multiple comparisons were Bonferroni-corrected to their number of comparisons. All the statistical analyses were conducted with SPSS v. 20.0 statistical software.

RESULTS

Table 3 presents a summary of the response latencies, error rates, and composite switch and mixing costs per experimental and linguistic condition, and **Figure 2** shows the response latencies by task version and trial type for monolinguals and bilinguals.

Switch Costs as a Function of Task Version

Shifting attention to a new task requires more cognitive resources than the repetition of the same task. Switch costs are defined as the difference in performance on switch trials as opposed to repeat trials, within the mixed-task blocks. In our study, mixed-task blocks were either memory-based (task switches occurred after every second trial without the appearance of a cue) or cue-based (task switches occurred in random order and were triggered by a pictorial cue). To analyze the effect of both types of task settings on switch costs, we conducted a 2 (Group: monolinguals and bilinguals) × 2 (Task type: cued vs. memory-based) × 2 (Trial type: switch vs. repeat) mixed ANOVA on TR as the dependent variable, with Group as a between-subjects factor and Task and Trial type as within-subjects factors. The main effect of Task type was significant ($F_{(1,38)} = 26.996$, MSE = 22,250.218, $p < 0.001$, $\eta_p^2 = 0.415$, $1 - \beta = 0.999$). Also, response latencies were larger on the switch than on repeat trials ($F_{(1,38)} = 101.077$, MSE = 18,637.808, $p < 0.001$, $\eta_p^2 = 0.727$, $1 - \beta = 1$, confirming that both task versions elicited switch costs for shifting attention. As indicated by a significant Task × Trial type interaction ($F_{(1,38)} = 30.334$, MSE = 7,045.376, $p < 0.001$, $\eta_p^2 = 0.444$, $1 - \beta = 1$), response latencies increased from the memory-based to the cued version. This was especially the case in repeat trials, leading to smaller switch costs in the cued condition. We found a significant Group × Task interaction ($F_{(1,38)} = 8.569$, MSE = 22,250.218, $p = 0.006$, $\eta_p^2 = 0.184$, $1 - \beta = 0.814$), suggesting that monolinguals and bilinguals adjusted in a different way to cued vs. memory-based task blocks. The magnitude of switch costs in both tasks was similar for monolinguals and bilinguals, as indicated by a non-significant main effect of Group ($p = 0.219$), and a non-significant three-way interaction Group × Trial × Task type ($p = 0.383$). To further investigate the significant Group × Task interaction, we performed Bonferroni corrected pairwise comparisons on the Group × Trial × Task interaction. Results revealed that, whereas monolinguals' RTs were significantly larger on cued switch trials when compared to memory-based switch trials (mean difference = 123 ms, $p = 0.001$), the performance of bilinguals did not differ on switch trials of both task versions (mean difference = −33 ms, $p = 0.35$). See **Figure 3**. On repeat trials, both groups showed a similar pattern, with higher RTs in the cued than in the memory-based condition (mean difference = 260 ms, $p = 0.01$ and 140 ms, $p = 0.01$ for monolinguals and bilinguals, respectively).

An analysis of the error rates confirmed that the task repetition was more demanding for monolinguals than for bilinguals in a setting of unpredictable cued task switches. Monolinguals committed significantly more errors than bilinguals on cued repeat trials [monolinguals: 7.3% bilinguals: 2.6% ($U = 109.5$, $z = -2.145$, $p = 0.012$)]. The performance of the two groups did not differ in accuracy in the remaining factor levels, and error rates were overall lower in the memory-based condition (repeat trials: 4.28%, $p = 0.665$; switch trials: 5.1%, $p = 0.455$; switch trials: 5.1%, $p = 0.455$) than in the cue-based condition (switch trials: 6.95%, $p = 0.494$).

In sum, these results suggest that, when rule changes were triggered by external cues, bilinguals switched more efficiently

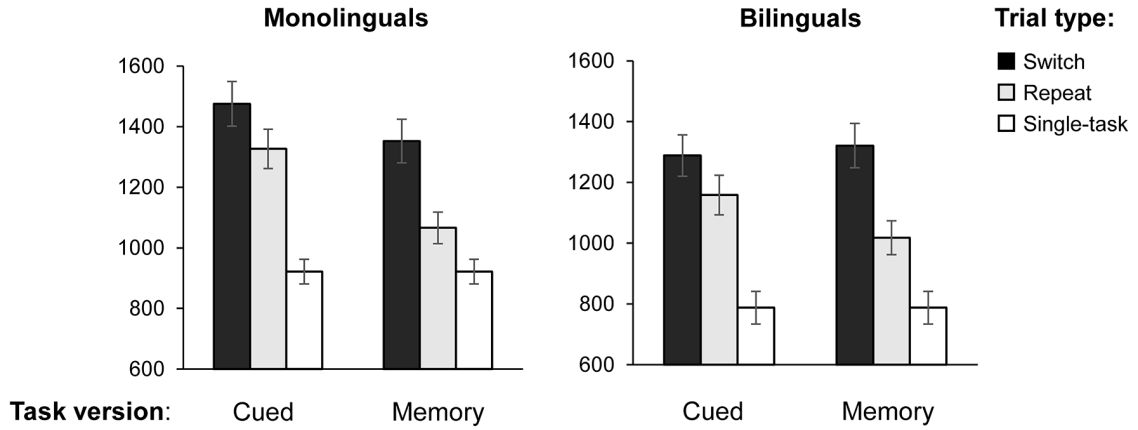


FIGURE 2 | Mean reaction times (RTs) on the switch, repeat, and non-switch trials by task version (cued, memory-based, and single-task) for monolinguals and bilinguals. Error bars: ±1 SE.

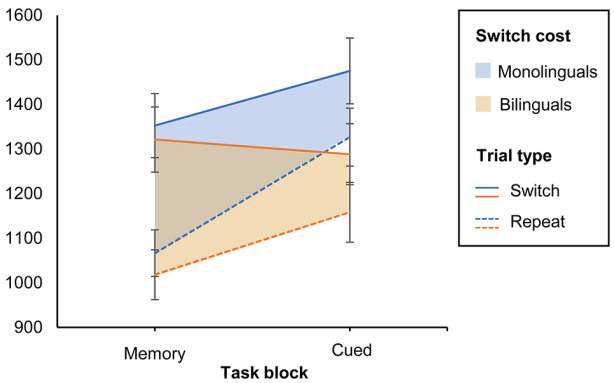


FIGURE 3 | Switch costs by task version for monolinguals and bilinguals. The continuous lines indicate switch trials, and the discontinuous lines indicate repeat trials. The shadowed areas represent switch costs (i.e., the difference between both trial types). Error bars: ±1 SE.

between task sets across trials than monolinguals. These findings are congruent with the previously discussed literature in that bilinguals may allocate their cognitive resources in a more parsimonious way when task demands increase.

Mixing Costs as a Function of Task Version

The repetition of a task rule in a context of set-shifting is always more effortful than performing the same task in a single-task context due to more complex task-set monitoring requirements (Monsell, 2003). This is what is indexed as "mixing costs" (i.e., the difference between repeat trials of a mixed task block and non-switch trials of a single-task block). To analyze the effect of single-task trials vs. repeat trials of both task versions, we conducted a 2 (Group: monolinguals and bilinguals) × 3 (Task type: single-task vs. memory-repeat trials vs. cued repeat trials) mixed ANOVA, with Group as a between-subjects factor and Trial type as a within-subjects factor. The main effect of Trial type was significant ($F_{(1,38)}$ = 94.618, MSE = 16.082, $p < 0.001$, $\eta_p^2 = 0.711$, $1 - \beta = 1$), indicating that the repetition of a trial in a mixed task block was overall more demanding than performing one task at a time. Neither the main effect of Group ($p = 0.116$), nor the Trial type × Group interaction resulted statistically significant ($p = 0.094$), suggesting that both groups produced similar mixing costs in both conditions. Bonferroni corrected pairwise comparison showed a trend for bilinguals being faster on single-task trials ($F_{(1,38)}$ = 3.982, $p < 0.052$, $\eta_p^2 = 0.095$, $1 - \beta = 0.494$) and on cued repeat trials ($F_{(1,38)}$ = 3.271, $p < 0.078$, $\eta_p^2 = 0.079$, $1 - \beta = 0.422$) whereas, as mentioned earlier, the performance on memory-based repeat trials was similar for both groups ($p < 0.534$).

To compare the magnitude of mixing costs as a function of task version, we ran an additional ANOVA, with Group as a between-subjects factor and Mixing cost (memory-based vs. cued) as within-subjects factors. The main factor of Mixing cost was significant ($F_{(1,38)}$ = 44.353, MSE = 18,066.958, $p < 0.001$, $\eta_p^2 = 0.539$, $1 - \beta = 1$), confirming that Mixing costs were overall higher in the cued condition (406 ms and 371 ms) than in the memory-based condition (145 ms and 231 ms, for monolinguals and bilinguals, respectively). A marginally significant Group × Mixing cost interaction ($F_{(1,38)}$ = 4.028, MSE = 18,066.958, $p = 0.052$, $\eta_p^2 = 0.096$, $1 - \beta = 0.498$) suggested that monolinguals experienced a larger increase in composite mixing costs from the cued to the memory-based task version (261 ms increase for monolinguals and 140 ms increase for bilinguals). Altogether, it seemed that both groups experienced an increase in the magnitude of mixing costs when task switches were unpredictable and externally cued and that this increase was slightly larger for monolinguals.

DISCUSSION

The results of the present study suggest that bilinguals shift their attention more efficiently than monolinguals when the task requirements mimic context-related dual-language management (i.e., aleatory and externally triggered task switches). The difference in response latencies between cued and memory-based switch trials was significantly larger in monolinguals

than in bilinguals. The performance of bilinguals did not differ across task versions, whereas monolinguals experienced a pronounced increase in response latencies when set-shifting was unpredictable and triggered by an external cue. Task performance also differed in terms of accuracy, as monolinguals had a significantly higher error rate than bilinguals on cued repeat trials, suggesting that it was overall more effortful for them to shift attention under unpredictable task-switching conditions than it was for bilinguals. However, the magnitude of composite switch and mixing costs was similar for monolinguals and bilinguals, suggesting that composite scores might not sufficiently capture fine-grained differences in performance.

To compare task-switching abilities under different cognitive demands, in the present study we adapted a task-switching paradigm that contained both memory-based and cued task-switching blocks. This procedure served to tax slightly different underlying control mechanisms. The memory-based task-switching paradigm involves predictable sequences of rule changes and requires primarily the monitoring of information in working memory. By contrast, cued task-switching, like language-switching, additionally requires context-dependent attentional reorientation and increased cognitive control demands. Thus, we predicted that a bilingual advantage would only be found when set shifting was triggered externally. The results of this pilot study confirmed only partially this hypothesis. Monolinguals and bilinguals did not differ significantly in response latencies within each task version, but significant group differences were found in the dynamics between the two versions. The two groups performed almost identically in the memory-based switch task; hence this variable could be taken as baseline performance. Contrary to monolinguals, whose performance decreased, bilinguals maintained the same performance in the cued condition. Bilinguals had lower response latencies on cued switch trials and lower error rates on cued repeat trials, suggesting a bilingual advantage in the flexible adjustment to task-relevant context processing. These results are congruent with the existing literature regarding the similarity between cued task switching and linguistic code-switching (Christoffels et al., 2007; Prior and Gollan, 2011). Bilinguals might be more trained in efficiently interpreting contextual requirements to flexibly adjust their behavior. Previous research has shown that explicit cueing in a set of random-switching facilitates the task-set reconfiguration when enough time is given to prepare for the next trial (Tornay and Milán, 2001). Our experimental design included a cue-target interval of 800 ms, thus providing enough time for task preparation. Differences in efficient preparatory task-set activation are related primarily to individual differences in cognitive control, whereas age-related changes mainly appear to affect target response selection and task performance in general (Adrover-Roig and Barceló, 2010). In this line, our results suggest that cognitive aging affects the working-memory processes of monolinguals and bilinguals similarly, but that bilinguals might use contextual cues more efficiently and start the task-set reconfiguration earlier than monolinguals.

Our results also suggest that a long period of second-language immersion might parallel the cognitive benefits produced by an early age of acquisition. In our study, late bilinguals had been immersed in their second-language environment for more than 40 years on average and were highly balanced. However, dual-language exposure alone does not seem enough to modulate cognitive control. The balance in language use has been widely discussed as a core factor to explain the bilingual advantage (Verreyt et al., 2016; Yang et al., 2016; Hartanto and Yang, 2020). Even in balanced bilinguals, only high-frequency language switchers showed an advantage over monolinguals in tasks that measure cognitive flexibility (Barbu et al., 2020).

Long-time balanced dual-language immersion might lead to changes related to a more efficient reorientation to stimuli-driven task demands. As mentioned earlier, memory-based task switching requires more implication of WM sustained by an interaction of frontoparietal areas that are very sensitive to aging. Previous research has shown that, contrary to the so-called age-related posterior-anterior shift (PASA; Davis et al., 2008), this shift is reversed in some bilinguals to more subcortical/posterior regions during the performance of executive function tasks (Rodríguez-Pujadas et al., 2013; Grundy et al., 2017). Context-dependent reorientation (as in cued task-switching) relies on the interaction of frontostriatal loops with the special implication of the basal ganglia (Shulman et al., 2009; Van Schouwenburg et al., 2010). Several authors have proposed that at the initial stages of bilingualism, language control is mostly managed by prefrontal areas (Ullman, 2001; Stocco et al., 2014). Then, as dual-language management becomes more automatic, its neural processing shifts partly to subcortical areas (Lieberman, 2000; Tettamanti et al., 2005) as occurs in procedural knowledge (Packard and Knowlton, 2002). Bilinguals show expanded morphology in basal ganglia (Burgaleta et al., 2016). Damage to this brain area produces pathologic code-switching (Lieberman, 2000; Abutalebi and Green, 2008) similarly as it affects task-switching abilities in early Parkinson disease patients (Packard and Knowlton, 2002). Neuroimaging findings suggest that age-related changes in prefrontal areas affect bilinguals to a similar degree as monolinguals. However, bilinguals instead of overrecruiting those areas rely more on subcortical areas developed by life-long dual-language management. Our behavioral results fit with the current knowledge on bilingual neural processing and suggest that in older adults, processes that rely heavily on WM are affected similarly in monolinguals and bilinguals, but that bilingualism might improve processes that require a flexible reorientation to environmental cues.

Bilingualism is just one of the many components that might contribute to cognitive reserve. Numerous other factors and lifestyle habits can counteract their hypothetical benefits. Also, findings are heavily influenced by study design, and while retrospective studies tend to a protective effect of bilingualism on cognition, prospective studies often fail to find differences between monolinguals and bilinguals (Paap et al., 2016; Watson et al., 2016). The best alternative to investigate the effect of bilingualism on aging is to conduct powered randomized controlled trials that enable adequate control of baseline characteristics, psychological assessment, and experimental manipulations. To date, there are no results from

such studies, but several promising study protocols, especially on the effect of foreign language learning in older adults, have recently been registered, and we can thus hope to obtain more insight into these important research questions in the near future.

LIMITATIONS AND FUTURE DIRECTIONS

A limitation of the present study is the small sample size. Possible differences between monolinguals and bilinguals, especially in composite switch and mixing costs, could be missed due to low statistical power. Small samples also increase the risk of type I errors, and the statistically significant interaction effect found in switch trials across conditions would need replication. However, the present study provides an innovative approach, contributing to the ongoing debate on the reliability of a bilingual advantage and prepares the ground for a larger-scale investigation, focusing not only on bilingual balance and language use but also on specific task characteristics.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Universidad Nacional de Educación a Distancia (UNED). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JAR and SB: conceptualization and study design. JAR: enrolled the participants, collected, and analyzed the data. JAR and JMR: data analysis. All the authors: interpretation and final approval. JAR with support from SB: manuscript preparation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Spanish Ministry of Economy and Competitiveness (grant # PSI2016-80377-R) to SB and JMR, and by an FPI Fellowship from the Spanish Ministry of Economy and Competitiveness (grant # BES-2017-079760) to JAR. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

REFERENCES

Abutalebi, J., and Green, D. (2007). Bilingual language production: the neurocognition of language representation and control. *J. Neurolinguistics* 20, 242–275. doi: 10.1016/j.jneuroling.2006.10.003

Abutalebi, J., and Green, D. W. (2008). Control mechanisms in bilingual language production: neural evidence from language switching studies. *Lang. Cogn. Process.* 23, 557–582. doi: 10.1080/01690960801920602

Adesope, O. O., Lavin, T., Thompson, T., and Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Educ. Res. Rev.* 80, 207–245. doi: 10.3102/0034654310368803

Adrover-Roig, D., and Barceló, F. (2010). Individual differences in aging and cognitive control modulate the neural indexes of context updating and maintenance during task-switching. *Cortex* 46, 434–450. doi: 10.1016/j.cortex.2009.09.012

Ansaldo, A. I., Ghazi-Saidi, L., and Adrover-Roig, D. (2015). Interference control in elderly bilinguals: appearances can be misleading. *Clin. Exp. Neuropsychol.* 37, 455–470. doi: 10.1080/13803395.2014.990359

Antón, E., Fernández García, Y., Carreiras, M., and Duñabeitia, J. A. (2016). Does bilingualism shape inhibitory control in the elderly? *J. Mem. Lang.* 90, 147–160. doi: 10.1016/j.jml.2016.04.007

Antoniou, M. (2019). The advantages of bilingualism debate. *Annu. Rev. Linguist.* 5, 395–415. doi: 10.1146/annurev-linguistics-011718-011820

Baddeley, A. D., and Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology* 8, 485–493. doi: 10.1037/0894-4105.8.4.485

Baene, W. D., Duyck, W., Brass, M., and Carreiras, M. (2015). Brain circuit for cognitive control is shared by task and language switching. *J. Cogn. Neurosci.* 27, 1752–1765. doi: 10.1162/jocn_a_00817

Ballesteros, S., Voelcker-Rehage, C., and Bherer, L. (2018). Editorial: cognitive and brain plasticity induced by physical exercise, cognitive training, video games, and combined interventions. *Front. Hum. Neurosci.* 12:169. doi: 10.3389/fnhum.2018.00169

Barac, R., Bialystok, E., Castro, D. C., and Sanchez, M. (2014). The cognitive development of young dual language learners: a critical review. *Early Child. Res. Q.* 29, 699–714. doi: 10.1016/j.ecresq.2014.02.003

Barbu, C.-A., Gillet, S., and Poncelet, M. (2020). Investigating the effects of language-switching frequency on attentional and executive functioning in proficient bilinguals. *Front. Psychol.* 11:1078. doi: 10.3389/fpsyg.2020.01078

Barulli, D., and Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. *Trends Cogn. Sci.* 17, 502–509. doi: 10.1016/j.tics.2013.08.012

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Mendes de Leon, C. F., Arnold, S. E., et al. (2003). Education modifies the relation of AD pathology to level of cognitive function in older persons. *Neurology* 60, 1909–1915. doi: 10.1212/01.wnl.0000069923.64550.9f

Bialystok, E., Craik, F. I., and Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends Cogn. Sci.* 16, 240–250. doi: 10.1016/j.tics.2012.03.001

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Can. J. Exp. Psychol.* 60, 68–79. doi: 10.1037/cjep2006008

Bialystok, E. (2017). The bilingual adaptation: how minds accommodate experience. *Psychol. Bull.* 143, 233–262. doi: 10.1037/bul0000099

Bialystok, E., Craik, F. I. M., and Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia* 45, 459–464. doi: 10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I. M., Klein, R., and Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: evidence from the Simon task. *Psychol. Aging* 19, 290–303. doi: 10.1037/0882-7974.19.2.290

Birdsong, D., Gertken, L. M., and Amengual, M. (2012). Bilingual language profile: an easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin.* Available online at: https://sites.la.utexas.edu/bilingual/. Accessed May 15, 2018.

Blanco-Elorrieta, E., and Pylkkänen, L. (2018). Ecological validity in bilingualism research and the bilingual advantage. *Trends Cogn. Sci.* 22, 1117–1126. doi: 10.1016/j.tics.2018.10.001

Burgaleta, M., Sanjuán, A., Ventura-Campos, N., Sebastian-Galles, N., and Ávila, C. (2016). Bilingualism at the core of the brain. Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage* 125, 437–445. doi: 10.1016/j.neuroimage.2015.09.073

Calvo, N., García, A. M., Manoiloff, L., and Ibáñez, A. (2016). Bilingualism and cognitive reserve: a critical overview and a plea for methodological innovations. *Front. Aging Neurosci.* 7:249. doi: 10.3389/fnagi.2015.00249

Carlson, S. M., and Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Dev. Sci.* 11, 282–298. doi: 10.1111/j.1467-7687.2008.00675.x

Casey, B. J., Giedd, J. N., and Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biol. Psychol.* 54, 241–257. doi: 10.1016/s0301-0511(00)00058-2

Christoffels, I. K., Firk, C., and Schiller, N. O. (2007). Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208. doi: 10.1016/j.brainres.2007.01.137

Coderre, E. L., Smith, J. F., van Heuven, W. J. B., and Horwitz, B. (2016). The functional overlap of executive control and language processing in bilinguals. *Biling* 19, 471–488. doi: 10.1017/S1366728915000188

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi: 10.1016/j.neuron.2008.04.017

Costa, A., Hernández, M., and Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: evidence from the ANT task. *Cognition* 106, 59–86. doi: 10.1016/j.cognition.2006.12.013

Costa, A., Hernández, M., Costa-Faidella, J., and Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: now you see it, now you don't. *Cognition* 113, 135–149. doi: 10.1016/j.cognition.2009.08.001

Craik, F. I. M., Bialystok, E., and Freedman, M. (2010). Delaying the onset of Alzheimer disease: bilingualism as a form of cognitive reserve. *Neurology* 75, 1726–1729. doi: 10.1212/WNL.0b013e3181fc2a1c

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., et al. (2006). Language control in the bilingual brain. *Science* 312, 1537–1540. doi: 10.1126/science.1127761

Davis, S. W., Dennis, N. A., Daselaar, S. M., Fleck, M. S., Cabeza, R. (2008). Qué PASA? The posterior–anterior shift in aging. *Cereb. Cortex* 18, 1201–1209. doi: 10.1093/cercor/bhm155

de Bruin, A., Bak, T. H., and Della Sala, S. (2015). Examining the effects of active vs. inactive bilingualism on executive control in a carefully matched non-immigrant sample. *J. Mem. Lang.* 85, 15–26. doi: 10.1016/j.jml.2015.07.001

Donnelly, S., Brooks, P., and Homer, B. (2019). Is there a bilingual advantage on interference-control tasks? A multiverse meta-analysis of global reaction time and interference cost. *Psychon. Bull. Rev.* 26, 1122–1147. doi: 10.3758/s13423-019-01567-z

Ferreira, N., Owen, A., Mohan, A., Corbett, A., and Ballard, C. (2015). Associations between cognitively stimulating leisure activities, cognitive function and age-related cognitive decline. *Int. J. Geriatr. Psychiatry* 30, 422–430. doi: 10.1002/gps.4155

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Gavrilov, L. A., and Heuveline, P. (2003). "Aging of population," in *The Encyclopedia of Population*, eds P. Demeny and G. McNicoll (New York, NY: Macmillan Reference), 32–37.

Gertken, L. M., Amengual, M., and Birdsong, D. (2014). "Assessing language dominance with the bilingual language profile," in *Measuring l2 Proficiency: Perspectives from SLA*, eds P. Leclercq, A. Edmonds, and H. Hilton (Bristol: Multilingual Matters), 208–225.

Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., and Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *J. Neurosci.* 33, 387–396. doi: 10.1523/JNEUROSCI.3837-12.2013

Goral, M., Campanelli, L., and Spiro, A. (2015). Language dominance and inhibition abilities in bilingual older adults. *Biling* 18, 79–89. doi: 10.1017/S1366728913000126

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Biling: Lang. Cogn.* 1, 67–81. doi: 10.1017/S1366728998000133

Green, D. W., and Abutalebi, J. (2013). Language control in bilinguals: the adaptive control hypothesis. *J. Cogn. Psychol.* 25, 515–530. doi: 10.1080/20445911.2013.796377

Green, D. W., and Wei, L. (2014). A control process model of code-switching. *Lang. Cogn. Neurosci.* 29, 499–511. doi: 10.1080/23273798.2014.882515

Grundy, J. G., Anderson, J. A., and Bialystok, E. (2017). Neural correlates of cognitive processing in monolinguals and bilinguals. *Ann. N. Y. Acad. Sci.* 1396, 183–201. doi: 10.1111/nyas.13333

Hartanto, A., and Yang, H. (2016). Disparate bilingual experiences modulate task-switching advantages: a diffusion-model analysis of the effects of interactional context on switch costs. *Cognition* 150, 10–19. doi: 10.1016/j.cognition.2016.01.016

Hartanto, A., and Yang, H. (2020). The role of bilingual interactional contexts in predicting interindividual variability in executive functions: a latent variable analysis. *J. Exp. Psychol.* 149, 609–634. doi: 10.1037/xge0000672

Hernández, M., Martin, C. D., Barceló, F., and Costa, A. (2013). Where is the bilingual advantage in task-switching? *J. Mem. Lang.* 69, 257–276. doi: 10.1016/j.jml.2013.06.004

Houtzager, N., Lowie, W., Sprenger, S., and de Bot, C. (2015). A bilingual advantage in task-switching? Age-related differences between German monolinguals and Dutch/Frisian bilinguals. *Biling: Lang. Cogn* 20, 69–79. doi: 10.1017/S1366728915000498

Huff, M. J., Balota, D. A., Minear, M., Aschenbrenner, A. J., and Duchek, J. M. (2015). Dissociative global and local task-switching costs across younger adults, middle-aged adults, older adults, and very mild Alzheimer's disease individuals. *Psychol. Aging* 30, 727–739. doi: 10.1037/pag0000057

Hurd, M. D., Martorell, P., Langa, K. M., Delavande, A., and Mullen, K. J. (2013). Monetary costs of dementia in the United States. *N. Engl. J. Med.* 368, 1326–1334. doi: 10.1056/NEJMsa1204629

Kaamaa, P., Szewczyk, J., Chuderski, A., Senderecka, M., and Wodniecka, Z. (2020). Patterns of bilingual language use and response inhibition: a test of the adaptive control hypothesis. *Cognition* 204:104373. doi: 10.1016/j.cognition.2020.104373

Kapa, L. L., and Colombo, J. (2013). Attentional control in early and later bilingual children. *Cogn. Dev.* 28, 233–246. doi: 10.1016/j.cogdev.2013.01.011

Kray, J. (2006). Task-set switching under cue-based versus memory-based switching conditions in younger and older adults. *Brain Res.* 1105, 83–92. doi: 10.1016/j.brainres.2005.11.016

Kray, J., and Lindenberger, U. (2000). Adult age differences in task-switching. *Psychol. Aging* 15, 126–147. doi: 10.1037/0882-7974.15.1.126

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., and Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychol. Bull.* 144, 394–425. doi: 10.1037/bul0000142

Lieberman, P. (2000). *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax and Thought*. Cambridge: Harvard University Press.

Luk, G., Green, D., Abutalebi, J., and Grady, C. (2011). Cognitive control for language switching in bilinguals: a quantitative meta-analysis of functional neuroimaging studies. *Lang. Cogn. Process.* 27, 1479–1488. doi: 10.1080/01690965.2011.613209

Martínez de la Iglesia, J., Onís Vilches, M. C., Dueñas Herrero, R., Albert Colomer, C., Aguado Taberné, C., and Luque Luque, R. (2002). The Spanish version of the Yesavage abbreviated questionnaire (GDS) to screen depressive dysfunctions in patients older than 65 years. *Medifam* 12, 620–630.

Meuter, R. F. I., and Allport, A. (1999). Bilingual language switching in naming: asymmetrical costs of language selection. *J. Mem. Lang.* 40, 25–40. doi: 10.1006/jmla.1998.2602

Monchi, O., Petrides, M., Petre, V., Worsley, K., and Dagher, A. (2001). Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *J. Neurosci.* 21, 7733–7741. doi: 10.1523/JNEUROSCI.21-19-07733.2001

Monsell, S. (2003). Task-switching. *Trends Cogn. Sci.* 7, 134–140. doi: 10.1016/s1364-6613(03)00028-7

Monsell, S., Sumner, P., and Waters, H. (2003). Task-set reconfiguration with predictable and unpredictable task switches. *Mem. Cogn.* 31, 327–342. doi: 10.3758/bf03194391

Osorio, A., Fay, S., Pouthas, V., and Ballesteros, S. (2010). Aging affects brain activity in highly educated older adults: an ERP study using a word-stem priming task. *Cortex* 46, 522–534. doi: 10.1016/j.cortex.2009.09.003

Paap, K. R., and Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cogn. Psychol.* 66:232–258. doi: 10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., and Sawi, O. (2016). Should the search for bilingual advantages in executive functioning continue? *Cortex* 74, 305–314. doi: 10.1016/j.cortex.2015.09.010

Paap, K. R., Mason, L., Zimiga, B., Ayala-Silva, Y., and Frost, M. (2020). The alchemy of confirmation bias transmutes expectations into bilingual advantages: a tale of two new meta-analyses. *Q. J. Exp. Psychol.* 73, 1290–1299. doi: 10.1177/1747021819900098

Packard, M. G., and Knowlton, B. J. (2002). Learning and memory functions of the Basal Ganglia. *Annu. Rev. Neurosci.* 25, 563–593. doi: 10.1146/annurev.neuro.25.112701.142937

Park, D. C., and Reuter-Lorenz, P. A. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annu. Rev. Psychol.* 60, 173–196. doi: 10.1146/annurev.psych.59.103006.093656

Pliatsikas, C., DeLuca, V., Moschopoulou, E., and Saddy, J. D. (2016). Immersive bilingualism reshapes the core of the brain. *Brain Struct. Funct.* 222, 1785–1795. doi: 10.1007/s00429-016-1307-9

Pot, A., Keijzer, M., and de Bot, K. (2018). Intensity of multilingual language use predicts cognitive performance in some multilingual older adults. *Brain Sci.* 8:92. doi: 10.3390/brainsci8050092

Prior, A., and Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *J. Int. Neuropsychol. Soc.* 17, 682–691. doi: 10.1017/S1355617711000580

Qu, L., Low, J. J. W., Zhang, T., Li, H., and Zelazo, P. D. (2015). Bilingual advantage in executive control when task demands are considered. *Biling: Lang. Cogn* 19, 277–293. doi: 10.1017/S1366728914000376

Reimers, S., and Maylor, E. A. (2005). Task switching across the life span: effects of age on general and specific switch costs. *Dev. Psychol.* 41, 661–671. doi: 10.1037/0012-1649.41.4.661

Reuter-Lorenz, P. A., and Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychol. Rev.* 24, 355–370. doi: 10.1007/s11065-014-9270-9

Riley, K. P., Snowdon, D. A., and Markesbery, W. R. (2002). Alzheimer's neurofibrillary pathology and the spectrum of cognitive function: Findings from the nun study. *Ann. Neurol.* 51, 567–577. doi: 10.1002/ana.10161

Rodríguez-Pujadas, A., Sanjuán, A., Ventura-Campos, N., Román, P., Martin, C., Barceló, F., et al. (2013). Bilinguals use language-control brain areas more than monolinguals to perform non-linguistic switching tasks. *PLoS One* 8:e73028. doi: 10.1371/journal.pone.0073028

Rubin, O., and Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 1477–1491. doi: 10.1037/0278-7393.31.6.1477

Runnqvist, E., Fitzpatrick, I., Strijkers, K., and Costa, A. (2012). "An appraisal of the bilingual language production system: quantitatively or qualitatively different from monolinguals?" in *The Handbook of Bilingualism and Multilingualism*, 2nd Edn, eds T. K. Bhatia and W. C. Ritchie (Chichester, UK: John Wiley and Sons, Ltd.), 244–265. doi: 10.1002/9781118332382.ch10

Salvatierra, J. L., and Rosselli, M. (2010). The effect of bilingualism and age on inhibitory control. *Int. J. Biling* 15, 26–37. doi: 10.1177/1367006910371021

Scaltritti, M., Peressotti, F., and Miozzo, M. (2017). Bilingual advantage and language switch: what's the linkage? *Biling: Lang. Cogn.* 20:80. doi: 10.1017/S1366728915000565

Shulman, G. L., Astafiev, S. V., Franke, D., Pope, D. L. W., Snyder, A. Z., McAvoy, M. P., et al. (2009). Interaction of stimulus-driven reorienting and expectation in ventral and dorsal frontoparietal and basal ganglia-cortical networks. *J. Neurosci.* 29, 4392–4407. doi: 10.1523/JNEUROSCI.5609-08.2009

Soveri, A., Rodriguez-Fornells, A., and Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Front. Psychol.* 2:183. doi: 10.3389/fpsyg.2011.00183

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia* 47, 2015–2028. doi: 10.1016/j.neuropsychologia.2009.03.004

Stocco, A., Lebiere, C., and Anderson, J. R. (2010). Conditional routing of information to the cortex: a model of the basal ganglia's role in cognitive coordination. *Psychol. Rev.* 117, 541–574. doi: 10.1037/a0019077

Stocco, A., Yamasaki, B., Natalenko, R., and Prat, C. S. (2014). Bilingual brain training: a neurobiological framework of how bilingual experience improves executive function. *Int. J. Biling* 18, 67–92. doi: 10.1177/1367006912456617

Terry, C., and Sliwinski, M. (2012). Aging and random task switching: the role of endogenous versus exogenous task selection. *Exp. Aging Res.* 38, 87–109. doi: 10.1080/0361073X.2012.637008

Tettamanti, M., Moro, A., Messa, C., Moresco, R. M., Rizzo, G., Carpinelli, A., et al. (2005). Basal ganglia and language: phonology modulates dopaminergic release. *Neuroreport* 16, 397–401. doi: 10.1097/00001756-200503150-00018

Tornay, F. J., and Milán, E. G. (2001). A more complete task-set reconfiguration in random than in predictable task switch. *Q. J. Exp. Psychol.* 54, 785–803. doi: 10.1080/713755984

Ullman, M. T. (2001). A neurocognitive perspective on language: the declarative/procedural model. *Nat. Rev. Neurosci.* 2, 717–726. doi: 10.1038/35094573

Van den Noort, M., Vermeire, K., Bosch, P., Staudte, H., Krajenbrink, T., Jaswetz, L., et al. (2019). A systematic review on the possible relationship between bilingualism, cognitive decline, and the onset of dementia. *Behav. Sci.* 9:81. doi: 10.3390/bs9070081

Van Schouwenburg, M. R., den Ouden, H. E. M., and Cools, R. (2010). The human basal ganglia modulate frontal-posterior connectivity during attention shifting. *J. Neurosci.* 30, 9910–9918. doi: 10.1523/JNEUROSCI.1111-10.2010

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., and Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Biling: Lang. Cogn.* 19, 181–190. doi: 10.1017/S1366728914000352

Wasylyshyn, C., Verhaeghen, P., and Sliwinski, M. J. (2011). Aging and task switching: a meta-analysis. *Psychol. Aging* 26, 15–20. doi: 10.1037/a0020912

Watson, C. W.-M., Manly, J. J., and Zahodne, L. B. (2016). Does bilingualism protect against cognitive aging? Methodological issues in research on bilingualism, cognitive reserve, and dementia incidence. *Linguist. Approaches Biling.* 6, 590–604. doi: 10.1075/lab.15043.wat

World Health Organization. (2012). *Dementia: A Public Health Priority*. Available online at: https://www.who.int/mental_health/publications/dementia_report_2012/en/. Accessed October 15, 2018.

Woumans, E. V. Y., Santens, P., Sieben, A., Versijpt, J. A. N., Stevens, M., and Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer's disease. *Biling: Lang. Cogn* 18, 568–574. doi: 10.1017/S136672891400087X

Yang, H., Hartanto, A., and Yang, S. (2016). The complex nature of bilinguals' language usage modulates task-switching outcomes. *Front. Psychol.* 7:560. doi: 10.3389/fpsyg.2016.00560

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* 17, 37–49. doi: 10.1016/0022-3956(82)90033-4

Zelazo, P. D., Craik, F. I. M., and Booth, L. (2004). Executive function across the life span. *Acta Psychol.* 115, 167–183. doi: 10.1016/j.actpsy.2003.12.005

Zhang, H., Wu, Y. J., and Thierry, G. (2020). Bilingualism and aging: a focused neuroscientific review. *J. Neurolinguistics* 54, 100890. doi: 10.1016/j.jneuroling.2020.100890

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.

frontiers in Human Neuroscience

STUDY PROTOCOL
published: 20 January 2021
doi: 10.3389/fnhum.2020.611691

Development and Health of Adults Formerly Placed in Infant Care Institutions – Study Protocol of the LifeStories Project

Patricia Lannen[1], Hannah Sand[1], Fabio Sticca[1], Ivan Ruiz Gallego[1], Clara Bombach[1], Heidi Simoni[1], Flavia M. Wehrle[2] and Oskar G. Jenni[2]*

[1] Marie Meierhofer Children's Institute, University of Zurich, Zurich, Switzerland, [2] Child Development Center, University Children's Hospital Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Leonhard Schilbach, Ludwig Maximilian University of Munich, Germany

Reviewed by:
Mike Martin, University of Zurich, Switzerland
Ayşe İkinci Keleş, Niğde Ömer Halisdemir University, Turkey

***Correspondence:**
Patricia Lannen
lannen@mmi.ch

Specialty section:
This article was submitted to Health, a section of the journal Frontiers in Human Neuroscience

Received: *29 September 2020*
Accepted: *03 December 2020*
Published: *20 January 2021*

Citation:
Lannen P, Sand H, Sticca F, Ruiz Gallego I, Bombach C, Simoni H, Wehrle FM and Jenni OG (2021) Development and Health of Adults Formerly Placed in Infant Care Institutions – Study Protocol of the LifeStories Project. Front. Hum. Neurosci. 14:611691. doi: 10.3389/fnhum.2020.611691

A growing volume of research from global data demonstrates that institutional care under conditions of deprivation is profoundly damaging to children, particularly during the critical early years of development. However, how these individuals develop over a life course remains unclear. This study uses data from a survey on the health and development of 420 children mostly under the age of three, placed in 12 infant care institutions between 1958 and 1961 in Zurich, Switzerland. The children exhibited significant delays in cognitive, social, and motor development in the first years of life. Moreover, a follow-up of a subsample of 143 children about 10 years later revealed persistent difficulties, including depression, school related-problems, and stereotypies. Between 2019 and 2021, these formerly institutionalized study participants were located through the Swiss population registry and invited to participate once again in the research project. Now in their early sixties, they are studied for their health, further development, and life-course trajectories. A mixed-methods approach using questionnaires, neuropsychological assessments, and narrative biographical interviews was implemented by a multidisciplinary team. Combining prospective and retrospective data with standardized quantitative and biographical qualitative data allows a rich reconstruction of life histories. The availability of a community sample from the same geographic location, the 1954–1961 cohort of the Zurich Longitudinal Studies, described in detail in a paper in this issue (Wehrle et al., 2020), enables comparison with an unaffected cohort. This article describes the study design and study participants in detail and discusses the potential and limitations of a comparison with a community sample. It outlines a set of challenges and solutions encountered in the process of a lifespan longitudinal study from early childhood into the cusp of old age with a potentially vulnerable sample and summarizes the lessons learned along the way.

Keywords: longitudinal study, lifespan development, institutional care, adverse childhood experiences, early childhood, child development, compulsory social measures and placements

INTRODUCTION

In Switzerland, placing infants in institutions was not uncommon in the first half of the 20th century (Ryffel, 2013). The main reasons for having a child placed in an institution were either being an unmarried or underaged mother or having a migrant background, foremost Italian migrant worker status (German: *Gastarbeiter*) (Meierhofer and Keller, 1974). Having a child as a young, unmarried mother was, from the point of view of the authorities and society, slovenly (liederlich) and was to be "disciplined" (Ramsauer, 2000; Lengwiler and Praz, 2018; Unabhängige Expertenkommission Administrative Versorgungen, 2019). Migrant workers were subjected to serious prejudice and residence permit restrictions and were forced to work full time with long working hours to stay in Switzerland (D'Amato, 2012; Joris, 2012). Generally, infants were placed into institutions at a very young age, before the age of two weeks, due to the lack of paid maternity leave (Huber, 1995). At that time, the infant was seen as a simple reflex-driven being (Meierhofer, 1958), and a belief was prevalent that "there will be no harm to infants if they are cared for by strangers" (Meierhofer and Keller, 1974). Hard-earned success in reducing child mortality had made preventing the spread of germs a priority, so an institutional practice of "isolation" was the norm, involving as little physical contact as possible, feeding according to a rigid plan, and strict hygiene (Ryffel, 2013). Care practices were generally characterized by strict routines that did not take into consideration individual needs and an intense wariness of spoiling children (Gebhardt, 2009). This created the conditions of chronic deprivation found to be responsible for the profound negative effects on development described in more recent work (Nelson et al., 2014; Berens and Nelson, 2015).

A growing volume of international research shows that children placed in institutional care as it is commonly implemented are typically deprived of a supportive, intensive, one-to-one relationship with a primary caregiver. Such conditions of deprivation are profoundly damaging to children, particularly during the critical early years of development (Schore, 2001). Children who were placed into institutions shortly after birth and subjected to severe sensory, emotional, and sometimes physical neglect show a dramatic decrease in brain activity compared to children who were never institutionalized (Center on the Developing Child, 2007). They are more likely to suffer from growth delay, frequent infections, and hearing and vision problems. Furthermore, motor development is often delayed and stereotypical behaviors such as body rocking and head banging occur (Browne, 2009; Berens and Nelson, 2015; Sherr et al., 2017). Finally, a number of studies have found negative effects on cognitive and social development (Schore, 2001; Johnson et al., 2006). Although no data is currently available on the effects of institutionalization across the entire lifespan, a number of large studies have demonstrated the potential long-term negative effects of other adverse events and circumstances during childhood, such as child abuse and neglect, household substance abuse, and mental illness, on development and health into adulthood (Felitti et al., 1998; Werner, 2013). The impact of such events has also been shown to be especially severe during the critical developmental stages in early childhood, with possible permanent effects on the morphology of the brain (Shonkoff and Phillips, 2000).

However, not all individuals experience negative health outcomes in relation to stress. In his seminal work, Antonovsky (1979) coined the term "salutogenesis" after observing "how people manage stress and stay well." Similarly, findings from a 40-year longitudinal study by Werner (2013) showed that one third of all high-risk children exposed to adverse experiences early on displayed resilience that allowed them to develop into caring, competent, and confident adults. Particular protective factors and biographical events helped to balance out risk factors at critical periods in their development in a dynamic interaction of personal factors, environmental factors, and biographical developments. Further, evidence has started to emerge of what mitigates the negative impact of institutionalization, such as age at entry, stability of care, size of institution and children/staff ratio (Berens and Nelson, 2015; Sherr et al., 2017). Care circumstances and practices with children in institutions can vary greatly, with some of the most severe conditions of deprivation studied as part of the Bucharest Early Intervention Study with children in Romania (Nelson et al., 2014). This study was also able to show that when children are moved from institutional care into family based care, they have a chance of restoring brain development, highlighting the plasticity of the developing brain (Nelson et al., 2014; Bick et al., 2015).

The potential impact of this early institutionalization across the lifespan is unknown. Furthermore, whereas data is available on the effects of exposure to adversity in early childhood into early and mid-adulthood, the effects of adverse experiences into late adulthood remain as yet virtually unexplored. With a unique combination of historical and newly collected 60-year long-term follow-up data, the overall aim of the LifeStories project (German: *Lebensgeschichten*) is to examine the personal developmental trajectories of individuals that were affected by placement in institutions as infants in the late 1950s and early 1960s in Switzerland. The study builds on the pioneering work of Dr. Marie Meierhofer on the delayed development of infants placed in institutions for care at the end of the 1950s (Meierhofer and Keller, 1974). By simultaneously using data from an unaffected comparison sample of children growing up in families, the 1954–1961 cohort of the Zurich Longitudinal Studies of the University Children's Hospital Zurich (ZLS, $N = 445$) — one of the most significant data sets on child development globally (Ulijaszek et al., 1998)—the project will shed light on the impact of certain institutional care practices during infancy across the lifespan.

METHODS AND ANALYSIS

Design

This population-based study uses a combined prospective and retrospective, mixed-methods approach over a 60-year period to investigate how the lives of individuals who had been placed in an institution as infants in the late 1950s in Switzerland

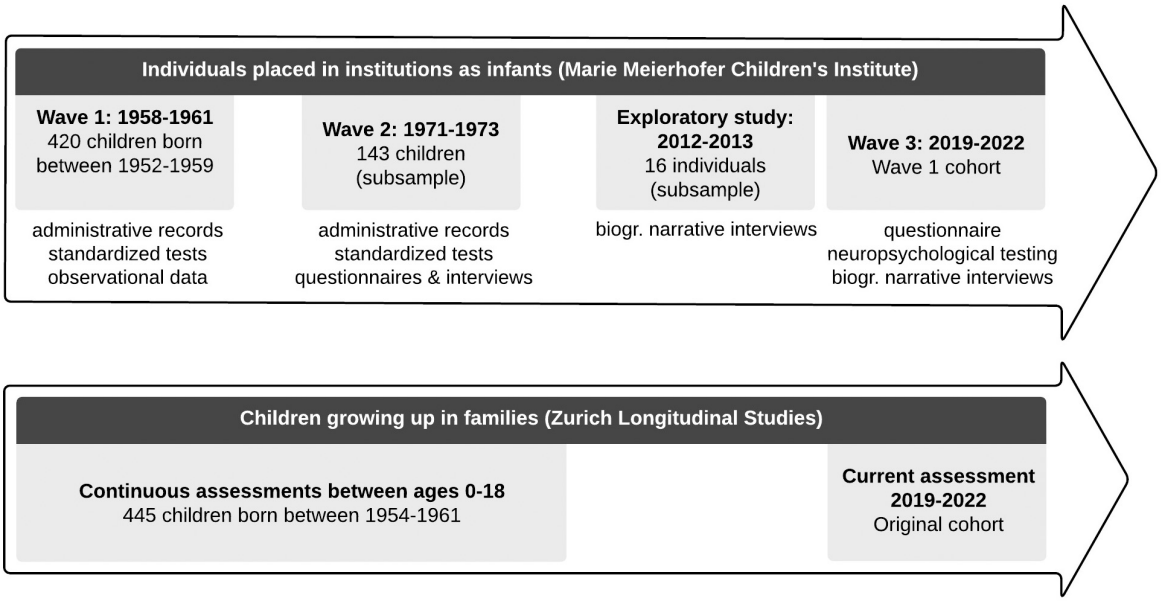


FIGURE 1 | Study design.

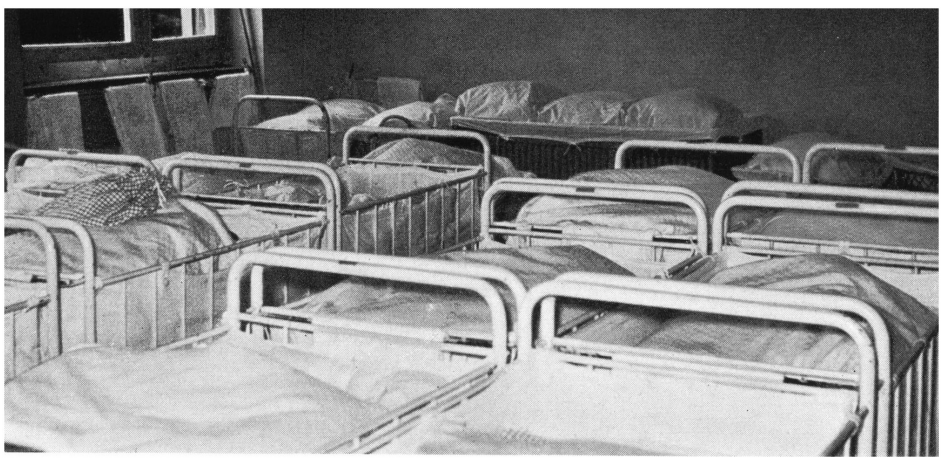


FIGURE 2 | Example of one of the rooms infants were kept in at one of the 12 institutions (Meierhofer and Keller, 1974).

developed subsequently. The availability of data from the 1954–1961 cohort of the Zurich Longitudinal Studies enables comparison with an unaffected group of children growing up in families at the same time and in the same geographic region. The overall study design is depicted in **Figure 1**. This article describes the cohort of children placed in institutions as infants. The ZLS are described in the paper in this issue (Wehrle et al., 2020).

The following sections describe the assessment at three time points of the cohort of infants placed in institutions.

Wave 1: 1958–1961

As a baseline, this project uses data from the study conducted by Dr. Meierhofer between 1958 and 1961. She conducted a survey on health and development in 12 infant and toddler care institutions in Zurich, Switzerland. Data collection took place over 16 months. During this period, around 630 children were placed in these 12 institutions. **Figure 2** provides an example of an infant room in one of these institutions to illustrate the living conditions at that time. Children who fulfilled the following eligibility criteria were included in the study: children were at least three months old at the time of the examination, they were not older than six months at the first time of in-care-placement, their institutional stay was never interrupted for more than three months, and they did not have any diagnosed medical disorder. In addition, children were excluded if they suffered from any acute infectious disease at the time of the examination or responded with

distress to the test situation. This resulted in a sample of $N = 429$.[1].

Two types of data were collected during Wave 1: person-centered data for each child and institution-centered data for each institution in which children were placed. Data was collected as a mixed-method assessment and included quantitative and qualitative data.

Person-centered data

For all children, demographic data, the psychosocial situation of the family, nature and frequency of current contact with the family, and information on children's institutional stay (e.g., age at in-care placement) were recorded. Because in most instances it was not possible to obtain this information directly from the parents, Dr. Meierhofer used the administrative records available (e.g., guardianship records) and information provided by staff members of the care institutions. These data sources were also used to record children's health status (e.g., previous diseases, birth and pregnancy history). Some of this information was recorded in standardized, self-developed study templates; other data was documented in narrative notes and underwent quantitative coding afterward. To assess children's development, a standardized developmental test was used [*Échelle de développement*, Brunet-Lézine for all children up to 2.7 years (Brunet and Lézine, 1955), Terman-Merrill developmental test for children up to 4.6 years (Terman and Merrill, 1937), and *Schweizertest* for children up to 7.6 years (Biäsch and Fischer, 1969)]. The individual test result was used to calculate an overall developmental quotient (DQ) for each child. Furthermore, height and weight were measured. In addition, children's behavior and interaction with staff members, other children, and the examiner were assessed by observations during daily routines in the institutions and during the test situation. In general, the observational data was collected and documented in a qualitative manner, and some of the data was subsequently quantitatively coded. In addition, original photographs, slides, and film has been preserved.

Institution-centered data

Information on childcare circumstances and practices in the different institutions was collected, such as interaction time, child-staff-ratio, and the educational background of staff members. In addition to information provided by the heads of the institutions, observations were recorded by the researchers during daily routines (see **Table 1** for an overview of the instruments used at Wave 1).

Today, most of the data from Wave 1 is stored in the Federal Archive in Aarau. The research team has been granted permission to access and analyze this historical data for scientific purposes related to the historic reappraisal of care practices before 1981, on condition that it is used in anonymized form, according to article 11c of the Swiss Federal Act on the Reappraisal of Compulsory Social Measures and Placements before 1981 (German: *Bundesgesetz über die Aufarbeitung der fürsorgerischen Zwangsmassnahmen und Fremdplatzierungen vor 1981*) and the Act on Public Information, Data Protection and Archives (German: *Gesetz über die Information der Öffentlichkeit, den Datenschutz und das Archivwesen; IDAG*) and the corresponding by-law (VIDAG).

Unfortunately, even though additional data from Wave 1 was retrieved from various private archives as well as the archive of the Marie Meierhofer Children's Institute, some of the data of this initial assessment remains missing. However, identifying data has been preserved for 98% of the original cohort: for 420 of the 429 children. Aggregated developmental data (e.g., DQ) is available for the majority of the children up to 3 years at the time of Wave 1 ($n = 322$).

To make this data accessible, all documents were retrieved from the archives as scans or digital photographs. Because this data is only available in analog form (mostly hand-written, some of it typed up with manual typewriters), data had to be entered manually into electronic form. A couple of issues posed a challenge to systematic data entry: firstly, information for the individual children was not available in a systematically compiled form but was distributed over many different datasheets. Secondly, the information was not recorded consistently across all of the 12 institutions but was collected using many slightly different datasheets (see **Figure 3** for an example of the available data). To deal with this and to reduce errors during data entry, standardized input templates were created for quantitative data (Limesurvey GmbH, 2020). Data was then imported into R (R Core Team, 2018) and merged for further processing and analysis. Qualitative data was typed up in Microsoft Word. Photographs, audio files, and film have been digitized for preservation.

In her analyses of the data at the time, Dr. Meierhofer found that the children placed in institutions had significant delays in cognitive, social, and motor development compared to the children growing up in families studied by the ZLS. Differences in children's development within the group of institutionalized children could not be explained by families' socio-economic status or the infants' contact with their family members. However, differences in the quality of care (i.e., interaction time and child–staff ratio) between the 12 institutions included in the study accounted for the variance in children's development. Meierhofer and Keller (1974) concluded that the developmental delays of the children were primarily caused by the poor conditions of care in the institutions.

Wave 2: 1971–1973

To assess the developmental progress and health status of the children, Dr. Meierhofer and her team conducted a follow-up study of a subsample of the children between 1971 and 1973, then aged between 13 and 15 years (Meierhofer and Hüttenmoser, unpublished).

Of the original cohort, Dr. Meierhofer only considered children born between 1957 and 1959, who were up to age three years at the time of Wave 1, for the follow-up

[1]In the further course of the study, Dr. Meierhofer excluded additional children because they either had organic disorders only diagnosed in the further course of the study, their data collection could not be completed (e.g., because information on family's social background was not available), or because a further examination of administrative records revealed that the children did not, after all, meet the selection criteria concerning age at in-care placement or continuity of the institutional stay. Some children were excluded because, due to an intermediate change of institution, they were accidently tested twice. Thus, the final sample for Dr. Meierhofer's analysis consisted of 391 children.

TABLE 1 | Overview of the assessment instruments (Wave 1 and 2).

Wave	Type of data	Instrument (References)	Subconstruct
1	Administrative records and information provided by staff members of the care institutions	n.a.	Demographic Variables/Family background
			Contact with the family
			Information on institutional stay
			Health status
			Childcare circumstances and practices (on institutional level)
	Standardized tests	Brunet–Lézine (Brunet and Lézine, 1955)	Overall development (children up to 2.7 years)
		Termann–Merill (Terman and Merrill, 1937)	Overall development (children up to 4.6 years)
		Schweizertest (Biäsch and Fischer, 1969)	Overall development (children up to 7.6 years)
		n.a.	Body Weight and Height
	Observational data	n.a. (part of the data scored according to Thalmann's symptom-burden scale, Thalmann, 1971)	Behavior
			Interaction
			Childcare circumstances and practices (on institutional level)
2	Administrative records and information provided by parents/caregivers	n.a.	Demographic Variables/Family background
			Care history
			Academic career
			Health status and pubertal development
	Standardized tests/questionnaires	WIP (Dahl, 1972)	Cognitive abilities
		Kinder-Angst-Test (Thurner and Tewes, 1969)	Mental health
		Rorschach-Test (Bohm, 1965)	Personality
		Foto-Hand-Test (Belschner et al., 1971)	Aggressive behavior
		Sohnaufsatz (Ungricht, 1955)	Graphology
		Baumtest (Koch, 1957)	Character/affective development
		Sociogram (Bastin, 1967)	Social status
		Polaritätenprofil (Instrument developed for the present study)	Personality, learning and academic behavior
		n.a.	Body Weight and Height
	Semi-structured interviews	n.a. (part of the data scored according to Thalmann's symptom-burden scale, Thalmann, 1971)	Care histories, development, social relationships, and education

study (N = 354). The address for 82 of the eligible children was not found at the time, and so it was not possible to contact them. An additional 13 children were excluded as they had moved to other parts of Switzerland and the distance from these children was deemed too great for their inclusion in the study. For 64 children, parents or legal guardians actively declined consent for participation. Another 24 did not respond to the invitation to participate (presumed passive decline). Two of the children had died. An additional ten children were excluded for other reasons (not further specified). Of the remaining 159 children, 16 participated in a preliminary study between 1969 and 1971 (Meyer-Schell, 1971)[2] and were then excluded from further study. The final sample of the full follow-up study therefore

[2]As part of a diploma thesis, Ingrid Meyer-Schell conducted a preliminary follow-up study between 1969 and 1971. From the total sample of the first assessment wave, she only selected children of Swiss nationality, born to unmarried mothers, and who were younger than 33 months at the first assessment wave. They also had to be between 11 and 12 years old at the time of the follow-up study. Of 45 children contacted, 16 children were allowed to participate. Standardized tests to assess intelligence (Stanford-Binet Intelligence Test, Terman and Merrill, 1965, and Goodenough-Draw-a-Man Test, Goodenough, 1926), frustration tolerance (Rosenzweig Frustration Test, Duhm and Hansen, 1957) and conflicts and motivations (Sceno Test, von Staabs, 1964) were used as well as additional questionnaires to include children's, parents', and teachers' perspectives.

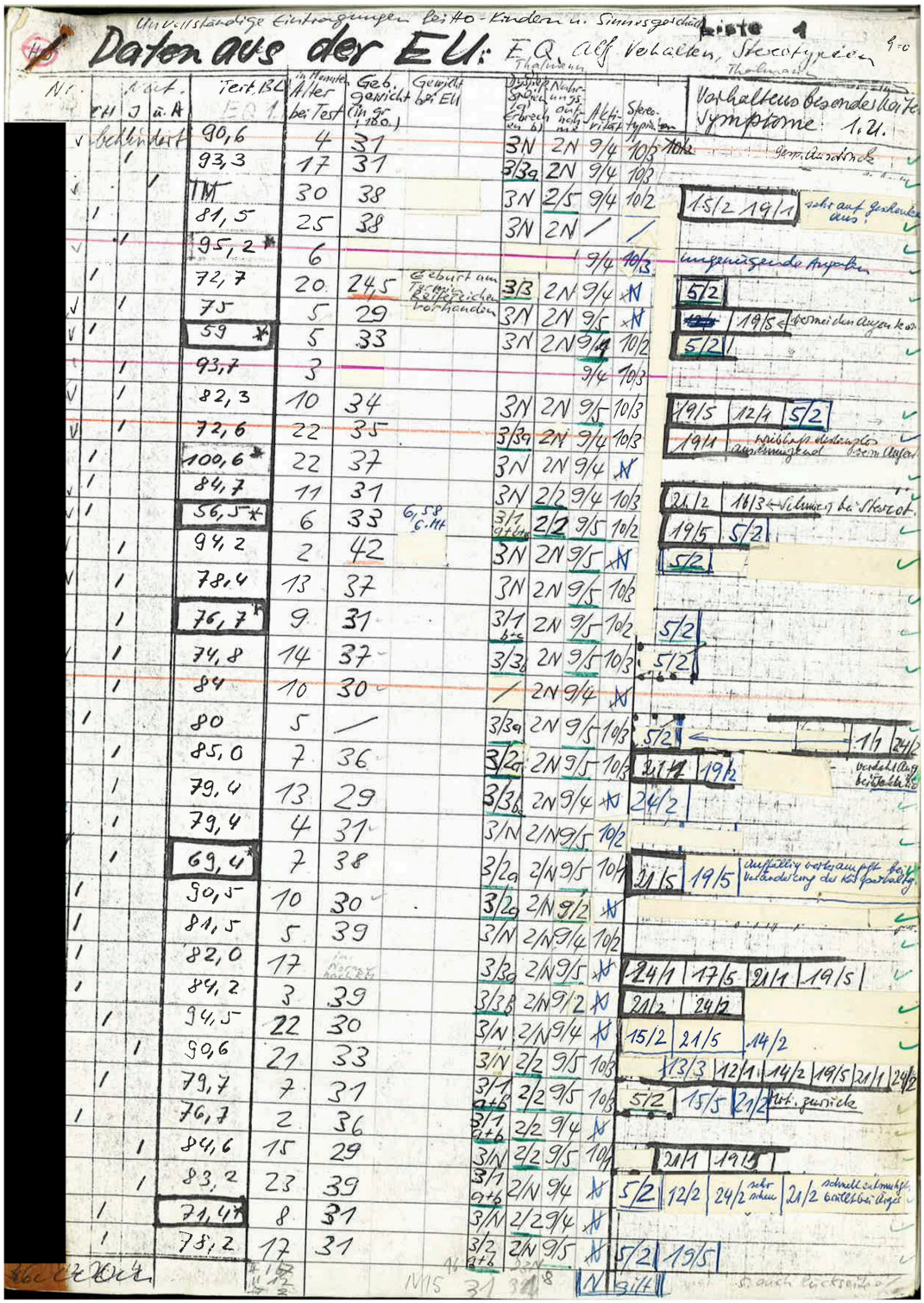


FIGURE 3 | Example of person-centered data (Wave 1).

consisted of N = 143 children. The distribution of gender, nationality, and marital status of the mother within the group of children examined was, according to Meierhofer and Hüttenmoser (unpublished) approximately comparable to that of the overall sample. In the further course of the study, 17 of these children with suspected epilepsy or other

neurological disorders were excluded from further analyses (Meierhofer and Hüttenmoser, unpublished).

Data collection during Wave 2 was carried out with a multi-method and multi-informant approach: quantitative and qualitative data were collected using administrative records, standardized tests and questionnaires, and interviews with parents or other primary caregivers, teachers, and the children themselves.

To obtain general information on the family background, children's care histories (e.g., number of different placements, reasons for placement changes), and their academic career, the study team used administrative data available, for example, from guardianship records. If necessary, parents or other caregivers were asked to provide further information. Most of the information was qualitative in nature and documented using standardized templates developed by the study team[3].

Standardized testing was conducted using a battery of neuropsychological tests and questionnaires: Children's cognitive abilities were assessed with the WIP, a short version of the Hamburg Wechsler Intelligence test for children (Dahl, 1968, 1972). Their mental health was assessed using the *Kinder-Angst-Test* (KAT), (Thurner and Tewes, 1969), a standardized questionnaire on children's anxiety. Furthermore, two projective tests were conducted: the Rorschach-Test (Bohm, 1965) to assess personality and the *Foto-Hand-Test* (Belschner et al., 1971) to assess aggressive behavior. The test battery also included the *Sohnaufsatz* (Ungricht, 1955), in which children had to write a short essay that was later analyzed for both content and graphology, and the *Baumtest* (Koch, 1957), which was intended to provide information on the child's character and affective development. In addition, children's height and weight were measured. Furthermore, data on pubertal development and general health was collected through information provided by the parents or caregivers. In addition, the social status of the children among their classmates was determined with a sociogram (Bastin, 1967): the children themselves and five of their classmates were asked to indicate with which children they would most like to spend time and with whom they would least like to spend time. The number of children's mutual choices was evaluated.

In addition, semi-structured interviews were conducted with primary caregivers, children themselves, and their teachers. For the interviews with the parents, key themes were predefined, such as pregnancy and birth history, care histories, children's development (motor skills, language, sleep, tidiness), social relationships, and education. Furthermore, interviews with the children contained questions about school, leisure activities, friendships, and their ideas about their future. The teacher survey focused on academic performance and social contact with classmates. If necessary, interviews with migrant workers were conducted in Italian and subsequently translated. The original audio-recordings and some of the transcripts have been preserved. The qualitative information collected during the interviews was scored according to Thalmann's symptom-burden scale (Thalmann, 1971) and used to assess children's behavior. In addition, teachers and researchers recorded information on the children's personality and learning and academic behavior (for example reliability, discipline, independence, pace of work) on a 5-point standardized scale (*Polaritätenprofil*). **Table 1** provides an overview of all instruments used at Wave 2.

At the time of Wave 2, data from the ZLS was not ready for comparison. Dr. Meierhofer and her team therefore selected the instruments for the follow-up-study so that data of a normative sample and, if possible, even comparative data from Switzerland was available that allowed basic comparisons (Meierhofer and Hüttenmoser, unpublished).

In her analysis, Dr. Meierhofer found that at the time of Wave 2, children who were placed in institutions as infants showed increased depression, school related-problems (e.g., significantly higher grade retention rate than the comparison group), and stereotypies (Meierhofer and Hüttenmoser, unpublished). Due to a fierce nature/nurture debate with one of her colleagues at the time, these results were never published (Wyss-Wanner, 2000) and exist only in an original research report in the archives of the Marie Meierhofer Children's Institute (Meierhofer and Hüttenmoser, unpublished).

In contrast to the Wave 1 data, the raw Wave 2 data has been fully preserved. Part of the data is stored in paper form at the Federal Archive in Aarau. It has been digitized in the same way as the Wave 1 data by first retrieving the data from the archive as digital photographs and then entering the data manually into data entry masks created with LimeSurvey (Limesurvey GmbH, 2020). The remaining data is available in paper form and on microfilm in the archive of the Marie Meierhofer Children's Institute and has now also been digitized for further processing. **Figure 4** provides an example of the available data. Just as for the Wave 1 data described above, data entry masks were created in LimeSurvey (Limesurvey GmbH, 2020), and further data processing were carried out with R (R Core Team, 2018).

Qualitative data will be typed up in Microsoft Word either from available audio files or from original transcripts.

Wave 3: 2019–2021

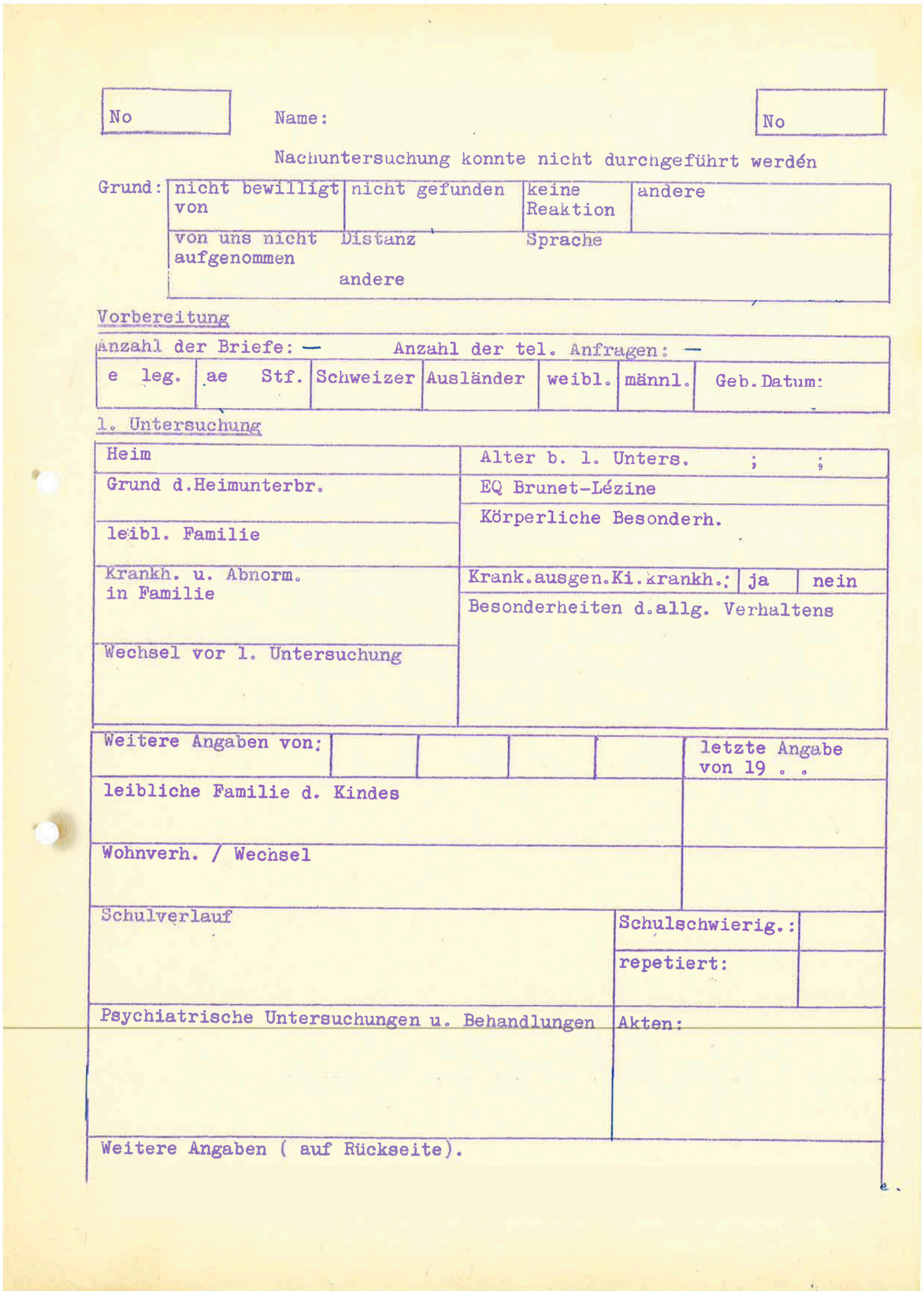
In a newly revived effort funded under the umbrella of the Swiss National Research Program 76, which is focused on scientific investigation into the compulsory social measures and placements before 1981, all individuals that took part in the study by Dr. Meierhofer at Wave 1 were once again located and contacted. These individuals were about 60 years when taking part in this third assessment of health and development (Wave 3).

Study preparations

Prior to Wave 3, two preparatory studies were conducted. They are separate, small studies but are briefly summarized here.

Exploratory study. Between 2012 and 2013, an exploratory study was implemented. Semi-structured narrative interviews were conducted with 16 individuals that had taken part in Waves 1 and 2. As part of the semi-structured interviews, participants talked about topics such as family, institution, foster or adoptive family, relationships, health, education, work life, hobbies, and aging (Ryffel and Simoni, 2016). Furthermore, findings suggested

[3]For some children, it was even possible to obtain basic information from administrative records or through parents or caregivers during telephone contact, even if the children did not participate in the follow-up study.



No | Name: | No

Nachuntersuchung konnte nicht durchgeführt werden

Grund: nicht bewilligt von | nicht gefunden | keine Reaktion | andere

von uns nicht aufgenommen | Distanz | Sprache | andere

Vorbereitung

Anzahl der Briefe: — Anzahl der tel. Anfragen: —

e leg. | ae Stf. | Schweizer | Ausländer | weibl. | männl. | Geb. Datum:

1. Untersuchung

Heim | Alter b. 1. Unters. ; ;

Grund d. Heimunterbr. | EQ Brunet-Lézine

leibl. Familie | Körperliche Besonderh.

Krankh. u. Abnorm. in Familie | Krank. ausgen. Ki. krankh.: ja | nein

Wechsel vor 1. Untersuchung | Besonderheiten d. allg. Verhaltens

Weitere Angaben von: | letzte Angabe von 19 . .

leibliche Familie d. Kindes

Wohnverh. / Wechsel

Schulverlauf | Schulschwierig.: | repetiert:

Psychiatrische Untersuchungen u. Behandlungen | Akten:

Weitere Angaben (auf Rückseite).

FIGURE 4 | Example of person-centered data (Wave 2).

that despite clear indications of resilience in some individuals, many generally struggled to find a sense of belonging and coherence in their lives. This seemed to be the case more often for individuals placed in institutions as a result of having been born to unmarried and/or underaged mothers than born to migrant workers. Children of migrant workers tended to reassured themselves that they were loved by their parents and that it was external circumstances that led to the placement. They remember it as a measure that "made sense" (Ryffel and Simoni, 2016; Sand and Gruber, 2017).

The experience of this exploratory study informed the search strategy, contact procedures, ethical considerations, and research questions for Wave 3 (for details see corresponding sections below). It also reaffirmed the need for a comprehensive long-term follow-up assessment, as it revealed how diverse the life trajectories can be within this cohort. It also showed that individuals are willing to participate in a research study and talk about their biographical trajectories.

Participatory research preparation. To prepare for Wave 3, in 2019, we conducted focused interviews with four individuals that had been placed in institutions as children to elicit feedback on procedures, documents, and assessment instruments to be used for contacting the cohort. This was in response to the request of those affected by the compulsory social measures and placements before 1981 to be included in research related to the reappraisal and reconciliation process (UEK Administrative Versorgungen, 2019; Unabhängige Expertenkommission Administrative Versorgungen, 2019). Interviewees' feedback helped to make documents more understandable and identified wording that might cause insecurities or trigger negative reactions. They also indicated shortcomings in some of the questionnaire items that they thought were potentially misleading or not appropriate for the situation of the participants (Lannen et al., 2020). In addition, they made significant contributions to how best to approach and work with the cohort. This participatory preparation showed that the inclusion of formerly institutionalized individuals in research is feasible and provides substantial benefits to the research quality on historic compulsory social measures and placements (Lannen et al., 2020).

Eligibility

Eligibility criteria were defined separately for (a) locating individuals and (b) contacting them.

In order to be eligible for the search, individuals had to have taken part in Dr. Meierhofer's study at Wave 1 and to have been assigned a study number for which identifying data was available. In addition, a minimum of three data points needed to be available for the search (full name, date of birth, and a location of residence from some point in their lives). A number of individuals were found to have moved abroad (25%, $n = 107$). If they were found to be residing in a country other than their country of origin, we were able to search them through the embassy or the Bureau of Foreign Affairs in charge. If a city they have moved to abroad is known, they can be searched for through the local population registry, if one exists. No search strategy exists for individuals who moved to their country of origin without any indication of location. In this case, individuals abroad are ineligible for the search.

A person who was not found and those deceased were obviously deemed ineligible for contact. Those that were found, were only eligible for contact, if there was no indication that they might have been adopted without their knowledge (happened early in life, was followed by a name change). This was based on the ethical concern of uncovering a potentially unknown adoption and causing distress to the individual. If there was an indication in the historical files that the individual knew about the adoption, or that the adoption happened within the family (e.g., by the new partner of the mother), individuals remained eligible for the study. In addition, individuals also became ineligible for contact if they had a data protection barrier with the municipalities (see below section "Locating individuals"). Finally, individuals who had actively declined study participation when contacted for the exploratory study (Ryffel and Simoni, 2016) were also deemed ineligible.

Locating individuals

The search process took place between October 2018 and March 2021.

Due to the sensitive nature of the request, the process of locating individuals after several decades was set up in a way that minimized the risk of false identification of individuals. Therefore, even though very labor intensive, the search for each individual's identity needed to be officially confirmed through the population registry. Luckily, this was possible, as in Switzerland, every individual is formally registered with the municipality where he or she resides. Individuals who relocate must give notice of departure with the old municipality and formally register with the new municipality, so individuals can be tracked through the system over time (municipal population registry). In addition, events (such as birth, marriage, divorce, or death) are registered in the civil population registry (*Zivilstandsregister*) in a person's commune of origin (*Heimatort*)[4].

In Switzerland, legislation differs slightly between the cantons[5], but generally, municipalities will provide full name, address, and dates of arrival in and departure from the municipality from the population registry to private individuals and organizations acting in the public interest without requiring a reason for the request[6]. If a "credible argument" is made, the municipality that the person moved to and from, date of birth, gender, marital status, and commune of origin (*Heimatort*) can also be released[7]. Before the information is released, the municipality is to verify that the request is made in the public

[4]Each resident in Switzerland with Swiss Nationality also has a "commune of origin" (*Heimatort*), similar in function to "place of birth" in other countries. The office of civil registry (*Zivilstandsamt*) in that commune of origin holds key documents such as birth certificates and marriage records. The civil registry records the place of residence at the time of events relevant to civil status (such as marriage, divorce, change of name).

[5]Cantons are the member states of the Swiss Confederation.

[6]In the Canton of Zurich, for example, this is stated in §18 section. 1 Law on registration and population registers (German: *Gesetz über das Meldewesen und die Einwohnerregister, MERG*).

[7]In the Canton of Zurich, for example, this would be *Bekanntgabe von erweiterten Personalien,* §18 Section 2 MERG.

interest[8]; if this is not case, the municipality is not allowed to release the information. Individuals that have actively instated a data protection barrier are excluded[9]. The municipalities also provide information on deceased individuals should the applicant make explicit an interest for the information[10]. Some municipalities requested more detailed information on the public interest of the request. In these cases, detailed information on the study and on data protection was provided. The study was described as a general survey on health and development and the fact that the individuals in question were placed in institutions as infants was not disclosed to protect their privacy.

A successful search for an individual through municipal population registries requires three data points: the full name, date of birth, and a municipality in which the individual has lived at some point. For a majority of individuals that took part in Wave 1, information on municipalities was available through the study archives. For some, we were able to gain access to municipalities of residence through the Zurich's City and Federal Archive (archived municipal registration files before 1976, archived supplementary files of birth registries, guardianship records, infant care institution files etc.). The municipal authorities released the current address if the individuals still lived in that municipality or released the name of the municipality the person moved to. The address information request was then iterated through as many municipalities as necessary until the current address was found.

For some cases, instead of a municipality of residence, the person's commune of origin (*Heimatort*) was available through the archives. A formal research request was submitted to the office of civil registry in Zurich and granted under Art. 60 of the *Zivilstandsverordnung vom 28.04.2004* (ZStV; SR 21 1.112.2) to inquire whether a place of residence can be found in the files for individuals still alive or a date of death for those deceased. Upon the precedent of Zurich, civil registries in other locations provided access to the information. For these individuals, the last known address through civil registries was thus identified and then the above-mentioned process of locating individuals through municipalities repeated for these individuals. An additional research request granted access to the information on deceased individuals whether death was of natural or unnatural (accident, crime, or suicide) causes. This information was made available by the civil registries in the municipalities where individuals resided at the time of death.

For a few individuals, their municipality of residence was found through moneyhouse.ch, an online platform providing information drawing among others from the cantonal commercial registers (since full name and date of birth was available through this portal) and then verified with the municipality.

Overall, the procedure for contacting individuals was very resource intensive and is still ongoing. The search was reiterated through up to 10 municipalities until a person's current address was identified. More than 300 municipalities were contacted, as well as more than 40 offices of civil registries in Switzerland alone. In total, more than 2000 emails were sent.

So far, 86% (n = 268 of total 313) of individuals residing in Switzerland were found, out of those, 78% (N = 208) were eligible for contact (not deceased or ineligible).

So far, 25% (n = 107) of individuals were found to have moved abroad. If a person had moved to a country other than their home country, we placed an inquiry with the embassy or the Bureau of Foreign Affairs. If the person was registered by their country as living abroad, he or she could be located. If a person was not registered but the municipality the person moved to was known, rather than just the country, then analogous to the search in Switzerland, individuals were searched for through the population registry (if some form of population registry existed in that country). This search strategy via municipalities was applied for all individuals who moved abroad regardless of their nationality or home country. Using these strategies, we have identified 47%, n = 50 individuals eligible for search abroad. Out of those, we have been able to locate and contact 24% (n = 12) so far.

Figure 5 shows the search path.

Contact Procedures

The contact procedure for this study was built on a number of key principles: first, it takes a stepwise approach, so that those individuals that do not wish to learn about any study details can opt out swiftly (ethical protocol, see section "Ethics"). Second, contact procedures were also staggered to ensure sufficient resources were available within the research team for personal contact with the cohort and data collection. Third, contact procedures were designed for a broad range of life trajectories, building on Antonovsky's salutogenetic approach (Antonovsky, 1987). This is expressed in the wording of documents and behavior of the researchers. Fourth, even though some participants might not remember that they had participated in the study in the past, overall, the study invitation has been framed as an invitation to remain in the study. Finally, the procedure has been set up as an opt-out procedure: at each step, the next point of contact through the research team will be announced, and individuals have to actively let the research team know if they want to disengage from the process. Passive decline has been clearly operationalized as no more contact after a final reminder for participation by mail if a participant cannot be reached by phone after three attempts.

Furthermore, the Swiss Federal Act on the Reappraisal of Compulsory Social Measures and Placements before 1981 (German: *Bundesgesetz über die Aufarbeitung der fürsorgerischen Zwangsmassnahmen und Fremdplatzierungen vor 1981*) stipulates that individuals have the right to view their archived files. Accordingly, the study team has created a leaflet with detailed information on how individuals can access any archived information, either related to their institutional placement or the study.

As an initial form of contact, a short letter with basic information about the study and the individual's name and year of birth found in the archives was sent to each prospective

[8]Canton of Zurich: §23 Law on information and data protection (German: *Gesetz über die Information und den Datenschutz, IDG*).

[9]Canton of Zurich: §22 IDG.

[10]Canton of Zurich: §19 IDV.

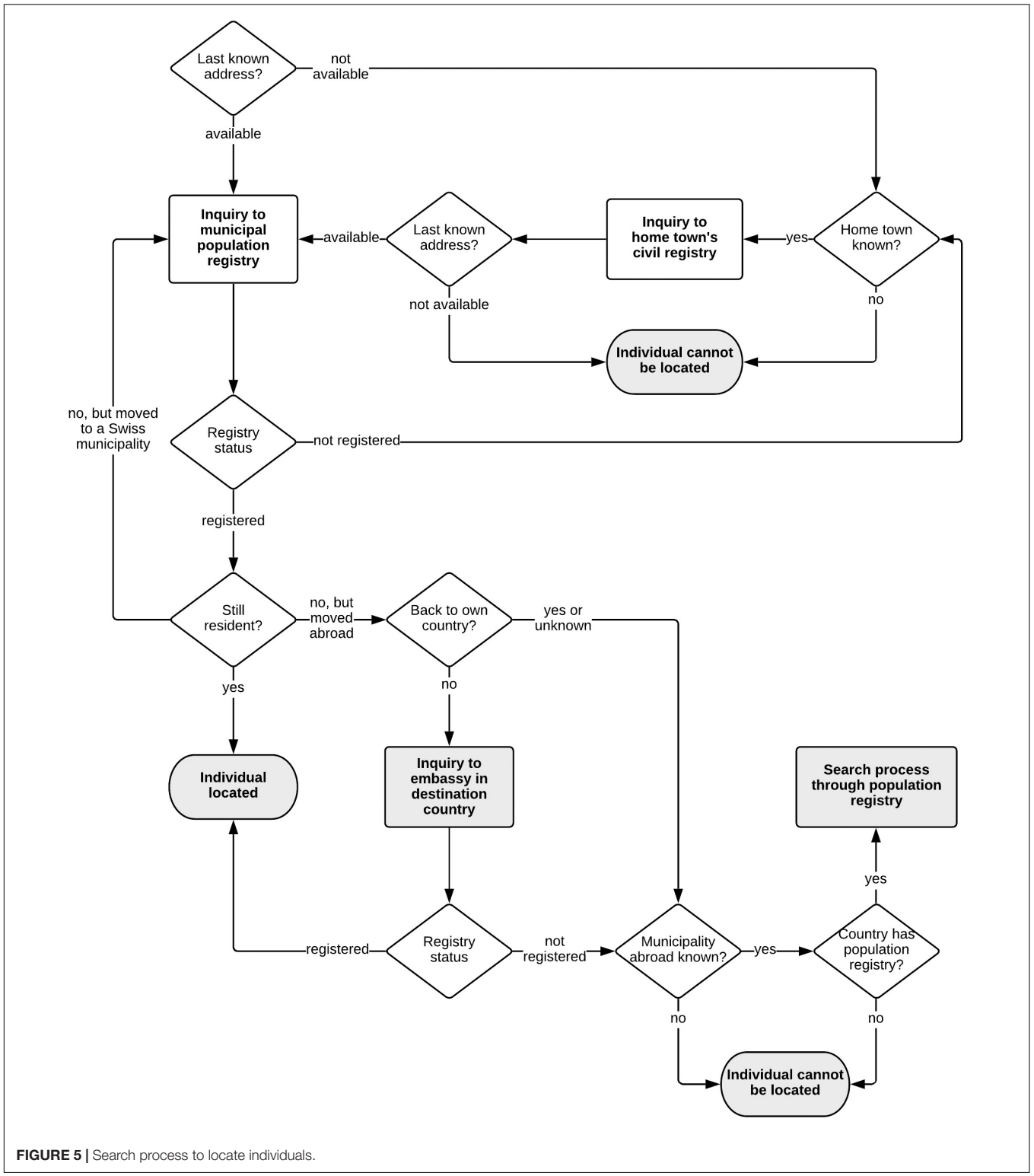


FIGURE 5 | Search process to locate individuals.

participant. It also included a request to let the researchers know by phone, email, or returning a slip (a stamped return envelope was included) if they did not want to participate. In that case, they were not contacted again. Two weeks later, anyone who had not actively declined was sent a second letter. This time, more detailed information was provided about the study, including a detailed leaflet that outlined its history, aims, and procedure. It reiterated that participation is voluntary and that they can withdraw at any time. It also outlined data protection measures. In addition, the leaflet specified the

components of the study (questionnaire, neuropsychological assessment, biographical narrative interview; for details see section "Data collection"), including expected duration and reimbursement. Individuals received CHF 80. – each for filling in the questionnaire and participating in the neuropsychological assessment, and CHF 40. – for participating in the biographical narrative interview. In addition, they received reimbursement for travel expenses. Again, if anyone indicated that they did not want to participate further (active decline), they were not contacted again. Otherwise, senior staff members called all the individuals for whom we were able to obtain a phone number to invite them to participate in the study and go over the consent form initially over the phone. Notes of the conversation were recorded. Those individuals for whom phone numbers were not available were invited to provide one with a return slip. It was also possible to indicate preferences for communication (mail only, for example, for those who do not want to talk on the phone but would like to participate in the study).

The same researcher remained available throughout the study to ensure consistency and to build trust. If an individual agreed to participate, the questionnaire was sent together with the consent form, and appointments for neuropsychological assessment and interviews were arranged. Help with completing questionnaires was offered when necessary over the phone as a standardized interview. A reminder call was made the day before the appointment. If necessary (e.g., due to poor health), the researcher offered to travel to the participants' home for data collection.

If the full questionnaire was not returned after 3–4 weeks, the researcher checked in with the participant either by phone or letter to remind the participant to fill in the questionnaire and to see if any assistance was needed. Those individuals who could not be reached by phone were sent a final reminder letter, including the short version of the questionnaire, a consent form, and a slip to indicate active decline or interest in further participation. Sending a final reminder letter with a questionnaire is based on the work of the *Zürcher Längsschnittstudie Von der Schulzeit bis ins mittlere Erwachsenenalter* (Schallberger and Spiess Huldi, 2001), which led to a substantial number of additional questionnaires returned.

For those individuals that were contacted as part of the exploratory study (Ryffel and Simoni, 2016), letters were tailored to the individuals' responses.

The recruitment process is depicted in **Figure 6**.

Data collection

Data collection started in September of 2019, is still ongoing, and is expected to last throughout Q1 of 2021.

A mixed-methods approach was chosen to maximize depth and breadth of data and to enable the accommodation of data collection to individuals' possibilities and preferences where necessary. Specifically, three types of data were collected: questionnaire data, data from neuropsychological assessments, and data from biographical narrative interviews. The full battery of questionnaire data covered physical and mental health, social abilities and demographics, information on work and family life, retrospectively captured education and professional background, critical life events, and transitions. It included constructs such as sense of coherence, self-efficacy, and the ability to fulfill one's basic psychological needs, which are all factors believed to be important moderators for adjustment and fulfilled, happy lives. All constructs and the corresponding operationalization and instruments used to assess them are listed in **Table 2**. The full questionnaire battery took about 120 min to complete. A pilot data collection phase had indicated that the nature and length of questionnaire was appropriate for the target population. For those unable or unwilling to fill in the full questionnaire battery, a short version of the questionnaire was provided, covering physical and mental health outcomes and demographic information (an estimated 60 min to complete).

In addition, the researchers invited participants to come to the Marie Meierhofer Children's Institute in Zurich for neuropsychological assessment (estimated duration: 120–150 min). A full list of assessment instruments is listed in **Table 2**.

Questionnaire and test data were digitalized by entering it into digital questionnaires programmed with LimeSurvey (Limesurvey GmbH, 2020). This procedure facilitated data entry and reduced the potential for errors. Inter-rater reliability checks will be performed for test scoring. In addition, data entry was double checked by a second researcher for both test data and questionnaire data. Data was exported from LimeSurvey using a comma-separated format (.csv) and cleaned with R (R Core Team, 2018).

Preliminary power analyses indicated that for a *t*-test for independent samples with an alpha-error of 0.05 and a power of 0.80, at least 102 participants are required to detect medium effect sizes, while the detection of large effect sizes requires 42 participants.

Finally, participants were invited for biographical narrative interviews (Rosenthal, 1995), focused on the individual narrative that participants present when asked to talk about their lives. These interviews were conducted without a predefined time limit, and the participants are asked to explore the topics and share the experiences that are relevant to them as they wished. The interviews were conducted at the location of the participants' choice. Interviews were recorded with an audio recorder and later transcribed and anonymized (Bohnsack et al., 2013). The audio files were then deleted to ensure data protection.

The time and nature of death of deceased participants was determined from the records of the municipal and civil registries.

Comparison With Zurich Longitudinal Studies

In the 1950s, Dr. Meierhofer was in close contact with Prof. Guido Fanconi, the medical director of the University Children's Hospital who initiated the ZLS study in 1954. Dr. Meierhofer aligned the assessment instruments used to study the infants in the institutions with those used in the ZLS (Wehrle et al., 2020). Dr. Meierhofer studied the children at two time points: when children were between birth and three years of age (Wave 1) (Meierhofer and Keller, 1974), and a subsample of the children about 10 years later (Meierhofer and Hüttenmoser, unpublished). Children in the ZLS, were continuously assessed until age 18 (Wehrle et al., 2020). In a parallel effort, both cohorts are

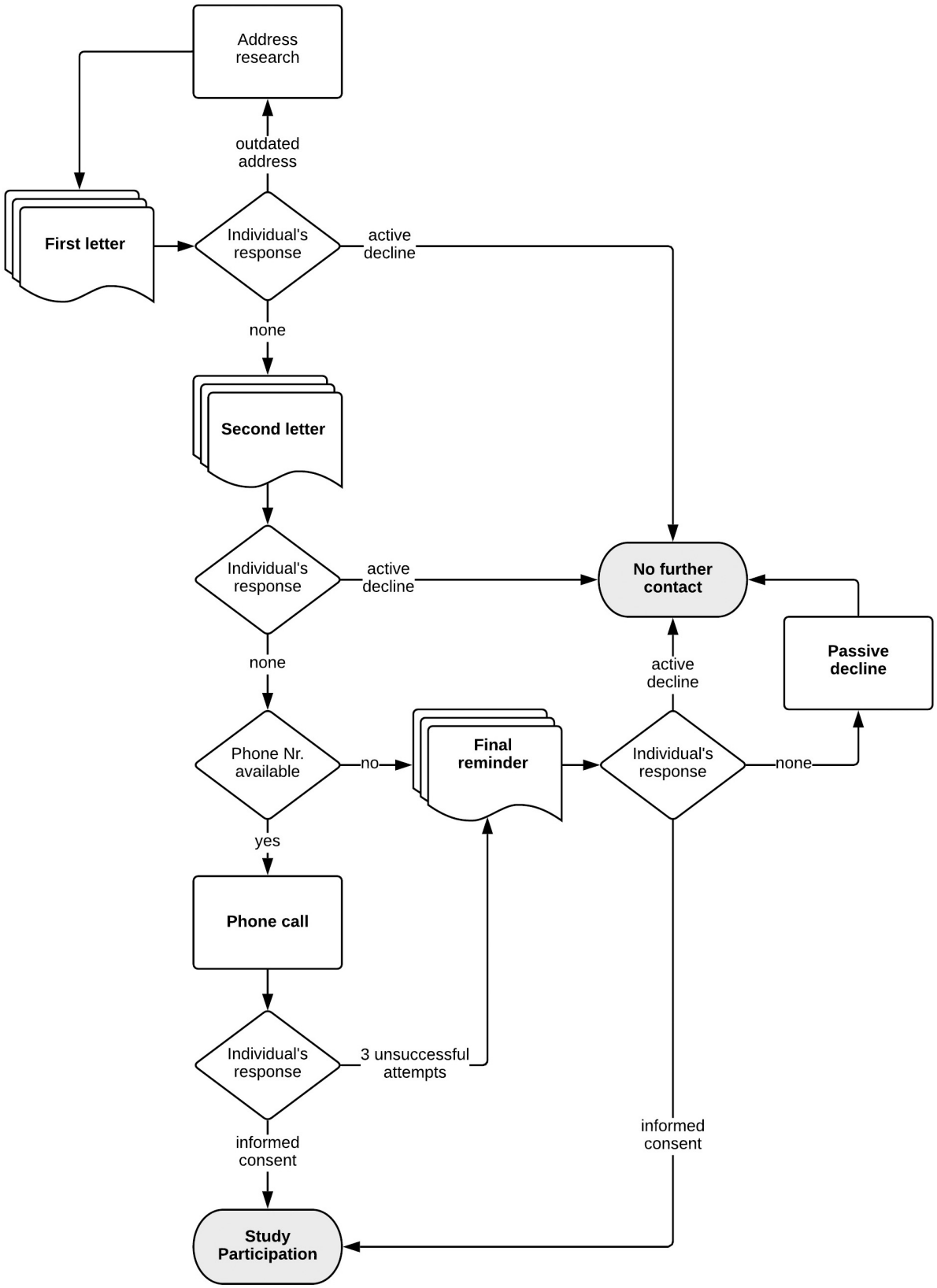


FIGURE 6 | Recruitment process.

TABLE 2 | Overview of the assessment instruments (Wave 3).

Format	Instrument (References)	Subconstruct
P&P	Instrument developed for the present study (n.a.)	Physical Health
		Physical Activity
		Sleep Patterns
		Nutrition
		Demographic Variables
		Biographical Cornerstones
		Evaluation of Study Participation
	Brief Symptom Checklist (Franke, 2017)	Psychiatric Symptoms
	NEO-Five Factor Inventory (Borkenau and Ostendorf, 2008)	Personality
	Cognitive Emotional Regulation Questionnaire (Loch et al., 2011)	Emotion Regulation
	Resilience Scale (Schumacher et al., 2004)	Resilience
	Health Questionnaire (Morfeld et al., 2011)	Subjective Physical and Psychological Functioning
	Pittsburgh Sleep Quality Index (Buysee et al., 1989)	Sleep Quality
	Satisfaction with Life Scale (Glaesmer et al., 2011)	Satisfaction with Life
	Sense of Coherence Scale (Schumacher et al., 2000)	Sense of Coherence
	General Self-Efficacy Scale (Beierlein et al., 2012)	Self-Efficacy
	Incongruence Questionnaire (Grosse Holtforth and Grawe, 2003)	Basic Psychological Needs
	Social Support Questionnaire (Fydrich et al., 2007)	Perceived Social Support
	Adult Attachment Scale – R (Schmidt et al., 2016)	Attachment Behavior
	Relationship Questionnaire (Bartholomew and Horowitz, 1991)	Attachment Style
	Retrospective 1-item measurement (Neumann, 2002)	Attachment to Own Parents
	Saarbrücken Personality Questionnaire (Paulus, 2011)	Empathy
NPA	Wechsler Adult Intelligence Scale (WAIS) (Petermann and Petermann, 2013)	Fluid Intelligence
		Crystalline Intelligence
		Verbal Intelligence
		Working Memory
		Numbers–Symbols Test
	Zurich Neuromotor Assessment II (Kakebeeke et al., 2018)	Fine and Gross Motor Skills
	Materialien und Normwerte für die neuropsychologische Diagnostik (Balzer et al., 2011)	Fluency
	Trail Making Test (Tombaugh, 2004)	Cognitive Flexibility
	Stroop Test (Strauss et al., 2006)	Inhibition
	Hopkins Verbal Learning Test (Brandt and Benedict, 2001)	Verbal Learning and Memory
	Rey Complex Figure Test (Rey, 1941)	Visual Construction, Figural Learning and Memory
	Instrument developed for the present study (n.a.)	Assessment-Related Health Issues
	n.a.	Body Weight and Height

currently being located and assessed again. Once again, the study instruments have been closely aligned, thus allowing a 60-year longitudinal design with a comparison group.

Adjustments Related to the Pandemic of COVID-19

About midway through data collection for Wave 3, the COVID-19 pandemic reached Switzerland, and measures to contain the spread of the virus were implemented. Neuropsychological assessments and interviews on site were suspended for three months between mid-March and mid-June 2020 and then recommenced with protection measures. Data collection for questionnaire data continued throughout, and now includes a questionnaire to assess the impact of the measures to contain the pandemic on a range of outcomes: work and finances, home and social life, daily routines, and mental and physical well-being. This additional questionnaire was also sent to participants who had already completed data collection. Furthermore, this same questionnaire was sent to participants of the ZLS (for details see this issue Wehrle et al., 2020). This will allow us to check for any possible bias in data collected before or after the onset of the pandemic as well as whether reactions differed between the two cohorts to the measures implemented by the government.

Data Analysis

Dr. Meierhofer and her team analyzed the historical data with the tools and the statistical techniques available at that time. For instance, tables and graphs were drawn by hand.

The historical data are now being re-analyzed with modern statistical methods to increase the reliability and validity of the historical results. Specifically, for data from Wave 1, we will seek to replicate Dr. Meierhofer's finding that institutional placement and especially care conditions are significant predictors of the differences in children's developmental outcomes. Our primary aim with the data of Wave 2 is to compare it with the data now available from the ZLS and to finally make the data available through publication.

Data will allow longitudinal analyses over two and three time points, both within group for children placed in institutions and between groups with children placed in institutions and children who grew up in families. The within-group analyses will examine the role of a set of potential predictors of inter-individual differences in growth in health and development. The between-group analyses will compare the two cohorts on the same outcomes.

We will apply structural equation modeling and multivariate regressions using a combination of variable- and person-centered statistical methods. These methods will allow us to account for measurement error (latent variable modeling), the multilevel structure of the data (multilevel modeling or correction of standard errors where multilevel modeling cannot be applied), and multiple covariates. A further important aspect of data analysis will be the implementation of multiple imputation methods to address missing data in general and selective drop-out in particular. A preliminary review of the raw data also indicates that historical data contains some errors that took place when transferring data between sheets or when rounding to the decimal. These errors will be corrected systematically.

For data from each wave, we will compute the deviation score of each participant of the cohort placed in institutions from the ZLS cohort mean. The resulting deviation score at Wave 3 will then be predicted by the deviation scores for Waves 1 and 2 by regression modeling. In addition, the influence of the individual variables on the longitudinal trajectory of health and development will be calculated using random-intercept autoregressive models (Hamaker et al., 2015) and latent growth models (Sticca and Perren, 2015). Multivariate associations among multiple constructs of interest will be examined using multivariate random-intercept cross-lagged models (Hamaker et al., 2015) and parallel process models (Sticca and Perren, 2015), depending on the research question at hand.

Qualitative data will be coded and descriptive analyses run and depicted in graphs where useful. Longer handwritten narratives and texts from Waves 1 and 2 will be subjected to content analysis (Mayring, 2000). Selected historical qualitative materials and narrative interviews conducted at Wave 3 will also undergo reconstructive and sequential analysis according to Rosenthal's (Rosenthal, 2015) method. The aim of this abductive analysis method is to identify the latent content of the text, make generalizations, and come to conclusions via in-depth analyses of individual cases (Rosenthal, 2015).

Qualitative and quantitative data will be triangulated and historical data contextualized in the historic discourse on education and upbringing, structural violence, welfare practices, discrimination of certain family models, and compatibility of employment and family.

Ethics

A number of measures identified ethical issues and mitigated risks for all three waves.

An independent ethics expert reviewed both historical studies (Wave 1: 1958–1961 and Wave 2 1971–1973), drawing on primary historical data and reports and publications. The review concluded that although research ethics practices have developed over time, Dr. Meierhofer adhered to the key ethical principles that hold today, chiefly the principle of not harming the subjects. The review even concluded that the children that took part in the study may have benefited; they certainly received some extra interaction time and attention as part of the study, in contrast to the deprivation prevalent in the institutions at that time. Assessment took place through observation and conversation and was non-invasive. At Wave 1, consent was provided by the head of the care institutions and at Wave 2 by the children's legal guardians. Without such consent, children were not included in the study. Data was anonymized for analysis and publication (Brauer, 2019a).

For Wave 3, a number of ethical issues were identified. These included, for example, the risk of disclosing a previously unknown or not remembered institutional placement, the risk of disclosure of an institutional placement to next-of-kin or other third party, distress caused by learning about previously unknown inclusion in a study, discussing potentially distressing

events from the past as part of data collection, and distress caused by accessing archived materials.

A comprehensive ethical framework and ethics protocol were developed that detailed every step of interaction with the individuals of the cohort with the aim of mitigating potential risks. In addition to detailed informed consent, voluntary participation, the option of withdrawing consent at any time, and protection of personal data, measures included a step-wise approach to contacting individuals with increasing amounts of information on the study, an option to claim mistaken identity, easy opt-out procedures, contact only with senior researchers, regular, standardized screening for distress, psychological support available on site for study participants, and psychological supervision available to researchers.

Furthermore, a thorough consent procedure was implemented, including seeking consent to link new and historical data. Participation was voluntary, and consent could be withdrawn at any time. Data will only be analyzed and published in an anonymized form.

These ethical considerations and ways to mitigate distress were developed and reviewed in consultation with an external, independent ethicist (Brauer, 2019b). Furthermore, findings from the exploratory study outlined in section "Exploratory Study" (Ryffel and Simoni, 2016) and the focus interviews conducted as part of the participatory study preparation (Lannen et al., 2020) (see section "Participatory research preparation") also shaped the ethical framework of the study.

Wave 3 was reviewed and approved by the Ethics Committee of the Faculty of Philosophy at the University of Zurich (Approval Number 19.4.7).

DISCUSSION

A number of longitudinal studies over several decades on human development have been conducted in different parts of the world (for an exemplary overview see this issue Wehrle et al., 2020) and emerging data from robust longitudinal studies of children placed in institutions is becoming available (Nelson et al., 2014). However, the 60-year span of the LifeStories project presents the longest follow-up of children who spent their early years in an institution. Central to the feasibility of this study are a number of significant opportunities and challenges, with several lessons emerging from each of them.

Historical Data

Comprehensive long-term longitudinal studies over several decades are rare, because they require continuous project leadership across several generations of researchers and an institutional home for the data to guarantee data access and data preservation over time. Other challenges of long-term longitudinal studies include the advancement of science overall: concepts and methodologies may change fundamentally after the initiation of the study. For instance, instruments used several decades ago might no longer be in use, thus making direct comparisons across the life span difficult. Some of these challenges are described and discussed in detail in the paper in this issue (Wehrle et al., 2020). For this study, even though it was dormant for over 40 years, access to data was possible through the continuity of institutional involvement and became accessible to the next generation of researchers at the Marie Meierhofer Children's Institute. However, the fact that the data was not preserved in its entirety and had to be reassembled from a variety of documents—some organized by institution, some by subject, and some by outcome—posed a significant challenge. Using modern tools such as LimeSurvey templates and having another researcher enter data a second time mitigated the risk of data entry errors.

Locating Individuals

A second challenge was to locate individuals after many decades while reducing possible false identification. However, being able to locate individuals through population registry is promising and has been successfully implemented, primarily by studies in the Nordic countries, where a central population registries exist (Kreicbergs et al., 2004; Lannen et al., 2008, 2010; Surkan et al., 2008; Lysholm and Lindahl, 2019). In Switzerland, the process is not centrally organized, and tracking individuals through the system over time proved to be very time consuming. Nonetheless, it was possible to find 86% ($n = 268$) of individuals as long as they still lived in Switzerland even decades after the previous contact. In fact, it proved possible to find individuals that Dr. Meierhofer was unable to locate at the time of Wave 2. It is plausible, but difficult to fully verify, that the reasons why individuals were not found in Switzerland included both death and a name change after an adoption.

Due to the substantial proportion of children of migrant workers (Meierhofer and Keller, 1974) in the cohort, almost one quarter of the individuals moved abroad. Finding these individuals is central to addressing some of the key research questions of the study, as the development trajectory of this subgroup of individuals might differ significantly due to familial backgrounds and reasons for placements. Tracing individuals who had moved abroad proved more difficult, but so far, 24% ($n = 12$) of eligible individuals living abroad were found.

Contacting and Recruiting Study Participants

In recent years, a growing number of studies have examined the experience of formerly institutionalized individuals in Switzerland (Bombach et al., 2017, 2018b; Ammann et al., 2019; UEK Administrative Versorgungen, 2019). However, the primary approach of these studies was an opt-in approach in response to a call for contemporary witnesses to share their experiences. In contrast, because this was a pre-existing, predefined cohort, the study team approached a set cohort of individuals and invited them to continue their participation in the study after many decades. Although it is the availability of data from this cohort that provides the opportunity to conduct this longitudinal study, the study presents a number of challenges. Some individuals will be unaware of their former institutional placement, as they might not remember it due to their age at the time or due to repression of memories (Freyd, 1996; Ryffel and Simoni,

2016). This also might be a period of their lives that was not discussed in the family because it was considered stigma or taboo (Sack and Ebbinghaus, 2012; Ryffel and Simoni, 2016; Lengwiler, 2018). Hence, contacting them might disclose this information. A number of measures were put in place to mitigate some of the risks when contacting the individuals, including a comprehensive ethical framework and ethics protocol and psychological support available on site. Furthermore, all contact with the individuals was made solely by senior staff members. This proved essential when having to respond to diverse biographical themes triggered during contact procedures. For example, as a result of their past experiences with authorities, concerns about being purposely misled or judged by the researcher surfaced regularly (Lannen et al., 2020). Working with senior researchers was also key to responding appropriately to any questions, concerns, or potential distress that arose during contact. Earning the trust of the participants was essential. This was achieved chiefly by consistency; the same researcher stayed in contact with the individual throughout the study. It was also achieved by the researchers' efforts to be empathetic no matter what situations came up and to be "humanly available" to the participants (Lannen et al., 2020). Trust was further built by the ability within the research team to deploy researchers with the same linguistic backgrounds as the study participants (Swiss-German, German, Italian, Spanish).

Another important element proved to be the participatory approach to study preparation that included individuals affected by institutionalization as children. Their input significantly shaped the narrative, the approach of the study, and the researchers' skills (Lannen et al., 2020). It also provided an opportunity to respond to the request of those affected to be included in research related to the historic reappraisal process of compulsory social measures and placements before 1981 in Switzerland (UEK Administrative Versorgungen, 2019).

Finally, the project built on Antonovsky (1987) and Werner (2013) findings and a belief that the life trajectory of an individual is determined by many factors of internal and external nature. When asked to provide a recommendation on what would be helpful to formerly institutionalized children in overcoming their experience, Gahleitner, as part of her expert mandate in the round-table discussions related to institutional placement in Germany, stated: "A salutogenetic approach (. . .) allows insights that are usually lost when the focus lies on a search for deficits"[11] (Gahleitner, 2009).

This was expressed in how letters to the participants were formulated and how researchers interacted with them. It was also expressed in the outcome measures included in data collection and an attempt to move away from a solely deficit-oriented narrative in relation to their starting conditions. While prepared for any distress expressed by the participants as a result of past events, the team was also able to acknowledge evidence of resilience and a sense of coherence (Antonovsky, 1987). Overall, the approach attempts to overcome the risk of categorizing the individuals in the group of individuals that started unfavorably and developed poorly (those in infant care) compared to a community sample of individuals that grew up at home and therefore are considered to be the norm (ZLS). The study is embedded in a framework of respect and empathy for each individual's life story. It is open to a multitude of life trajectories between and within each study arm without prejudging the outcomes. At the same time, the framework honors the sense of injustice felt by many of the individuals subjected to welfare practices implemented before 1981. It is an attempt to honor their experiences without passing judgment, and also seek the potential resilience and strength that may have grown from each individual's unique experience.

Methodological Approach

A number of factors fundamentally shaped the methodological approach of the study. These included the feedback on instruments from individuals affected by institutional placement as children during study preparation. Moreover, the research team applied a multidisciplinary and multimethod approach. Leveraging backgrounds in developmental psychology, pediatrics, educational sciences, sociology, and anthropology allowed the team to use validated instruments for assessment of health as well as psychological constructs and combine them with, for example, reconstructive methods. Although this approach required constant discussions and feedback loops in order to negotiate common terminology and framing, it enabled a holistic view of the interplay between individual development and a particular societal and educational context. Furthermore, the expertise combined in quantitative and qualitative approaches enabled the collection of both standardized and narrative data and thus reflect the complexity of the human experience as completely as possible.

However, this study also faced a number of key methodological challenges:

First, the availability of a community sample of individuals that grew up in the same geographic location at the same time (Wehrle et al., 2020) enables us to distinguish historical variables from biological, personal, and social ones and extract generalizable statements and recommendations. However, children placed in institutions in the late 1950s were exposed to multiple risks: both the societal, legal, and familial circumstances that put the children at risk of institutionalization in the first place and the deprivation the children experienced as part of the institutional placement (Meierhofer and Keller, 1974). This is relevant, because other studies have found evidence of a dose–response effect in that the more adverse events a child is exposed to, the more profound and lasting the effect on the individual will be (Felitti et al., 1998; Finkelhor et al., 2007; Huebner et al., 2016). As part of this study, it will be important to distinguish pre-existing risks as much as possible from the impact of institutionalization and possible compensatory effects arising from relationship conditions with parents or better institution quality by including variables in our models that account for possible additive and interaction effects. Fortunately for us, children were generally assigned to institutions independently of variables such as health, development, or potential covariates such as family background. Placement was decided by

[11] Original in German: "*Die salutogenetische Orientierung* (. . .) *bietet Einsichten, die in der reinen Suche nach Defiziten häufig verloren gehen.*"

geographic location or space availability, hence mirroring a quasi-experimental design of children with similar distribution of backgrounds and preconditions in each institution and thus increasing the probability that differences between development outcomes in the different institutions are due to circumstances in the care facility.

Second, individuals in this cohort are organized by institution in 12 clusters. This is a challenge, as 12 clusters is at the very low end, if not below the required amount of clusters from a statistical point of view. The other challenge arises when comparing data with the ZLS, as this multilevel structure of data is only present in one cohort. In addition, some information only exists for the entire institution even though it can be assumed that conditions within the institutions varied between units within a single institution. Combined expertise in the project team, the project board and through collaboration with specific experts will enable us to address this methodological challenge through multilevel modeling or correction of standard errors where multilevel modeling cannot be applied.

Third, there are a number of challenges inherent to longitudinal studies that run for several research generations. These challenges are described in detail in the paper of this issue (Wehrle et al., 2020), and include dealing with archived data and their sometimes spotty documentation, the predefined nature of the cohort or balancing continuity of measurements while keeping up with latest standards.

Fourth, the more time passes, the greater grows the chance of bias as a result of the possible impact of the events on mortality for those more severely affected, as described in seminal literature (Felitti et al., 1998). Furthermore, the results of the preliminary study (Ryffel and Simoni, 2016) also suggest that a self-selection bias might have operated toward participants with more adaptive long-term outcomes. In addition, women and the children of single mothers were over-represented compared to migrant families. Analyzing whether participation in the study is selective in relation to key variables from Wave 1 and 2 is essential.

Research Questions

Mastering these challenges will allow us to address a number of research questions.

Historical data:

(1) Circumstances of placements:
What information can be found on why infants were institutionalized as part of historic welfare practices, and how is this represented in the literature on relevant discourses from that time? What were the conditions and routines like within the institutions?

(2) Health and development of children
Can findings from the historical data be replicated? What additional findings related to the impact of infant institutionalization can be brought to light from the historical data? What additional and relevant insights can be found by comparing Wave 2 to the comparison group of the ZLS? How do familial and institutional contexts relate to children's health outcomes and abilities?

What deficits identified in Wave 1 and Wave 2 can still be detected in individuals 60 years after they were institutionalized as infants?

Newly collected and longitudinal data:

How have the lives of individuals that spent their infancy in institutions evolved? What is their morbidity (physical and mental) and mortality? What are their cognitive, social, and motor abilities today? What are their educational paths and socio-economic conditions? What deficits identified in Waves 1 and 2 can still be detected in individuals 60 years after they were institutionalized as infants? What specific vulnerabilities and/or strengths have emerged during the life course? How have some of the care circumstances and practices related to institutional placement affected long-term outcomes? What are the key individual variables and life events that moderate the relationship between early institutionalization and long-term outcomes? How do individuals talk about their life trajectories? What themes are relevant to them and how do they make sense of what happened?

Outlook and Conclusion

Overall, the study will allow us to better understand the effects of being placed in infant institutions under conditions of deprivation, will improve our knowledge of possible development trajectories, and will contribute to the reconciliation process and historical reappraisal process of compulsory social measures and placements before 1981 in Switzerland. It will reveal important insights for children placed in institutions today and allow us to better understand how children with different starting conditions can be supported in developing healthily. Finally, the present study will provide a unique contribution to our understanding of the interplay between individual and environmental promotive and protective factors for development over decades in individuals at risk.

A number of opportunities exist when considering potential outlooks for the study. The individuals of the LifeStories project are now at the cusp of old age. Some evidence has recently emerged demonstrating that exposure to adverse events in childhood is linked to premature biological aging in adulthood (Shalev et al., 2013; Puterman et al., 2016). However, this data was collected retrospectively. Longitudinal data is needed on how adverse events may affect biological aging.

Consent from all participants of Wave 3 was sought to contact them again in the future for additional prospective assessments. Thus, collecting biological and neuroimaging data will be important to further understanding the specific aging processes in this vulnerable cohort.

In addition, collecting information from additional perspectives will allow us to complement the efforts related to the historic reappraisal process of compulsory social measures and placements before 1981 in Switzerland. To better understand the circumstances that led to infant care placements at the time, the perspective of the parents of formerly institutionalized infants would be essential and is time sensitive due to their advanced age. This also applies to former employees of the institutions. In addition, interviews with children of those placed in institutions are relevant, as emerging evidence shows the

effect such measures may have from a generational perspective (Bombach et al., 2018a,b; UEK Administrative Versorgungen, 2019).

Making the details of the study process available as part of a study protocol allows the scientific community to learn what works to successfully conduct longitudinal studies and shed light onto the secrets of long-term adaptation processes across the lifespan.

DISSEMINATION

The results bear significant potential for scientific, practical, and social impact at national and international levels, and results will be disseminated accordingly.

Understanding the impact that measures related to compulsory social measures and placements before 1981 in Switzerland had on the children during infancy and childhood and the consequences 60 years later will contribute to the historical reappraisal and political reconciliation of such measures in Switzerland. Accessing historical records will enable a data-driven approach to understanding the causes, characteristics, and mechanisms that led to and surrounded infant care practices before 1981. It will allow society to make sense of how practices came about and enable it to uncover the zeitgeist and societal norms and values that framed the actions. The study will illuminate a piece of Swiss institutional history, provide a basis to reflect on today's practice critically, and might sensitize the society to the possible existence of blind spots in today's practices in Switzerland and around the world.

Due to its longitudinal design, the study is highly relevant and one of the few long-term follow-up studies of its kind. Through the existence of a comparison group from the same era (Wehrle et al., 2020), historical variables can be distinguished from biological, personal, and social ones, and generalizable statements and recommendations can be extracted that are relevant for today. This will improve our understanding of the detection and handling of difficult early life conditions and our support of the development of resiliency over a life-course perspective.

It will also improve our understanding of the long-term consequences of deprivation caused by institutionalization of young children. The study will reveal factors mitigating institutional care and protective processes for favorable development trajectories as a result of resiliency processes. It will also provide information on how the professional and policy community can best support children and young people in care and their families across the globe.

To this end, scientific publications focused on the key research questions will be published in peer-reviewed journals related to medicine, psychology and educational and other social sciences, and results will be presented at international scientific conferences of these disciplines. In addition, a significant effort will be made to make results accessible to a non-scientific expert audience, society and the survivors of compulsory social measures and placements before 1981.

Evidence briefs will serve as a key tool for disseminating results to a non-scientific audience. A set of evidence briefs will be published in German with a local focus related to the reappraisal and rehabilitation of compulsory social measures and placements before 1981. Additional briefs that focus on the results of international relevance will be published in German and English and target professionals and key organizations involved in out-of-family placements and a multidisciplinary audience of professionals and policy-makers working with individuals at all stages of life. Further, we will work toward a book publication with some of the historical images and materials available. In addition, roundtable discussions will be held with relevant professionals. Results will be presented at local conferences for a non-scientific audience. Efforts will be made to develop a curriculum that includes our findings in university-level training (psychology, social works, educational sciences) and with children in primary and secondary school settings.

All publications will be made available free of charge to participants if desired.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Faculty of Philosophy at the University of Zurich. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PL is the principal investigator of the study. She drafted the Abstract, Introduction, several sections of the description of Wave 3 (Study Preparations, Eligibility, Comparison with ZLS, Adjustments to COVID-19, and Ethics), the Discussion, and the Dissemination section. She edited all other sections of the manuscript. HSa is completing her doctoral work as part of this study. She prepared the historical data and drafted the section on Design and Analysis for Wave 1 and 2. She also contributed significantly to the section on locating individuals. She provided input to the Discussion section and the section on data collection of Wave 3. FS is a senior researcher in the project and drafted the section on data collection and analysis of Wave 3. He provided input to the discussion section, in particular the section on methodological challenges. IRG is a research assistant in the project. He drafted the section on locating individuals and provided input to the discussion section and the section on contact procedures. CB is a senior researcher in the project. She drafted the section on contact procedures and edited the manuscript. In particular, she provided input to the sections on Study Preparation, Ethics, Discussion, and Dissemination. HSi is a co-investigator of the study. She provided input to the Introduction and Discussion sections and edited a final version of the manuscript. FW is a senior researcher in the Zurich Longitudinal Studies (ZLS). She provided input to the sections about the ZLS and the Discussion section. OJ is a co-investigator of the study, provided input to the sections about the ZLS and the Introduction and Discussion sections, and edited the final version of the manuscript. He is the principal investigator of the Zurich

Longitudinal Studies. All the authors contributed to the article and approved the submitted version.

FUNDING

The project was funded by the Swiss National Science Foundation, under the umbrella of the National Research Programme 76 Welfare and Coercion; it was co-funded by the Maiores Foundation. Additional funds to expand the interview part of the study were provided by the Lotteriefonds of Zurich and the City of Zurich. A contribution to the publication was made by the Grueniger Foundation.

ACKNOWLEDGMENTS

First and foremost, we would like to thank the study participants for taking the time to participate in the study and for sharing their stories with us. We are grateful for the benevolence of the former care institutions, who have supported the study and our staff. Our gratitude goes to all the municipalities and civil registries who have supported us in our efforts to locate individuals. Specifically, we would like to thank Anja Huber at the Zurich City Archive, Markus Stoll from the Civil Registry in Zurich, Jeannine Rauschert at the Federal Archive in Aarau, and Verena Rothenbühler at the Federal Archive in Zurich. We are grateful to Marco Hüttenmoser for granting access to the data and for being available for numerous questions. Gaby Ryffel also supported us by being available for various issues that came up. We would like to thank the students who have supported the study: Corina Conrad, Alexandra Dübendorfer, Astrid Estermann, Naomi Gnägi, Nina Graf, Lea Kolly, Nicole Kradolfer, Linda Moresi, Manuela Rüdiger, Andrea Stamm, and Christina Widmer. Our thanks go to our MMI secretariat staff and our psychology staff for supporting the study. We would also like to thank our expert project board for their support (Matthias Allemand, Andrea Büchler, Moritz Daum, Miriam Gebhardt, Florence Martin, Pasqualina Perrig-Chiello, Delia Pop, Cornelia Rumo, Diana Wider, and Hansjörg Znoj) and the staff of the Zurich Longitudinal Studies Jon Caflisch, Dominique Eichelberger, Giulia Haller, and Tanja Kakebeeke. Our thanks also go to Simon Milligan for editing the manuscript. We are thankful to the funders that made this study possible.

REFERENCES

Ammann, R., Huonker, T., and Schmid, J. (2019). *Gesichter der Administrativen Versorgung. Porträts von Betroffenen. Vol. VOL. 1 Unabhängige Expertenkommission Administrative Versorgung*. Zürich: Chronos.

Antonovsky, A. (1979). *Health, Stress, and Coping*. San Francisco, CA: Jossey-Bass, 278.

Antonovsky, A. (1987). *Unraveling the Mystery of Health: How People Manage Stress and Stay Well. A Joint Publication in the Jossey-Bass Social and Behavioral Science Series and the Jossey-Bass Health Series*, 1st Edn. San Francisco, CA: Jossey-Bass, 218.

Balzer, C., Berger, J.-M., Caprez, G., Gonser, A., Gutbrod, K., and Keller, M. (2011). *Materialen und Normwerte für die Neuropsychologische Diagnostik (MNND)*. Rheinfelden: Verlag Normdaten.

Bartholomew, K., and Horowitz, L. M. (1991). Attachment styles among young adults: a test of a four-category model. *J. Pers. Soc. Psychol.* 61, 226–244. doi: 10.1037/0022-3514.61.2.226

Bastin, G. (1967). *Die Soziometrischen Methoden*. Bern: Huber.

Beierlein, C., Kovaleva, A., Kemper, C. J., and Rammstedt, B. (2012). *ASKU – Allgemeine Selbstwirksamkeit Kurzskala*. Available online at: https://psycharchives.zpid.de/handle/20.500.12034/431 (accessed December 17, 2020).

Belschner, W., Lischke, G., and Selg, H. (1971). *Foto-Hand-Test (FHT) zur Erfassung der Aggressivität*. Freiburg: Alber.

Berens, A. E., and Nelson, C. A. (2015). The science of early adversity: is there a role for large institutions in the care of vulnerable children? *Lancet* 386, 388–398. doi: 10.1016/S0140-6736(14)61131-4

Biäsch, H., and Fischer, H. (1969). *Testreihen zur Prüfung von Schweizer Kindern vom 4.-15. Altersjahr. 2*. Bern: Huber.

Bick, J., Zhu, T., Stamoulis, C., Fox, N. A., Zeanah, C., and Nelson, C. A. (2015). Effect of early institutionalization and foster care on long-term white matter development: a randomized clinical trial. *JAMA Pediatr.* 169, 211–219. doi: 10.1001/jamapediatrics.2014.3212

Bohm, E. (1965). *Lehrbuch der Rorschach-Psychodiagnostik*. Bern: Huber.

Bohnsack, R., Nentwig-Gesemann, I., and Nohl, A.-M.(eds). (2013). *Die dokumentarische Methode und ihre Forschungspraxis. Grundlagen Qualitativer Sozialforschung. 3*. Wiesbaden: Springer VS. doi: 10.1007/978-3-531-19895-8

Bombach, C., Gabriel, T., and Keller, S. (2018a). "Legitimieren" und "integrieren" – Die Auswirkungen von Heimerfahrungen auf den weiteren Lebensverlauf," in *Fremdplatziert Heimerziehung in der Schweiz, 1940–1990*, eds G. Hauss, T. Gabriel, and M. Lengwiler (Zürich: Chronos), 253–272.

Bombach, C., Gabriel, T., and Keller, S. (2018b). "Vulnerabilität und Anerkennung. Erzählte Biografie nach Heimplatzierungen zwischen 1950 und 1990," in *Zwischen Erinnerung und Aufarbeitung Fürsorgerische Zwangsmassnahmen an Minderjährigen in der Schweiz im 20 Jahrhundert*, eds B. Ziegler, G. Hauss, and M. Lengwiler (Zürich: Chronos), 83–109.

Bombach, C., Gabriel, T., Keller, S., Ramsauer, N., and Marx, A. S. (2017). *Zusammen Alleine: Alltag in Winterthurer Kinder- und Jugendheimen 1950–1990*. Zürich: Chronos, 224.

Borkenau, P., and Ostendorf, F. (2008). *NEO-FFI. NEO-Fünf-Faktoren-Inventar nach Costa und McCrae. 2., neu normierte und vollständig überarbeitete Auflage*. Göttingen: Hogrefe.

Brandt, J., and Benedict, R. H. B. (2001). *Hopkins Verbal Learning Test – Revised. Administration Manual*. Lutz, FL: Psychological Assessment Resources.

Brauer, S. (2019b). *Gutachten zum Geplanten NFP 76-Forschungsprojekt "The Impact of Infant Institutionalization in Switzerland – a 60-year follow up" des Marie Meierhofer Instituts für das Kind*. Zurich: Brauer & Strub.

Brauer, S. (2019a). *Ethische Einschätzung bezüglich der Einbindung von Probanden in den beiden Marie Meierhofer-Studien zur Entwicklung von Heimkindern von 1958–1961 und 1971–1973*. Zurich: Brauer & Strub.

Browne, K. (2009). *The Risk of Harm to Young Children in Institutional Care*. London: Save the children.

Brunet, O., and Lézine, I. (1955). *Échelle de Développement Psychomoteur de la Première Enfance*. Issy-les-Moulineaux: Editions scientifiques et psychotechniques.

Buysee, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry Res.* 2, 193–213. doi: 10.1016/0165-1781(89)90047-4

Center on the Developing Child (2007). *The Impact of Early Adversity on Child Development*. Available online at: www.developingchild.harvard.edu (accessed December 17, 2020).

Dahl, G. (1968). *Übereinstimmungsvalidität des HAWIE und Entwicklung Einer Reduzierten Testform*. Meisenheim (am Glan): Hain.

Dahl, G. (1972). *Reduzierter Wechsler-Intelligenztest. WIP*. Meisenheim (am Glan): Hain.

Duhm, E., and Hansen, J. P. F. T. (1957). *Der Rosenzweig-P-F Test. Deutsche Bearbeitung der Rosenzweig "Picture Frustration Study", Form für Kinder*. Göttingen: Verlag für Psychologie Dr. C. J. Hogrefe.

D'Amato, G. (2012). "Die durchleuchtete, unsichtbare Arbeitskraft. Die italienische Einwanderung in die Schweiz in den 50er Jahren," in *Schöner Leben, Mehr Haben Die 50er Jahre in der Schweiz im Geiste des Konsums*, ed. T. Pfunder (Zürich: Limmat), 237–252.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., et al. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The adverse childhood experiences (ACE) study. *Am. J. Prev. Med.* 14, 245–258. doi: 10.1016/S0749-3797(98)00017-8

Finkelhor, D., Ormrod, R. K., and Turner, H. A. (2007). Poly-victimization: a neglected component in child victimization. *Child Abuse Negl.* 31, 7–26. doi: 10.1016/j.chiabu.2006.06.008

Franke, G. H. (2017). *BSCL. Brief-Symptom-Checklist. Deutsche Version*. Göttingen: Hogrefe.

Freyd, J. J. (1996). *Betrayal Trauma: The Logic of Forgetting Childhood Abuse*. Cambridge, MA: Harvard University Press, 232.

Fydrich, T., Sommer, G., and Brähler, E. (2007). *F-SozU. Fragebogen zur Sozialen Unterstützung*. Göttingen: Hogrefe.

Gahleitner, S. B. (2009). *Was Hilft Ehemaligen Heimkindern bei der Bewältigung ihrer Komplexen Traumatisierung? Expertise im AUFTRAG des Runden Tisches Heimerziehung*. Berlin: Silke Birgitta Gahleitner.

Gebhardt, M. (2009). *Die Angst vor dem Kindlichen Tyrannen?: eine Geschichte der Erziehung im 20. Jahrhundert*. München: Deutsche Verlags-Anstalt (DVA).

Glaesmer, H., Grande, G., Braehler, E., and Roth, M. (2011). The german version of the satisfaction with life scale (SWLS): psychometric properties, validity, and population-based norms. *Eur. J. Psychol. Assess.* 27, 127–132. doi: 10.1027/1015-5759/a000058

Goodenough, F. L. (1926). *Measurement of Intelligence by Drawings*. Chicago: World Book Co.

Grosse Holtforth, M., and Grawe, K. (2003). Der Inkongruenzfragebogen (INK). *Z. Für Klin. Psychol. Psychother.* 32, 315–323. doi: 10.1026/0084-5345.32.4.315

Hamaker, E. L., Kuiper, R. M., and Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychol Methods* 20, 102–116. doi: 10.1037/a0038889

Huber, D. (1995). Ein halbes Jahrhundert in Erwartung. Zur Geschichte der in der Schweiz bis heute fehlenden Mutterschaftsversicherungen. *Olympe Fem. Arbeitshefte Zur Polit.* 3, 92–99.

Huebner, G., Boothby, N., Aber, L., Darmstadt, G., Diaz, A., Masten, A. S., et al. (2016). *Beyond Survival: The Case for Investing in Young Children Globally*. Washington, DC: National Academy of Medicine.

Johnson, R., Browne, K., and Hamilton-Giachritsis, C. (2006). Young children in institutional care at risk of harm. *Trauma Violence Abuse* 7, 34–60. doi: 10.1177/1524838005283696

Joris, E. (2012). "Dezenter sex appeal – eklatante Diskriminierung. Weiblichkeits- und Männlichkeitsmythen in Zeiten von Textilrevolution und kaltem Krieg," in *Schöner Leben, Mehr Haben Die 50er Jahre in der Schweiz im Geiste des Konsums*, ed. T. Pfunder (Zürich: Limmat), 105–119.

Kakebeeke, T., Knaier, E., Chaouch, A., Caflisch, J., Rousson, V., and Largo, R. H. (2018). Neuromotor development in children. Part 4: new norms from 3 to 18 years. *Dev. Med. Child Neurol.* 60, 810–819. doi: 10.1111/dmcn.13793

Koch, K. (1957). *Der Baumtest. Der Baumzeichenversuch als Psychodiagnostisches Hilfsmittel*. Bern: Huber.

Kreicbergs, U., Valdimarsdóttir, U., Steineck, G., and Henter, J.-I. (2004). A population-based nationwide study of parents' perceptions of a questionnaire on their child's death due to cancer. *Lancet* 364, 787–789. doi: 10.1016/S0140-6736(04)16939-0

Lannen, P., Bombach, C., and Jenni, O. G. (2020). Using participatory methods to develop and implement research on historic compulsory social measures and placements in Switzerland. *Int. J. Child Youth Fam. Stud.* 11, 96–120. doi: 10.18357/ijcyfs114202019940

Lannen, P., Wolfe, J., Mack, J., Onelov, E., Nyberg, U., and Kreicbergs, U. (2010). Absorbing information about a child's incurable cancer. *Oncology* 78:259. doi: 10.1159/000315732

Lannen, P., Wolfe, J., Prigerson, H. G., Onelov, E., and Kreicbergs, U. C. (2008). Unresolved grief in a national sample of bereaved parents: impaired mental and physical health 4 to 9 years later. *J. Clin. Oncol.* 26, 5870–5876. doi: 10.1200/JCO.2007.14.6738

Lengwiler, M. (2018). "Aufarbeitung und Entschädigung traumatisierender Fremdplatzierungen. Die Schweiz im internationalen Vergleich," in *Zwischen Erinnerung und Aufarbeitung Fürsorgerische Zwangsmassnahmen an Minderjährigen in der Schweiz im 20 Jahrhundert*, eds B. Ziegler, G. Hauss, and M. Lengwiler (Zürich: Verlag), 159–176.

Lengwiler, M., and Praz, A.-F. (2018). "Kinder- und Jugendfürsorge in der Schweiz. Entstehung, Implementierung und Entwicklung (1900-1980)," in *Fremdplatziert Heimerziehung in der Schweiz, 1940-1990*, eds G. Hauss, T. Gabriel, and M. Lengwiler (Zürich: Chronos), 29–52.

Limesurvey GmbH (2020). *LimeSurvey: An Open Source Survey Tool [Internet]*. Hamburg: LimeSurvey GmbH.

Loch, N., Hiller, W., and Witthöft, M. (2011). Der cognitive emotion regulation questionnaire (CERQ) erste teststatistische Überprüfung einer deutschen Adaption. *Z. Für Klin. Psychol. Psychother.* 40, 94–106. doi: 10.1026/1616-3443/a000079

Lysholm, J., and Lindahl, B. (2019). Strong development of research based on national quality registries in Sweden. *Ups J. Med. Sci.* 124, 9–11. doi: 10.1080/03009734.2018.1520761

Mayring, P. (2000). "Qualitative Inhaltsanalyse," in *Qualitative Forschung Ein Handbuch*, eds U. Flick, E. von Kardorff, and I. Steinke (Reinbeck: Rowohlt).

Meierhofer, M. (1958). Pädagogische Probleme der Säuglings – und Kleinkinderpflege. *Pro Juventute* 10, 1–3.

Meierhofer, M., and Keller, W. (1974). *Frustration im Frühen Kindesalter. 3. Aufl.* Bern: Hans Huber.

Meyer-Schell, I. (1971). *Nachuntersuchung von Sechzehn Schulkindern, die ihre frühe Kindheit in Einem Heim Verbrachten [Lizenzarbeit]*. Zürich: Universität Zürich.

Morfeld, M., Kirchberger, I., and Bullinger, M. (2011). *SF-36. Fragebogen zum Gesundheitszustand. 2., Ergänzte und Überarbeitete Auflage*. Göttingen: Hogrefe.

Nelson, C. A., Fox, N. A., and Zeanah, C. H. (2014). *Romania's Abandoned Children: Deprivation, Brain Development, and the Struggle for Recovery*. Cambridge, MA: Harvard University Press, 402. doi: 10.4159/harvard.9780674726079

Neumann, E. (2002). "KBEK – Kurzmessung der Bindung an die Eltern in der Kindheit," in *Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Hrsg.), Elektronisches Testarchiv*. Trier: ZPID.

Paulus, C. (2011). *Der Saarbrücker Persönlichkeitsfragebogen zur Messung von Empathie (SPF-IRI)*. Available online at: http://bildungswissenschaften.uni-saarland.de/personal/paulus/homepage/empathie.html (accessed December 17, 2020).

Petermann, F., and Petermann, U. (2013). *WAIS-IV. Wechsler Adult Intelligence Scale – Fourth Edition. Deutschsprachige Adaptation der WAIS-IV von D. Wechsler*. Frankfurt/Main: Pearson Assessment.

Puterman, E., Gemmill, A., Karasek, D., Weir, D., Adler, N. E., Prather, A. A., et al. (2016). Lifespan adversity and later adulthood telomere length in the nationally representative US health and retirement study. *Proc. Natl. Acad. Sci. U.S.A.* 113, E6335–E6342. doi: 10.1073/pnas.1525602113

R Core Team (2018). *R: A Language and Environment for Statistical Computing [Internet]*. Vienna: R Foundation for Statistical Computing.

Ramsauer, N. (2000). *"Verwahrlost" Kindswegnahmen und die Entstehung der Jugendfürsorge im schweizerischen Sozialstaat 1900-1945*. Zürich: Chronos.

Rey, A. (1941). L'examen psychologique dans les cas d'encephalopathie traumatique. *Arch. Psychol.* 28, 286–340.

Rosenthal, G. (1995). *Erlebte und Erzählte Lebensgeschichte. Gestalt und Struktur biographischer Selbstbeschreibungen*. Frankfurt: Campus Verlag.

Rosenthal, G. (2015). "Interpretative Sozialforschung. Eine Einführung. 5., aktualisierte und ergänzte Auflage," in *Grundlagentexte Soziologie*, eds M. Diewald and K. Hurrelmann (Weinheim: Beltz Juventa).

Ryffel, G. (2013). *Resilienz und Defizit. Entwicklung nach Einem Frustrierenden Start im Säuglingsheim*. Zürich: Marie Meierhofer Institut für das Kind.

Ryffel, G., and Simoni, H. (2016). *Forschungsdokumentation zum Projekt Lebensverläufe nach Strukturell Bedingter Frustration der Frühesten Entwicklungsbedürfnisse unter dem Aspekt der Resilienz. Retrospektive Teilstudie des Projektes Leben mit Brüchen in der Frühen Beziehungsumwelt*. Zürich: Marie Meierhofer Institut für das Kind.

Sack, M., and Ebbinghaus, R. (2012). "Was hilft ehemaligen Heimkindern der DDR bei der Bewältigung ihrer komplexen Traumatisierung?," in *Aufarbeitung der*

Heimerziehung in der DDR Expertisen, ed. Beauftragter der Bundesregierung für die Neuen Bundesländer (Berlin: Arbeitsgemeinschaft für Kinder und Jugendhilfe – AGJ), 299–395.

Sand, H., and Gruber, L. (2017). *Leben in Brüchen: Thematische Inhaltsanalyse der Interviews. (Projektarbeit im Rahmen des Masterstudiengangs Frühe Kindheit, Pädagogische Hochschule Thurgau)*. Zürich: Marie Meierhofer Institut für das Kind.

Schallberger, U., and Spiess Huldi, C. (2001). Die Zürcher Längsschnittstudie von der Schulzeit bis zum mittleren Erwachsenenalter. *Z Für Soziol. Erzieh Sozial.* 21, 80–89.

Schmidt, S., Muehlan, H., and Brähler, E. (2016). *AAS-R. Revised Adult Attachment Scale. Deutsche Version*. Göttingen: Hogrefe.

Schore, A. N. (2001). The effects of early relational trauma on right brain development, affect regulation, and infant mental health. *Infant Ment. Health J.* 22, 201–269. doi: 10.1002/1097-0355(200101/04)22:1<201::AID-IMHJ8>3.0.CO;2-9

Schumacher, J., Leppert, K., Gunzelmann, T., Strauß, B., and Brähler, E. (2004). Die Resilienzskala. Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Personmerkmal. *Z. Für Klin. Psychol. Psychiatr. Psychother.* 53, 16–39.

Schumacher, J., Wilz, G., Gunzelmann, T., and Brähler, E. (2000). Die sense of coherence scale von antonovsky. Teststatistische Überprüfung in einer repräsentativen Bevölkerungsstichprobe und Konstruktion einer Kurzskala. *Psychother. Psychosom. Med. Psychol.* 50, 472–482. doi: 10.1055/s-2000-9207

Shalev, I., Entringer, S., Wadhwa, P. D., Wolkowitz, O. M., Puterman, E., Lin, J., et al. (2013). Stress and telomere biology: a lifespan perspective. *Psychoneuroendocrinology* 38, 1835–1842. doi: 10.1016/j.psyneuen.2013.03.010

Sherr, L., Roberts, K. J., and Gandhi, N. (2017). Child violence experiences in institutionalised/orphanage care. *Psychol. Health Med.* 22, 31–57. doi: 10.1080/13548506.2016.1271951

Shonkoff, J. P., and Phillips, D.A. (eds). (2000). *From Neurons to Neighborhoods: the Science of Early Child Development*. Washington, DC: National Academies Press.

Sticca, F., and Perren, S. (2015). The chicken and the egg: Longitudinal associations between moral deficiencies and bullying: a parallel process latent growth model. *Merrill Palmer Q.* 61, 85–100. doi: 10.13110/merrpalmquar1982.61.1.0085

Strauss, E., Sherman, E. M. S., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. 3*. New York, NY: Oxford University Press.

Surkan, P. J., Steineck, G., and Kreicbergs, U. (2008). Perceptions of a mental health questionnaire: the ethics of using population-based controls. *J. Med. Ethics* 34, 545–547. doi: 10.1136/jme.2007.022418

Terman, L. M., and Merrill, M. A. (1937). *Measuring Intelligence*. New York, NY: Houghton Mifflin.

Terman, L. M., and Merrill, M. A. (1965). *Stanford-Binet Intelligenztest. S-I-T. Deutsche Bearbeitung von Prof. Dr. Heinz-Rolf Lückert. Handanweisung*. Göttingen: Verlag für Psychologie Dr. C. J. Hogrefe.

Thalmann, H. C. (1971). *Verhaltensstörungen bei Kindern im Grundschulalter*. Stuttgart: Klett.

Thurner, F., and Tewes, U. (1969). *Kinder-Angst-Test (KAT)*. Göttingen: Hogrefe.

Tombaugh, T. N. (2004). Trail making Test A and B: normative data stratified by age and education. *Arch. Clin. Neuropsychol.* 19, 203–214. doi: 10.1016/S0887-6177(03)00039-8

UEK Administrative Versorgungen (2019). *UEK Administrative Versorgungen – Unabhängige Expertenkommission zur Wissenschaftlichen Aufarbeitung der Administrativen Versorgungen [Internet]*. Available online at: http://www.uek-administrative-versorgungen.ch/de/Startseite.1.html (accessed September 4, 2017).

Ulijaszek, S. J., Johnston, F. E., Preece, M. A., and Tanner, J. (1998). *The Cambridge Encyclopedia of Human Growth and Development*. Cambridge: Cambridge University Press, 509.

Unabhängige Expertenkommission Administrative Versorgungen (2019). *Organisierte Willkür. Administrative Versorgungen in der Schweiz 1930-1981. Schlussbericht. Vol. VOL. 10 A*. Zürich: Chronos.

Ungricht, J. (1955). Der Sohnaufsatz als psychodiagnostisches Hilfsmittel. *Schweiz Z. Für Psychol. Ihre Anwendungen.* 14, 27–45.

von Staabs, G. (1964). *Der Scenotest. Beitrag zur Erfassung unbewusster Problematik und charakterologischer Struktur in Diagnostik und Therapie*, 4th Edn. Bern: Hans Huber.

Wehrle, F. M., Caflisch, J. A., Eichelberger, D., Kakebeeke, T. H., Latal, B., and Jenni, O. G. (2020). Health and development across the lifespan – the Zurich longitudinal studies. *Front. Hum. Neurosci. Res. Top. Longitud. Aging Res. Cogn. Behav. Neurosci.*

Werner, E. E. (2013). "What can we learn about resilience from large-scale longitudinal studies?," in *Handbook of Resilience in Children*, eds S. Goldstein and R. B. Brooks (Boston, MA: Springer), 87–102. doi: 10.1007/978-1-4614-3661-4_6

Wyss-Wanner, M. (2000). *Ein Leben für Kinder: Leben und Werk von Marie Meierhofer, 1909-1998. (Zürcher medizingeschichtliche Abhandlungen)*. Zürich: Juris Druck + Verlag.

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.

The reviewer MM declared a shared affiliation, though no other collaboration with the authors to the handling editor.

frontiers in Human Neuroscience

STUDY PROTOCOL
published: 28 January 2021
doi: 10.3389/fnhum.2020.612453

The Importance of Childhood for Adult Health and Development—Study Protocol of the Zurich Longitudinal Studies

*Flavia M. Wehrle[1,2], Jon Caflisch[1,2], Dominique A. Eichelberger[1], Giulia Haller[1], Beatrice Latal[1,2,3], Remo H. Largo[1†], Tanja H. Kakebeeke[1,2] and Oskar G. Jenni[1,2,3]**

[1] Child Development Center, University Children's Hospital Zurich, Zurich, Switzerland, [2] Children's Research Center, University Children's Hospital Zurich, Zurich, Switzerland, [3] University of Zurich, Zurich, Switzerland

Evidence is accumulating that individual and environmental factors in childhood and adolescence should be considered when investigating adult health and aging-related processes. The data required for this is gathered by comprehensive long-term longitudinal studies. This article describes the protocol of the Zurich Longitudinal Studies (ZLS), a set of three comprehensive cohort studies on child growth, health, and development that are currently expanding into adulthood. Between 1954 and 1961, 445 healthy infants were enrolled in the first ZLS cohort. Their physical, motor, cognitive, and social development and their environment were assessed comprehensively across childhood, adolescence, and into young adulthood. In the 1970s, two further cohorts were added to the ZLS and assessed with largely matched study protocols: Between 1974 and 1979, the second ZLS cohort included 265 infants (103 term-born and 162 preterm infants), and between 1970 and 2002, the third ZLS cohort included 327 children of participants of the first ZLS cohort. Since 2019, the participants of the three ZLS cohorts have been traced and invited to participate in a first wave of assessments in adulthood to investigate their current health and development. This article describes the ZLS study protocol and discusses opportunities, methodological and conceptual challenges, and limitations arising from a long-term longitudinal cohort recruited from a study about development in early life. In the future, the ZLS will provide data to investigate childhood antecedents of adult health outcomes and, ultimately, will help respond to the frequent call of scientists to shift the focus of aging research into the first decades of life and, thus, to take a lifespan perspective on aging.

Keywords: longitudinal study, study protocol, development, growth, childhood, lifespan, aging

OPEN ACCESS

Edited by:
Lutz Jäncke,
University of Zurich, Switzerland

Reviewed by:
Sascha Zuber,
Université de Genève, Switzerland
Thomas Halvorsen,
Haukeland University Hospital, Norway

***Correspondence:**
Oskar G. Jenni
oskar.jenni@kispi.uzh.ch

†*Deceased on November 11th 2020*

Specialty section:
This article was submitted to Health, a section of the journal Frontiers in Human Neuroscience

Received: *30 September 2020*
Accepted: *23 December 2020*
Published: *28 January 2021*

Citation:
Wehrle FM, Caflisch J, Eichelberger DA, Haller G, Latal B, Largo RH, Kakebeeke TH and Jenni OG (2021) The Importance of Childhood for Adult Health and Development—Study Protocol of the Zurich Longitudinal Studies. Front. Hum. Neurosci. 14:612453. doi: 10.3389/fnhum.2020.612453

INTRODUCTION

The Zurich Longitudinal Studies (ZLS) are a set of comprehensive studies on child and adolescent growth, health, and development that are currently expanding into adulthood. This article describes the ZLS study protocol and discusses opportunities, methodological and conceptual challenges, and limitations arising from these cohorts; the participants were recruited for this long-term

longitudinal study from a previous study about development in early life. In the future, the ZLS will provide data to investigate childhood antecedents of adult health outcomes and, ultimately, will help respond to the frequent call of scientists to shift the focus of aging research into the first decades of life and, thus, to take a lifespan perspective on aging (Sanders, 2016; Moffitt et al., 2017; Williamson and Leroi, 2019).

Childhood and Adolescence Are Important for Adult Health and Aging-Related Processes

Evidence is accumulating that early life requires consideration when investigating adult health and aging-related processes. In fact, even prenatal life has been shown to be relevant in this regard: for example, adults who were exposed to *in utero* malnutrition during the Dutch famine in 1944/1945 were reported to suffer from higher rates of coronary heart disease, higher levels of stress reactivity, and poorer mental health and premature brain aging (Roseboom et al., 2006; Franke et al., 2018; Van den Broek and Fleischmann, 2019). Moreover, adult outcome is known to be shaped by the socioeconomic and psychosocial environment of childhood. Individuals who grew up in a suboptimal environment have been found to experience poorer health as adults, independent of their current living conditions; this includes an increased risk of various chronic conditions and poorer mental well-being in mid-adulthood (McCrory et al., 2015; Stafford et al., 2015), lower cognitive abilities (Cermakova et al., 2018; Aartsen et al., 2019), more subjective memory complaints (Nishizawa et al., 2019), an increased risk of frailty (Gale et al., 2016), and generally less successful aging (Brandt et al., 2012) in late adulthood. Other studies have focused on how characteristics of the child, most prominently cognitive abilities and personality traits, are related to adult health outcome and aging. Interestingly, lower childhood and adolescent intelligence has been found to predict more advanced biological age in mid- and late adulthood (Schaefer et al., 2015; Stevenson et al., 2019). Potentially, this explains the increased rates of later-life morbidity and mortality in individuals with lower IQ earlier in life (Batty et al., 2007; Schaefer et al., 2015). Furthermore, complex links have been described between personality traits in childhood and adolescence and adult outcome: For example, childhood sociability was found to be related to better subjective well-being and family relationships in mid-adulthood, and better relationships in mid-adulthood and childhood conscientiousness were predictive of a longer life (Kern et al., 2014).

These exemplary findings demonstrate how early life may shape adult health and development and, ultimately, aging-related processes. Appropriate long-term longitudinal studies are required to advance this understanding.

Examples of Long-Term Longitudinal Studies

Although a comprehensive overview is beyond the scope of this article, this section describes some well-known longitudinal studies, all of which have provided insights into how early life impacts adult health, development, and aging-related processes.

A number of birth cohort studies have been ongoing for decades and accordingly have tracked their participants from early age into adulthood. For example, the MRC National Survey of Health and Development (NSHD), the oldest of several British birth cohort studies, was initiated in 1946 and included more than 5,000 infants. The focus of the study has shifted from an initial interest in the costs associated with pregnancy and birth to how family and environmental factors impact child growth, educational attainment, and cognitive development. As participants have aged, pathways to physical and cognitive aging have become of increasing interest (for a comprehensive cohort profile, see Wadsworth et al., 2006). Currently, the NSHD cohort members are in their eighth decade of life and are still being followed (Kuh, 2016; Kuh et al., 2016). In the early 1970s, the Dunedin Multidisciplinary Health and Development Study enrolled more than 1,000 infants born in the city of Dunedin, New Zealand. They have been tested repeatedly across childhood, adolescence, and adulthood with comprehensive assessments of their health and developmental status at each time-point. Most recently, the 45-year follow-up was completed (Poulton et al., 2015). Although participants have not yet reached old age, the data of the Dunedin study has already contributed to longitudinal aging research by quantifying the pace of aging in mid-adulthood from parameters of biological age (Belsky et al., 2015).

The Fels Longitudinal Study was initiated in Ohio, USA, in 1929 and aimed to describe normal child growth and development. It enrolled 15 to 20 individuals annually over the course of 70 years, either prenatally or at birth, and followed them across life (Roche, 1992). The data continues to provide insights into early-life antecedents of chronic diseases in adulthood (e.g., Sun et al., 2008; Sabo et al., 2017). The Intergenerational Studies (IGS) comprise datasets of three studies (the Berkeley Growth Study, the Berkeley Guidance Study, and the Oakland Growth Study) on typical patterns of child and adolescent development. They were initiated in the late 1920s and early 1930s in California, USA. The studies used different protocols to follow their participants into young adulthood. All assessed comprehensive data on physical, cognitive, and social development. To increase statistical power, the three studies were then merged, and collaborative follow-up assessments of the 500 individuals were added to the dataset in mid- and late adulthood (see e.g., Grimm et al., 2011 for details on the study protocols). The Lothian Birth Cohorts of 1921 and 1936 include individuals who had participated in the Mental Surveys of 1932 and 1947, two Scottish surveys that during the months of June 1932 and 1947 tested the intellectual abilities of all children born in 1921 and 1936, respectively. More than 1,500 individuals residing in Edinburgh and the surrounding area were traced seven decades later and have been assessed repeatedly since to study differences in and causes of cognitive aging (Deary et al., 2012). The Terman Study of the Gifted, another study set in California, included more than 1,000 school-aged children born in the 1910s with an IQ above 135 and assessed their health and development across childhood, adolescence, and adulthood (Frey, 2018). A new team of researchers then linked the archived datasets with publicly

available death records of these individuals and, thus, established a lifespan dataset (Friedman and Martin, 2011).

Reviving old studies has remained popular to date: Currently, several research groups are working on identifying former participants of childhood studies and following them up in adulthood. For example, the Louisville Twin Study, a study on child development related to multiple birth status, is currently piloting the assessment of their former members, who are now middle-aged (Davis et al., 2019). Moreover, a cohort of individuals who were placed into infant care institutions in the 1950s in Switzerland are currently being tracked and invited to participate in a study to describe their life stories (Lannen et al., 2021).

The ZLS joins these long-term longitudinal studies in the quest of understanding the importance of early life for later health and development: The ZLS—initiated in the mid-1950s—are a unique set of three cohort studies born two decades apart that assess the development across a range of domains in more than 1,000 individuals. They span the time between birth and the transition to old age and include typically developing individuals and individuals at risk for neurodevelopmental impairments (i.e., born preterm) and dyads of parents and their children. This makes the ZLS highly valuable for assessing early-life antecedents of adult health and aging in different generations.

The Zurich Longitudinal Studies – a Brief History

The International Children's Center (ICC) Coordinated Longitudinal Studies

In the first half of the 20th century, the majority of studies on child health and development were initiated in North America and were largely shaped by the US child welfare movement at the time (Tanner, 1981, 1998). In Europe, the first cohort studies were initiated only after WWII. For instance, in 1951, the Institute of Child Health at the University of London started a multidisciplinary study on growth and development from birth to maturity under the guidance of the pediatrician Alan Moncrieff (1901–1971). About 2 years later, the Center Internationale de L'Enfance (International Children's Center, ICC), an institution founded in 1950 by the French government and UNICEF, began a similar endeavor in Paris under the direction of Robert Debré (1883–1978). In 1954, these two eminent child health experts joined with the American pediatrician Frank Falkner (1918–2003) to initiate a set of harmonized European longitudinal studies on child health and development: the ICC Coordinated Longitudinal Studies. In the following years, more cohorts were added from Zurich, Brussels, Stockholm and two centers in Africa (Dakar, Senegal and Kampala, Uganda). Falkner contributed broad expertise in longitudinal research as coordinator of the Fels study and became the central figure of the ICC Studies. By 1957, Falkner had moved from the Fels Research Institute in Yellow Springs, Ohio to Louisville, Kentucky. Louisville was the home of the Louisville Twin Study, still regarded as the largest and most comprehensive twin study worldwide (Davis et al., 2019). Subsequently, the Louisville cohort joined the ICC cohorts.

The ICC studies aimed to assess child health and development in various cultures and included various disciplines such as pediatrics, psychology, education, and social sciences. The ICC studies may be considered the first multidisciplinary longitudinal cohort studies on child health and development to include several nations across Europe and Africa. The ICC studies were concluded when the participants reached young adulthood in the mid-1970s; however, some of the involved centers had ceased data acquisition before then.

The ZLS as Part of the ICC Coordinated Studies and Beyond

The first principal investigator of the Zurich branch of the ICC studies and the first ZLS cohort (ZLS-1) was the renowned Swiss pediatrician Guido Fanconi (1892–1979), medical director of the University Children's Hospital Zurich, Switzerland. Besides collaborating with the investigator of the other branches of the ICC studies, he was also in close contact with Dr. Marie Meierhofer, the city physician of Zurich, who initiated a study on the health and development of individuals placed in infant care institutions at that time (see Lannen et al., 2021 for details of that study). Fanconi was succeeded in 1962 by pediatric endocrinologist Andrea Prader (1919–2001), who was keenly interested in understanding the growth mechanisms of children and adolescents. Ultimately, Prader was able to turn ZLS-1 into the most successful of all the ICC studies – as the ZLS-1 became the largest cohort to have been followed up from birth to maturity (Tanner, 1981, 1998). In 1974, the developmental pediatrician Remo H. Largo (1943–2020) continued the ZLS and aimed to expand the scope of the studies beyond the understanding of growth mechanisms. Thus, he added two additional cohorts, largely following the initial protocol of ZLS-1: ZLS-2 enrolled children born preterm and at term, and ZLS-3, the Generation Study, enrolled children of the participants of ZLS-1. Although the vast majority of the ZLS-2 and ZLS-3 participants reached young adulthood in the 1990s and early 2000s, the assessments were only completed in 2020, when the youngest ZLS-3 participant turned 18 years old. The data assessed within the scope of these three studies will together be termed ZLS-Childhood. Since 2005, the last author of this article (OGJ) has been the principal investigator of the ZLS.

Aims and Selected Findings of ZLS-Childhood

The aims of ZLS-Childhood were (1) to describe developmental trajectories of individual children in the physical, motor, cognitive, and social domains from birth into young adulthood in three independent cohorts born approximately two decades apart; (2) to understand the mechanisms of growth and development through longitudinal analysis; (3) to compare patterns of typical and atypical development (e.g., in preterm children); and (4) to investigate generational effects between parents and their children. Selected findings that have emerged from the ZLS-Childhood datasets to date are presented in the following section.

Intraindividual trajectories of child and adolescent developmental were a key interest of many analyses. For example, the most innovative statistical techniques at that time

were applied to understand growth patterns and mechanisms across development; **Figure 1A** illustrates growth velocity and height acceleration in individual participants of the ZLS-1 using spline functions and kernel estimation (Largo et al., 1978; Gasser et al., 1984). Further, infant motor milestones were found to be poor predictors of neuromotor and intellectual functions at school age and beyond (Jenni et al., 2013), and neuromotor development between early school age and young adulthood was shown to exhibit only moderate stability whereas intellectual abilities and growth variables were more stable (Jenni et al., 2011). Intraindividual developmental trajectories were also described for behavioral variables [e.g., sleep behavior (Jenni et al., 2007)].

Moreover, the development of term- and preterm-born children was compared in detail and, for instance, differences in language development in the first 5 years were described (Largo et al., 1986). Further, comparing the different ZLS cohorts provided insights into the effect of parenting on development: For instance, the timing and intensity of toilet-training changed markedly between the cohorts of ZLS-1 (early and intensive training) and ZLS-2 (later and less intensive training) whereas the development of becoming dry and clean did not differ between the two generations. The conclusion drawn was that the development of bowel and bladder control is a maturational process that cannot be accelerated by early onset and high intensity toilet-training (Largo et al., 1996).

Clinical Applications and Further Studies From ZLS-Childhood

Over the years, a number of clinical applications have been developed with ZLS-Childhood data. For example, the personal research interest of the principal investigator Remo H. Largo led to the development of a test battery to assess neuromotor function in clinical practice: the Zurich Neuromotor Assessment (ZNA; Largo et al., 2001a,b). The test battery was recently updated and is applied in the current assessment wave in adulthood (ZNA-2; Kakebeeke et al., 2018). BoneXpert©, another frequently used clinical tool, was developed by Danish researchers from the hand X-rays of ZLS-1 participants. This enabled the automatic determination of bone age (Thodberg et al., 2009). Further, reference charts have been published for various developmental domains and continue to inform health care professionals and parents to date [e.g., on growth (Prader et al., 1989), sleep duration—see also **Figure 1B** (Iglowstein et al., 2003), the Draw-a-Person Test (Jenni, 2013) and play behavior (Bonhoeffer and Jenni, 2018)]. In fact, Zurich Play Behavior is an informal test battery which is used for developmental testing in clinical practice (Bonhoeffer and Jenni, 2018).

In tradition of the ZLS, a number of other comprehensive studies on child development were initiated at the University Children's Hospital Zurich over the past decades. These include the Research and Child Health Outcome (ReachOut) study, a prospective cohort study on children with congenital heart disease who underwent cardiopulmonary bypass surgery (Naef et al., 2019), and a prospective cohort study on children with congenital hypothyroidism (Dimitropoulos et al., 2009).

The Future of the ZLS: ZLS-Adulthood and ZLS-Lifespan

The dataset of ZLS-Childhood is very rich but to date, only a fraction of the available data has been analyzed and published. Although analyzing the archived ZLS-Childhood data will provide further insights into child and adolescent development, the main goal of the next years is to expand the studies into older ages and examine how factors in early life contribute to aging-related processes. Thus, the ZLS-Childhood data is currently complemented with data on health and development assessed in adulthood, namely in mid-adulthood and at the transition to old age. These ZLS-Adulthood assessments, eventually, set the basis to initiate a lifespan study, ZLS-Lifespan.

METHODS AND ANALYSIS

In the following sections, the recruitment process, study procedure, and assessment protocols are detailed separately for assessments between birth and young adulthood (i.e., ZLS-Childhood) and the first assessment wave in mid- and later adulthood (i.e., ZLS-Adulthood). A final paragraph illustrates how the data will be linked and expanded to form a lifespan data set: ZLS-Lifespan. **Figure 2** illustrates the set-up of the three ZLS cohorts.

Recruitment and Study Procedure

ZLS-Childhood

Between 1954 and 1961, a total of 445 infants were enrolled in ZLS-1. They were eligible for the study if they were born into a Swiss family living in the area of Zurich, Switzerland, and were born at the former Women's Clinic of the Cantonal Hospital of Zurich (today the University Hospital Zurich). Eligible parents were approached at random immediately after the birth of their child and informed about the study (Falkner, 1960). If they agreed to participate, they were enrolled, and demographic information was recorded. Because some parents inquired about the enrollment of later-born children into the study, a number of younger siblings were also enrolled at birth.

Between 1974 and 1979, participants for the ZLS-2 were again recruited from the Women's Clinic of the Cantonal Hospital of Zurich and other women's clinics in the Zurich area. A total of 265 infants, 162 born preterm (i.e., before 37 weeks of gestational age) and 103 born at term, were enrolled immediately after birth. Families were eligible if the parents were fluent in German. Parents were approached at random. If parents agreed to participate, the infants were enrolled, and data was collected at the time of birth from preterm and term-born infants and again at term-equivalent age from most preterm infants. Similarly to ZLS-1, some younger siblings were also enrolled at birth.

ZLS-3 enrolled the children of the ZLS-1 participants. After the ZLS-1 participants had reached young adulthood and, thus, had completed their own study participation, they were sent yearly letters inquiring whether they had become parents. If they indicated that they had, they were asked to enroll them in ZLS-3. This resulted in a total of 327 infants of 174 ZLS-1 participants enrolled in ZLS-3 between 1970 and 2002.

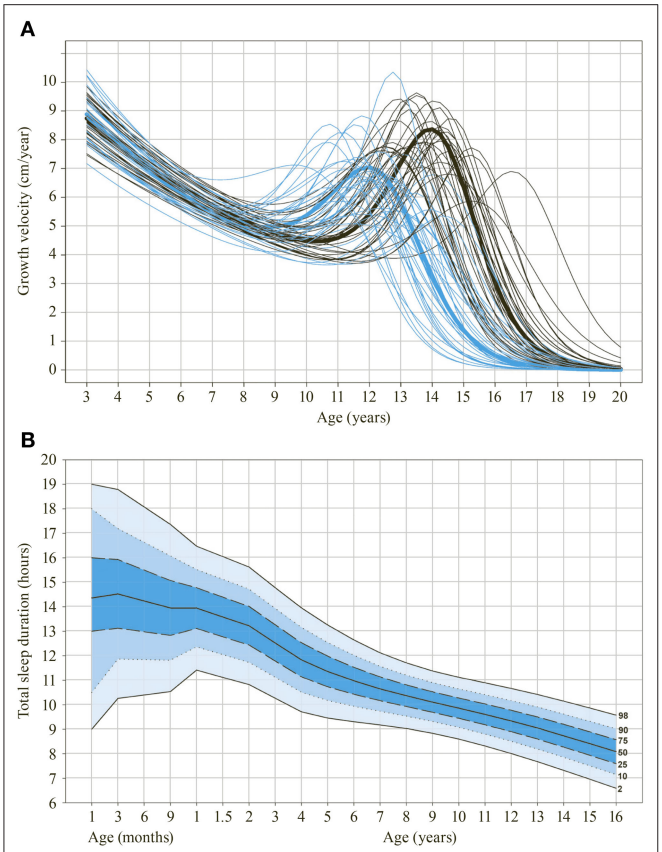


FIGURE 1 | Selected findings derived from ZLS-Childhood data. **(A)** Growth velocity (in cm/year) of individual participants of the first ZLS cohort between 3 and 18 years of age using spline functions and kernel estimation. Blue lines indicate girls, black lines indicate boys. Thick lines indicate group means. **(B)** Reference chart on sleep duration between 3 months and 16 years of age. Data is derived from the second and third ZLS-cohorts (see also Iglowstein et al., 2003).

In all three cohorts, regular assessments were then scheduled at 1, 3, 6, 9, 12, 18, and 24 months of age and annually thereafter until the age of 9 years. For the remaining duration of the studies, the timing of the visits differed between the cohorts: For ZLS-1, the children were assessed every 6 months (girls from the age of 9.5 years, boys from the age of 10.5 years) until the annual increment in height was <0.5 cm per year. Thereafter, the children were seen annually and finally discharged when the increment in height had become <0.5 cm in 2 years (Prader et al., 1989). For ZLS-2, children were assessed at 10, 14, and 18 years of age onsite and were sent a number of questionnaires at the ages of 12 and 16 years. For ZLS-3, assessments continued annually (except at age 17 years) until children turned 18 years old. For all three cohorts, assessments were scheduled at intervals of ±2 days (1 month assessment), ±1 week (3–18 months assessments), and ±2 weeks (for all remaining assessments) from the birthday of the child (Falkner, 1960). For preterm infants, the assessments were planned at the child's age corrected for prematurity across the entire study period.

Figure 3 provides an overview of the number of individuals who were enrolled in the three ZLS cohorts and how the participation rate developed across childhood and adolescence (note: in the processes of digitizing the ZLS-Childhood archives, minor adaptations of these numbers may be necessary if further information on study participants becomes available). After the initial enrollment, the participation rate dropped markedly because some families failed to participate in any of the assessments (for these infants, only information of enrollment is available; ZLS-1: $n = 13$; ZLS-2: $n = 0$; ZLS-3: $n = 31$). Additionally, in the ZLS-2 cohort, the initial drop is partly explained by the fact that 30 preterm and 1 term infants died in the neonatal period. In the ZLS-3 cohort, a number of children were only enrolled at 3 months of age with data on birth being assessed retrospectively. In the ZLS-1 cohort, 2 participants and in the ZLS-3 cohort, 1 participant died during childhood. At the 18-year final assessment common to all three cohorts, of the initially enrolled participants, 66.1% participants of ZLS-1 ($n = 294$), 73.6% participants of ZLS-2 ($n = 195$), and 73.4% participants of ZLS-3 ($n = 240$) attended.

Study participants were assessed at the former Growth and Development Center (today the Child Development Center) at the University Children's Hospital Zurich. Visits lasted approximately half a day. The children's physical, motor, cognitive, and social health and development was assessed by a pediatrician. **Figure 4A** illustrates some of the tools that were used. The accompanying parent, mostly the mother, was interviewed by a study coordinator and reported on a variety of aspects of the child and the family (see "ZLS-Childhood" for details of the assessment protocol). Parents were not compensated monetarily, but children received a small gift or a small amount of money when they were older, and travel expenses were reimbursed. Feedback on the developmental and health status of the child was provided to the parents, and in most cases the family physician or pediatrician received a short report.

Data was collected on standardized paper record forms and as drawings, analog hand X-rays, and photographic negatives. A large amount of data was subsequently saved as punched card data or on magnetic bands and in some cases, manually typed into Excel files over time. To date, no overarching digital study database has been established. Hard copies of the data and documentation of the study procedure are stored at the University Children's Hospital Zurich, Switzerland. **Figure 5** illustrates one of the archives hosting the raw data. Since August 2020, the study data of all ZLS-Childhood waves is being scanned to make digital copies available for data extraction and subsequent inclusion into a study database. Digitization of the ZLS-Childhood data, likely, continues for several years to come. An inventory of the study documentation material is underway.

The ZLS Over the Years

After the completion of assessments for ZLS-Childhood in young adulthood, contact with former participants was maintained for some years. Initially, yearly letters were sent to inquire about current contact details, and the study-specific address database was updated accordingly. Participants of the ZLS-1 were further asked whether they had become parents during the past year,

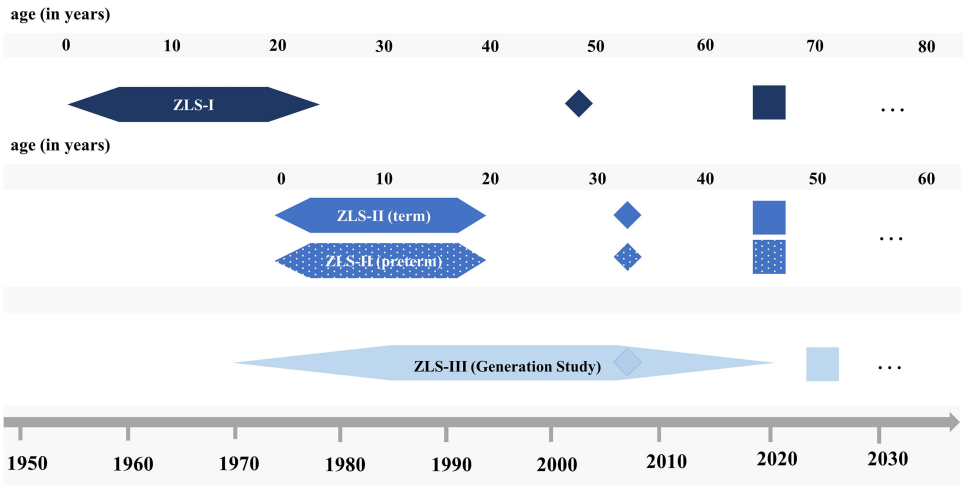


FIGURE 2 | Schematic illustration of the set-up of the three ZLS cohorts. Hexagons indicate ZLS-Childhood assessments, diamonds indicate sub-studies on motor abilities and rectangles indicate the first assessment wave of ZLS-Adulthood. Dots refer to potential future assessments (please refer to text for details).

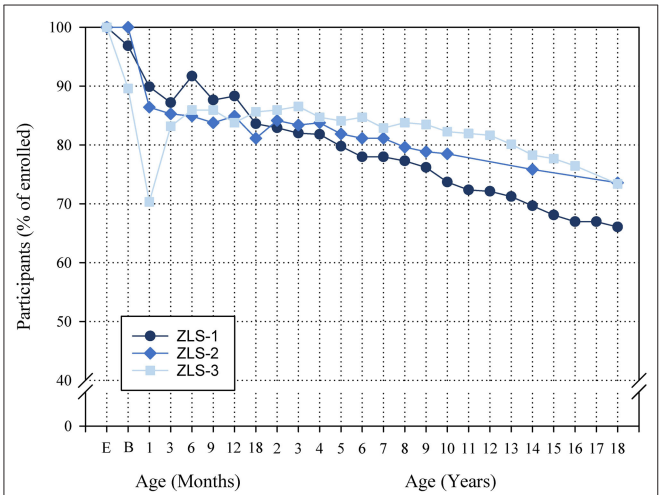


FIGURE 3 | Overview of the number of individuals enrolled in the three ZLS cohorts and the participation rate over the course of the studies. Semi-annual assessments and assessments after 18 years of age in the ZLS-1 cohort are not shown. In the process of digitizing the ZLS-Childhood archives, minor adaptations of these numbers may be necessary if further information on study participants becomes available. E, Enrollment; B, Birth.

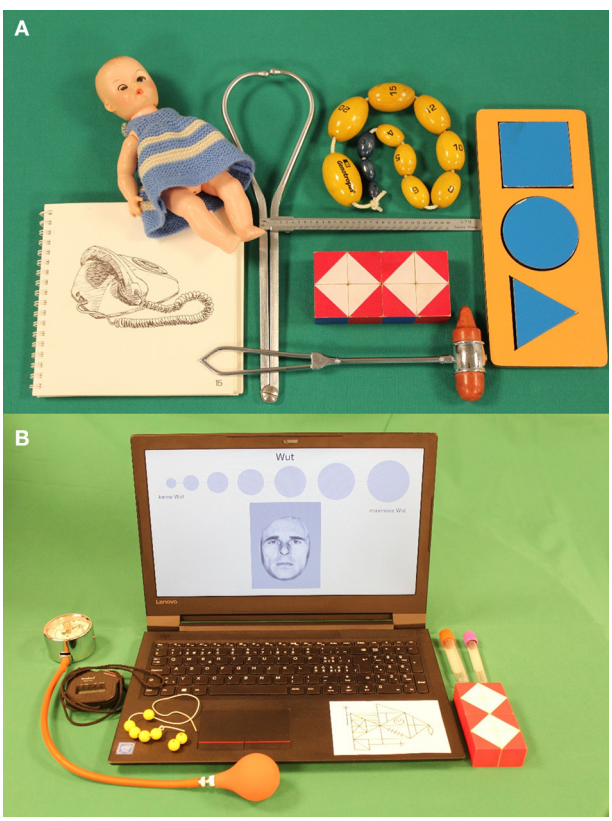


FIGURE 4 | Illustration of example assessment tools used for the ZLS assessments. **(A)** ZLS-Childhood. **(B)** ZLS-Adulthood (first assessment wave).

and if so, whether they were willing to enroll their children in ZLS-3. Those who did remained closely engaged with the ZLS as their children grew up and were assessed as study participants. At selected time-points, for some of them, information on their own health was assessed as part of the assessment of their children (e.g., IQ assessment of one parent when the participating child was 14 years old). Over the years, a number of events were organized to inform former participants about the study results. Between 2002 and 2004, 98 ZLS-1 participants took part in a substudy investigating the motor abilities of parents of ZLS-3 participants, and in 2007/2008, a subset of ZLS-2 ($n = 77$) and ZLS-3 participants ($n = 50$) took part in an independent study on motor abilities (both unpublished).

FIGURE 5 | One of the archives hosting the raw data of ZLS-Childhood.

The data acquired in ZLS-Childhood was continued to be analyzed, with the main focus shifting from physical development to the development of cognitive and motor abilities (see "Aims and Selected Findings of ZLS-Childhood" for example findings).

In 2014, substantial changes in the legal regulation of research projects with humans came into effect in Switzerland (see "Ethics" for details). As a consequence, all former ZLS participants were contacted by letter and informed about their options to consent to or decline the continued use of their childhood data for future research projects. To achieve this, the study-specific database containing the last-known contact details required comprehensive updating. This was done in two ways. In Switzerland, individuals are registered with their municipality of residence and are required to inform the respective residents' registration office of their new address if they move to a different town. This information is accessible to public institutions, including universities (for details on the legal aspects concerning this search strategy, see Lannen et al., 2021). The current contact details of the majority of individuals were identified in this way, through a number of iterations in case of multiple relocations. A second source of information was other ZLS participants: Alongside the letter with information on the legal changes, participants were sent a return form to indicate whether they knew any other ZLS participants and asked to provide their current contact details. This approach proved fruitful because the cohorts include many pairs of parents and their children (ZLS-1 – ZLS-3) or siblings, particularly within ZLS-3. Information letters were sent to all participants whose current contact details were retrieved (refer to "Ethics" for details on the response options for the continued use of data for future research projects).

The tracing of former study participants is still ongoing, and in the future, the contact details of further individuals are likely to be retrieved (e.g., disclosed by other ZLS participants or through contact with the Swiss Federal Department of Foreign Affairs for participants who emigrated). To date, of the 445 participants enrolled in ZLS-1, 399 participants have been located. Of these, 213 consented and 18 declined consent to the continued use of their data. Of the 265 participants enrolled in ZLS-2, 255 have been located. Of these, 125 consented, and three declined consent. Finally, of the 327 participants enrolled in ZLS-3, 299 have been located. Of these, 173 have consented and one has declined consent (note: individuals who have died either during ZLS-Childhood or since are considered in the number of located individuals but not in the number of individuals who provided or declined consent). For participants who do not reply to the information letter within 2 months, whose current contact details cannot be retrieved despite all efforts, or who have died, the data is retained in the dataset for further analyses (see "Ethics" for details).

Consequently, childhood data of 427 ZLS-1, 262 ZLS-2, and 326 ZLS-3 participants is available for further analyses. Placing the ZLS datasets on a sound legal basis is of utmost importance for the continued use of the data and the recruitment of participants into ZLS-Adulthood, as described below.

ZLS-Adulthood

The information on health and development assessed between birth and young adulthood (i.e., ZLS-Childhood) is currently being complemented with additional information in adulthood: At the time of the first assessment wave as described in this article, ZLS-1 participants are approximately between 60 and 65 years old, and ZLS-2 participants are approximately between 40 and 45 years old. The first assessment wave for ZLS-3 has not yet been scheduled.

Individuals who fulfill the following inclusion criteria are eligible for assessments in adulthood: They have participated at least once in a ZLS-Childhood assessment (i.e., those who were only enrolled but never participated are not eligible), their current contact details can be identified, they have not declined continued use of their childhood data for research, and they are able to provide written informed consent for study participation.

Eligible prospective participants are contacted through a letter and invited to participate in the assessment of ZLS-Adulthood. Detailed information is provided about the aim of the study and the study procedure. Participants are asked to contact the study team to schedule an assessment if they agree to participate. If they do not reply, eligible participants are contacted by phone or email a few weeks after the information letter has been sent. Those who cannot be reached after several weeks or for whom no phone number or email is available are sent a brief final letter that highlights the importance of their participation in the study. They are asked to complete and return the enclosed short-form study questionnaire with the signed consent form.

If eligible individuals agree to participate, an assessment is scheduled. Similarly to the assessments during childhood, the onsite assessment of physical, cognitive, motor, and social health parameters lasts about half a day. **Figure 4B** illustrates some of the tools used for the onsite testing. Assessments are scheduled either in the morning or in the afternoon, whichever is most convenient for the participants. Prior to the onsite assessment, participants complete a questionnaire on their health and well-being, their personality, and their current living situation. They also keep a sleep–activity diary for seven consecutive days prior to the onsite assessment. Participants are asked to participate

in specific parts only if they do not want to complete the full assessment protocol (e.g., only the short-form questionnaire).

Onsite testing is conducted in a quiet room specifically designed and equipped for this study in a venue provided by a private foundation (see "Acknowledgments") by research staff extensively trained in neuropsychological and motor testing. This provides optimal conditions for standardized assessments. Participation is compensated with CHF 125.- for onsite testing and questionnaire and CHF 25.- for questionnaire only, and travel expenses are reimbursed. Participants receive a summarized written feedback on their assessment if they wish. The recruitment of the ZLS-1 and ZLS-2 cohorts was initiated in April 2019, and assessments are planned to be completed within 2 years. However, assessments were suspended for 3 months due to restriction measures issued by the Swiss authorities to halt the spread of the Coronavirus disease 2019 (COVID-19) pandemic. The assessment of the ZLS-3 cohort is planned for after the completion of ZLS-1 and ZLS-2 assessments.

The data is entered and stored in a study-specific electronic case report form implemented in Redcap (Harris et al., 2009, 2019). Video recordings and photographs are saved on a server at the University Children's Hospital Zurich. Hand X-rays are taken at a private health clinic close to the onsite testing venue and stored on a server at the University Children's Hospital Zurich. Blood samples are drawn at the onsite testing venue and analyzed and stored at the University Children's Hospital Zurich. Due to logistic reasons, these measures are not always taken at the same time-point during the assessment. The exact time of the assessment is recorded to take potential effects of timing into consideration for further analyses as appropriate. Both measures are voluntary subparts of the assessment protocol. For individuals who have deceased, the time and nature of death is recorded.

All participants are being informed about and consent to the consolidation of the databases of ZLS-Childhood and ZLS-Adulthood into one joint database: ZLS-Lifespan.

Assessment Battery

ZLS-Childhood

The cataloging and preprocessing of the archived ZLS-Childhood datasets is an ongoing process, and thus, no final overview of the data is presented here. Instead, the structure of the assessment batteries applied in the three cohorts is described, and examples of the instruments used are provided. Details on the changes and adaptations of the assessment protocols within the individual cohorts over the course of the studies and the overlap and differences between the three cohorts are detailed.

The general architecture of the study was the same for all three ZLS-Childhood cohorts: The participants were tracked regularly and at frequent intervals from birth to young adulthood to capture the rapid developmental changes across the first two decades of life. Further, all three studies assessed detailed aspects of the child's proximal and distal psychosocial environment. **Table 1** illustrates the domains assessed. The study team gladly provides details on specific parameters that were traced across childhood and adolescence as they become available in the course of the ongoing cataloging, preprocessing and digitization of the ZLS-Childhood data; readers are encouraged to contact the authors. Notably, the developmental domains and environmental factors that were assessed in ZLS-Childhood largely overlap with those suggested more recently as requiring consideration when establishing longitudinal birth cohort studies (see Golding, 2009a for details). This highlights the foresight of the ZLS researchers at the time. They understood the importance of considering the central drivers of long-term development, which remain important to the present.

The study design and assessment instruments were initially decided upon within the context of the ICC Coordinated Longitudinal Studies (see "The Zurich Longitudinal Studies – A Brief History" for details on the ICC studies), of which the first ZLS cohort was part. Subsequently, they were largely retained for ZLS-2 and ZLS-3. The instruments were primarily quantitative in nature and included standardized physical, motor, and cognitive examinations by pediatricians and structured interviews with the parents. Examinations were documented on standardized record forms (see example for anthropometric measures in **Figure 6A**; **Supplementary Material 1** for English translation). Parents reported on child health (e.g., illnesses and accidents that had occurred since the previous visit), child development (e.g., sleep and eating behavior) and environmental factors (e.g., sociodemographic variables) by answering questions with specified response options. **Figure 6B** illustrates this standardized format for the assessment of sleep behavior (**Supplementary Material 1** for English translation). At specific time-points, teachers reported on child personality, behavior, and development through questionnaires sent by mail. Projective tests [e.g., the Rosenzweig Picture Frustration Study (P-F Test) (Rosenzweig et al., 1948), the Draw-a-Person Test (Goodenough, 1926; Harris, 1963)], open-ended questions to the parents and teachers (e.g., "What are the child's strengths?," "What are the child's difficulties?") and unstructured notes by the examiners and study coordinators complemented the quantitative assessments. Children prepared drawings at various ages (**Figure 7A**), and hand X-rays to assess bone age (**Figure 7B**) and full-body photographs to document the development of the physical appearance were taken at various time-points. Although the potential importance and value of collecting biological material in longitudinal cohort studies has been noted recently (Jones, 2009), this was not done with ZLS-Childhood participants except for the unsystematic collection of teeth from a subset of ZLS-2 and ZLS-3 participants.

Within the individual studies, assessment instruments were retained from birth into young adulthood if possible and reasonable. For example, the questions in the parental interviews on the current health status of the child remained the same over time, the assessment of anthropometric measures was repeated at every visit in the same way, and sociodemographic data was assessed repeatedly with an equivalent form. Questions and tests were added, adapted, or removed as developmentally appropriate. For example, questions related to breast-feeding were complemented and finally replaced by questions on general nutritional habits (e.g., intake of vegetables) as the child got older. Similarly, standardized tests were chosen to accurately estimate cognitive and motor abilities at specific ages (see **Figure 8** for

TABLE 1 | Overview of assessed domains of health and development during ZLS-childhood.

Physical health and development
Pregnancy and birth
Neonatal development
Illnesses and accidents
General physical health status
Anthropometry
Neurology
Hearing and vision
Pubertal development
Bone age
Dental status
Orthopedic information and podogram
Bowel and bladder control
Feeding and nutrition
Sleep
Motor development
Motor milestones
Fine and gross motor abilities
Cognitive development
General intellectual abilities[a]
Visuo-motor abilities
Language development
Academic abilities
Social development
Attachment, parent-child interaction
Emotional development
Psychosocial development and integration
Environment
Socio-economic status
Living situation
Parenting behavior
Critical life-events

The study team gladly provides details on specific parameters that were traced across childhood and adolescence as they become available in the course of the ongoing cataloging, preprocessing and digitization of the ZLS-Childhood data; readers are encouraged to contact the authors. [a]*See* ***Figure 8*** *for detailed information on how intellectual abilities were assessed across childhood and adolescence in the three ZLS-cohorts.*

an illustration of the tests applied to assess cognitive abilities at different ages).

The assessment instruments overlapped considerably between the three cohorts. In particular, the structured interviews with parents to assess the current health status, various aspects of child development, and factors of the psychosocial environment remained the same across all three studies. In addition, the longitudinal assessment of physical development across childhood, which was a primary interest of the researchers who initiated the ICC studies, remained an integral part of the assessment protocol in ZLS-2 and ZLS-3. Accordingly, parameters of physical development were documented in the same way across the three cohorts (e.g., anthropometric measures, see **Figure 6A**). Retaining the same instruments across the three cohorts ensured comparability of data and allows the investigation of developmental patterns between generations (see "Aims and Selected Findings of ZLS-Childhood" for an example of the comparison of parenting practices over time).

Despite this general overlap in the assessment protocols between the three cohorts, a number of adaptations were implemented over time to integrate advances in research and technology or for other reasons.

The frequency of assessments was reduced in the second and third cohorts compared to the first: In ZLS-1, more than 30 assessments were conducted in most children. Visits were scheduled at half-year intervals during puberty to track growth patterns. Importantly, this allowed the growth spurt in adolescence to be described (Largo et al., 1978; Gasser et al., 1984). It was considered that such an intense assessment protocol was no longer essential with the ZLS-2 and ZLS-3 cohorts to answer questions about physical development. Consequently, the semi-annual assessments, and in the ZLS-2 cohort also a number of yearly assessments, were omitted from the assessment protocol.

Some of the assessments of physical development were adjusted or omitted from the study protocols of the ZLS-2 and ZLS-3 cohorts. To describe pubertal development, children's physical appearance in ZLS-1 was documented by taking naked photographs from different angles. From this data, the Tanner stages were derived and pubertal development quantified (Tanner, 1962). Ethical considerations related to children's privacy led to the suspension of taking naked photographs in ZLS-2 and ZLS-3. Instead, children were photographed in underwear and later, full-body photographs were omitted altogether.

Cognitive development was assessed with standardized tests in all three cohorts, but the specific tests at particular ages differed between the cohorts and, in some cases, even between participants of one cohort. Reasons for the replacement of specific tests included the availability of updated or more appropriate tests or the personal preferences of the principal investigator at the time. For example, in the ZLS-1 and ZLS-2 cohorts, in accordance with the protocol of the ICC studies, the Brunet-Lézine test of infant development (Brunet and Lézine, 1951) was applied in all participants for the assessments between 1 and 24 months of age. This test provides information on four domains—posture, oculo-manual coordination, language, and socialization—and it summarizes general infant development as a developmental quotient. In the ZLS-3 cohorts, the Brunet-Lézine test was used initially but was later replaced with the Bayley Scales of Infant Development (BSID; Bayley, 1969). The BSID provides separate estimates for mental and psychomotor development.

New assessment instruments were also introduced with advancements in technology. For example, a computerized tapping task to assess motor abilities was applied with the ZLS-2 and ZLS-3 cohorts.

In summary, the data that is available for the three cohorts incorporated in ZLS-Childhood provide a detailed and comprehensive description of child and adolescent development. The study protocols have been designed to allow the investigation of a multitude of research questions, and the data was carefully

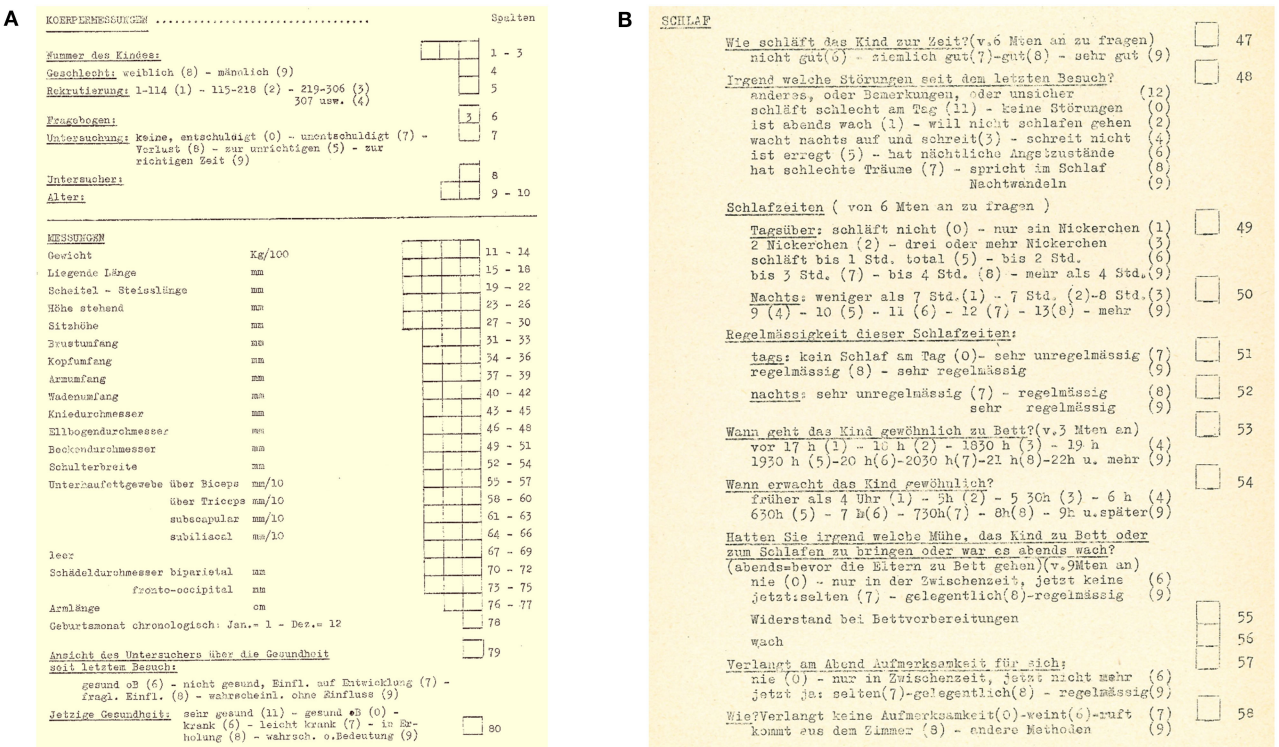


FIGURE 6 | Illustration of standardized record forms used for all three cohorts during ZLS-Childhood. **(A)** Assessment of anthropometric measures. **(B)** Assessment of sleep quantity and quality (see **Supplementary Material 1** for English translation).

recorded and archived. Nonetheless, their detailed understanding and processing requires time, money, and creativity. It has been noted that the time spent to acquire fresh data in a primary data collection study may be equivalent to the time spent understanding old, archived data (Jones, 2010). Future publications using the ZLS datasets to study child and adolescent development or to investigate how factors at the beginning of life impact adult outcome will have to detail how the childhood data was processed and made useable for such analyses.

ZLS-Adulthood

The assessment protocol of the first wave of ZLS-Adulthood is equivalent for participants of the ZLS-1 and the ZLS-2 cohorts. Potential adaptations for the assessment of ZLS-3 participants will be kept minimal to ensure comparability between the three cohorts. The assessment battery for ZLS-Adulthood was designed to include several practical and theoretical considerations. In childhood, development was assessed across a number of domains, including physical, mental, motor, and social domains. Following this precedents, the instruments for ZLS-Adulthood sought to capture indicators of health and development across a range of domains. Further, the assessment battery was intended to incorporate the understanding of health "as a state of physical, mental and social well-being and not merely the absence of disease or infirmity" as defined by the World Health Organization (WHO; World Health Organization, 1948). Thus, alongside instruments for assessing physical and mental symptoms, instruments were included to gauge physical, mental, and social well-being and quality of life. In the future, the datasets of the ZLS seek to provide information about how developmental factors contribute to aging-related processes later in life. To establish a basis for this, the WHO's definition of "healthy aging" was considered: Healthy aging describes the process of developing and maintaining functional abilities, that is, intrinsic capacities such as physical and mental abilities and characteristics of the proximal and distal living environment of individuals (World Health Organization, 2016). Consequently, a number of physical and mental abilities such as executive functions were assessed and detailed information on the environment was sought (e.g., social relations). Further definitions of "healthy aging" (see e.g., Aronson, 2020 for a review), that include aspects such as satisfaction with life and well-being were also considered. Lastly, in recent years, reports have provided evidence that aging-related processes may be more closely related to biological age than to chronological age (e.g., Belsky et al., 2015; Horvath and Raj, 2018). Aligning with this research, biological markers that allow the quantification of state-of-the-art aging parameters were added to the assessment protocol.

Tables 2 and **3** provide an overview of the assessment instruments employed in ZLS-Adulthood. They were selected because (i) they had been applied as part of previous assessments in childhood and thus allowed intraindividual longitudinal comparison of performance [e.g., Rey-Osterrieth Complex

Figure, (Rey, 1941)]; (ii) they are established clinical tools, enabling the comparison of results with clinical populations [e.g., S-words (Aschenbrenner et al., 2000)]; (iii) they have previously been applied in the Swiss Household Panel (Voorpostel et al., 2016) or the Swiss Health Survey (Bundesamt für Statistik, 2014), two large, representative studies in Switzerland, and thus, their results may be compared with the general Swiss population [e.g., socio-demographic questions, Satisfaction with Life Scale (Glaesmer et al., 2011)]; (iv) they have previously been described in the literature as sensitive to aging-related changes [e.g., dual task walking (Beurskens and Bock, 2012); grip strength (Cooper et al., 2011)]; or (v) they are being used in other long-term longitudinal cohort studies on health and development [e.g., in the Dunedin study, (Poulton et al., 2015)].

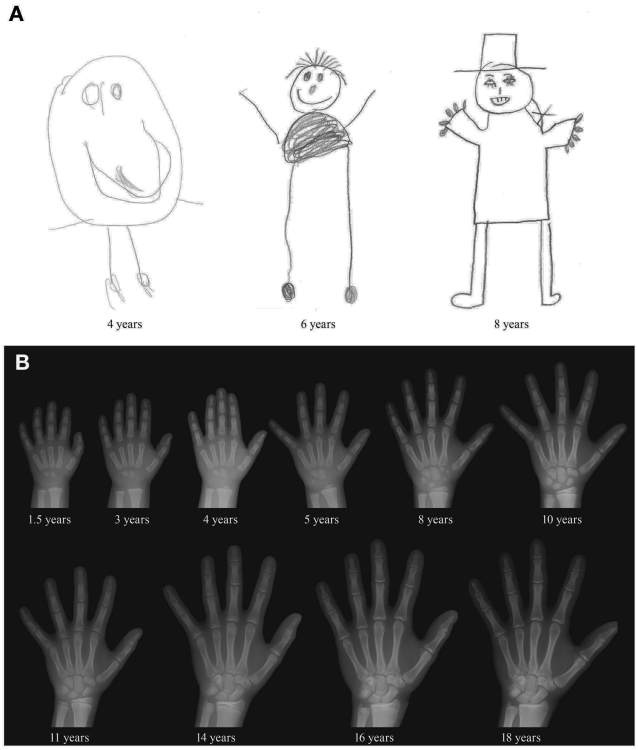


FIGURE 7 | Illustration of drawings and hand X-rays. **(A)** Draw-a-Person-Test of one male participant at 4, 6, and 8 years of age. **(B)** Serial hand X-rays of one male participant.

The onsite physical examination and neuropsychological and motor tests are applied in fixed order. This allowed the definition of a short protocol by dividing the full protocol into two parts (see **Table 2**). It also ensures similar cognitive load between the encoding and recall trials of memory tasks across all participants. The examination starts with a task to assess initial alertness [i.e., Tests of Attentional Performance, Alertness subtest; (Zimmermann and Fimm, 2009)]. Breaks are taken at fixed time-points during the assessment and at the request of participants. The assessment may be distributed across two appointments if this is more convenient for the participants. The start and end times of assessments are documented to later investigate potential effects of daytime on performance. The full protocol lasts between 4 and 6 h and the short protocol approximately 2 h, including breaks. During the COVID-19 pandemic, measures have been implemented to ensure a safe environment for participants and staff. This includes the installation of a plexiglass wall to shield participants from staff for all assessments at the table and wearing of face masks for all other tasks if the mandatory distance cannot be kept.

The study questionnaire is subdivided into several thematic parts (see subheadings in **Table 3**). For the short form, instruments or individual items are omitted from each part, but the general order is retained. Participants receive a link to the

Cohort	age at assessment in months						age at assessment in years																
	1	3	6	9	12	18	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
ZLS-1	BL	BL	BL	BL	BL	BL	BL	SB-L		SB-M	SB-M[a]		SB-M	RPM		SB-M			AH4				
ZLS-2	GR BL	GR BL	GR BL	GR BL	GR BL	GR BL	GR BL SB-L	GR SB-16	SO SB-16[a]	PET		WPPSI	Bender	WISC-R	ROCF				WISC-R ROCF				WISC-R[b] ROCF
ZLS-3	GR[c] BL[c] BSID[d]	GR[c] BL[c] BSID[d]	GR[c] BL[c] BSID[d]	GR[c] BL[c] BSID[d]	GR[c] BL[c] BSID[d]	GR[c] BL[c] BSID[d]	GR[c] BL[c] BSID[d] SB-L[c]	GR[c] BL[c] SB-60/16[e]	GR[c] SO	PET		WPPSI	Bender	WISC-R	ROCF	AID	WISC-R[a] ROCF	AID	WISC-R ROCF		AID ROCF		WISC-R[b] ROCF

FIGURE 8 | Instruments to assess cognitive abilities in ZLS-Childhood. AH4: AH4 Group Test of General Intelligence (Heim, 1970). AID: Adaptives Intelligenz-Diagnostikum ("Adaptive Intelligence Diagnostics") (Kubinger and Wurst, 1985). Bender: Bender's Visual Motor Gestalt Test (Clawson, 1962). BL, Brunet-Lézine Test (Brunet and Lézine, 1951); BSID, Bayles Scales of Infant Development (Bayley, 1969); GR, Griffiths Test (Griffiths, 1954); PET, Psycholinguistischer Entwicklungstest (German adaptation of the "Illinois Test of Psycholinguistic Abilities") (Angermaier, 1974); RPM, Raven's Progressive Matrices (Raven, 1947, 1956); ROCF, Rey-Osterrieth Complex Figure (Rey, 1941); SB-L/SB-M, Stanford-Binet Test of Intelligence L-Form/M-Form (Terman and Merrill, 1937; Lückert, 1957); SB-16, Stanford-Binet Test of Intelligence 1916-version (Terman, 1916); SB-60, Stanford-Binet Test of Intelligence 1960-version (Terman and Merrill, 1960; Terman et al., 1965); SO, Snijders-Oomen Non-Verbal Intelligence Tests (Snijders-Oomen, 1977); WISC-R, Wechsler-Intelligence Scale for Children-Revised (German version; Tewes and Titze, 1983; Willich and Friese, 1994). [a]Applied only in a small subgroup of study participants. [b]Selected subtests only (Similarities, Vocabulary, Digit Span, Picture Arrangement, Block Design, Object Assembly). [c]Tests were omitted from the assessment protocol over the course of the study (i.e., approx. after the 140th participants). [d]Task was only introduced in the course of the study (i.e., approx. after the 120th participant). [e]Until 1983: 1916-version was applied, after 1984: 1960-version was applied.

TABLE 2 | Test battery of onsite assessment for the first wave of ZLS-Adulthood (grouped by domain).

Domain	Assessment instrument	Order of assessment[a]
Physical health		
Audiometry	Interacoustics audiometer (AD629)	
Blood pressure and heart rate	Boso-Carat professional	
Anthropometric measures	Portable stadiometer (Seca 213), scale (Seca)	16
Visual acuity	Snellen Eye Test, Lang-Stereo Test	
Lung function	CareFusion MicroLab spirometer	
Grip strength	Martin Vigorimeter	
Blood sample	–	na[b]
Hand X-ray	Standardized X-ray of left hand	na[b]
Motor abilities		
Fine motor abilities	Zurich Neuromotor Assessment – 2 (Kakebeeke et al., 2018)	3
Pure motor abilities	Zurich Neuromotor Assessment – 2 (Kakebeeke et al., 2018)	11
Balance	Zurich Neuromotor Assessment – 2 (Kakebeeke et al., 2018)	18
Cognitive abilities		
Verbal comprehension	Similarities (Petermann and Petermann, 2012)	7
Perceptual organization	Block Design (Petermann and Petermann, 2012)	13
	Matrices (Petermann and Petermann, 2012)	14
Alertness	Tests of Attentional Performance (TAP), (Zimmermann and Fimm, 2009)	1
Processing speed	Coding (Petermann and Petermann, 2012)	25
	Symbol Search (Petermann and Petermann, 2012)	24
Working memory	Digit Span (Petermann and Petermann, 2012)	5, 22[c]
Fluency	Phonetic (S-Words) (Aschenbrenner et al., 2000)	21
	Semantic (Animals, Fruits) (Aschenbrenner et al., 2000)	4
	Design (Regard et al., 1982)	8
Cognitive flexibility	Trail Making Test (Strauss et al., 2006)	9
Inhibition/Interference	Stroop (Strauss et al., 2006)	23
	TAP Go/No-Go (Zimmermann and Fimm, 2009)	2
Planning	Tower Task (Unterrainer et al., 2019)	26
Verbal memory	Hopkins Verbal Learning Test (Brandt and Benedict, 2001)	12, 15[d]
Visuo-spatial memory	Rey-Osterrieth Complex Figure (Rey, 1941)	6, 10[e]
Further assessments		
Dementia screening	Montreal Cognitive Assessment (Nasreddine et al., 2005)	20
Dual-Task Walking	Subtracting 7 from 250 while walking (Abernethy, 1988); automatic assessment of gait parameters with Optogait (Lienhard et al., 2013)	19
Emotion perception	Computerized emotional rating task (Chiu et al., 2018)	27
Photograph	Nonstandardized and standardized facial photograph and full-body photograph	17
Current mood	Multidimensional Mood Questionnaire (Steyer et al., 1997)	28

The assessment protocol is equivalent for participants of the ZLS-1 and the ZLS-2 cohorts. Potential adaptations for the assessment of ZLS-3 participants (initiated after recruitment of ZLS-1 and ZLS-2 participants has been completed) will be kept minimal to ensure comparability between the three cohorts. [a]Tasks are applied in fixed order (1 = assessed first). Tasks 1 through 18 (and 28) are part of the short-form assessment protocol. [b]Due to logistic reasons, these measures were not always taken at the same time-point during the assessment. The exact time of the assessment is recorded to take potential effects of timing into consideration for further analyses as appropriate; both measures are voluntary subparts of the assessment protocol. [c]12: Digit Span forward/backward, 22: Digit Span sequential. [d]Learning and recall/recognition condition. [e]Copy and delayed recall condition.

online questionnaire implemented in RedCap (Harris et al., 2009, 2019) or are sent a paper version, whichever is most convenient for them. Paper questionnaires are entered manually into the database by the study staff. Completion time is between 90 and 120 min for the full version and approximately 45 min for the short form. In the course of the COVID-19 pandemic, the questionnaire has been expanded by a number of items to inquire about the impact of the pandemic and the measures taken by the authorities to halt its spread on the lives of the study participants (see **Supplementary Material 2**). This will help to systematically quantify a potential impact on health and well-being over the course of the data collection.

ZLS-Lifespan

Linking the rich datasets of ZLS-Childhood with the comprehensive data on health and development assessed during the first wave of ZLS-Adulthood allows the investigation of how early life impacts adult outcome. By adding additional waves of ZLS-Adulthood in the future, a ZLS-Lifespan dataset will develop and eventually provide information about health and development across the lifespan. A number of considerations will support this process. Firstly, the ZLS datasets need to be placed on a solid ethical and legal basis, including considerations related to making data available for other researchers [see e.g., the Concordat on Open Research Data (Concordat on Open Research Data, 2016)]. The effort to collect written informed consent from participants for the continued use of their

TABLE 3 | List of questionnaires completed by participants during the first wave of ZLS-Adulthood assessments.

Topic or construct	Questionnaire	Abbrev.	Included in short form
General information about yourself			
Socio-demographic information	Questions based on the SHP and SHS or study specific: Education, current working and living situation, income, civil status, children	–	yes (less detailed)
General information about personal life	Questions based on the SHP and SHS or study specific: Leisure activities, political and religious engagement	–	yes
Your health			
General perception of own health	Question of the SHS	–	yes
Information on current health	Questions based on the SHP and SHS or study specific: Drinking, smoking, drug consumption, medication intake, current illnesses, support measures (e.g., hearing aid)	–	yes
Information on previous illnesses and accidents	Based on the questionnaires applied as part of ZLS-Childhood	–	yes
Health-related quality of life	Short Form Health Survey (Morfeld et al., 2011; Wirtz et al., 2018)	SF-12	yes
Handedness	Edinburgh Handedness Inventory – Short Form (Veale, 2014)	EHI	no
Physical activity	Based on the questionnaires applied as part of ZLS-Childhood and on the SHP and SHS	–	yes
Diet	Based on the questionnaires applied as part of ZLS-Childhood and a questionnaire of the "Swiss Society for Nutrition"	–	no
Sleep quantity and quality	Pittsburgh Sleep Quality Index (Hinz et al., 2017)	PSQI	yes
Subjective memory complaints	Memory Complaint Questionnaire (Crook et al., 1992)	MAC-Q	no
Your life			
Biographical transitions	Study specific questions: Age at and appraisal of different biographical transitions and critical life events (e.g., puberty, transition to parenthood, serious illness, unemployment,…)	–	yes
Sense of Coherence	Sense of Coherence L9 (Schuhmacher et al., 2000)	SOC-L9	yes
Your well-being			
Life satisfaction	Satisfaction with Life Scale (Glaesmer et al., 2011)	SWLS	yes
Psychological Wellbeing	Ryff Inventory (Ryff and Keyes, 1995; Staudinger et al., 1999)	–	no
Domain-specific satisfaction	Study specific questions, partly based on the SHP and the PFB-K: Evaluation of satisfaction with various life domains (e.g., health, work, personal relationships…)	–	yes
Beliefs about Aging	Essentialist Beliefs About Aging Scale (Weiss et al., 2016)	–	no
Mental health	Mini-Symptom-Checklist (Franke, 2017)	Mini-SCL	no
Describing your personality			
Personality traits	NEO-Five-Factor Inventory – 30-item short form (Körner et al., 2008)	Neo-FFI-30	yes
Empathy	Saarbrücker Persönlichkeits-Fragebogen (based on the Interpersonal Reactivity Index) (Davis, 1983; Paulus, 2011)	SPF/IRI	no
Coping with difficult situations			
Resilience	Resilience Scale – 13-item form (Schuhmacher et al., 2000)	RS-13	yes
Basic psychological needs	Incongruence Questionnaire (Grosse Holtforth and Grawe, 2003)	–	no
Cognitive Emotion Regulation	Cognitive Emotional Regulation Questionnaire (Garnefski et al., 2001; Loch et al., 2011)	CERQ	no
Self-Efficacy	General Self-Efficacy Scale (Beierlein et al., 2012)	–	yes
Your relationships			
Attachment behavior	Adult Attachment Scale – Revised (Schmidt et al., 2016)	AAS-R	no
Attachment style	Adapted from Relationship Scales Questionnaire (Bartholomew and Horowitz, 1991)	RS-Q	yes
Attachment to Parents	Retrospective 1-item assessment (Neumann, 2002)	–	yes
Loneliness	6-Item Scale for Overall, Emotional, and Social Loneliness (Gierveld and Tilburg, 2006)	–	no
Social Support	Social Support Questionnaire (Fydrich et al., 2007)	F-SozU	yes

(Continued)

TABLE 3 | Continued

Topic or construct	Questionnaire	Abbrev.	Included in short form
Additional information			
Evaluation of study participation	Study specific questions	–	yes
Impact of COVID-19 pandemic[a]	Study specific questions to assess the impact of the Covid-19 pandemic on the lives of study participants (**Supplementary Material 2**)	–	yes
Sleep–activity diary[b]	Study specific diary on sleep (at night and naps during the day) and activity (e.g., sports, alcohol consumption) to be completed in the mornings and evenings for seven consecutive days.	–	no

The assessment protocol is equivalent for participants of the ZLS-1 and the ZLS-2 cohorts. Potential adaptations for the assessment of ZLS-3 participants (initiated after recruitment of ZLS-1 and ZLS-2 participants has been completed) will be kept minimal to ensure comparability between the three cohorts. Questionnaires are grouped thematically and presented after the respective sub-headings. PFB-K, Partnerschaftsfragebogen Kurzform (Relationship Questionnaire Short Form) (Kliem et al., 2012); SHP, Swiss Household Panel (Voorpostel et al., 2016); SHS, Swiss Health Survey (Bundesamt für Statistik, 2014). [a]These questions were added to the questionnaire in June 2020. [b]The diary is not part of the questionnaire but completed on paper by all participants.

childhood data (see "Ethics" for details), to have them consent to the assessment of further data in adulthood, and to link these datasets is currently ongoing. Secondly, the individual data storage units need to be replaced by one common database that allows childhood and adulthood data to be linked and is readily expandable in the future. The digitization of the ZLS-Childhood data has been initiated in August 2020 by scanning all paper forms. In the course of the next years, this data will be extracted from the digital copies and merged in a common data base with the data assessed as part of the first wave of ZLS-Adulthood (e.g., in RedCap). Efforts are undertaken to ensure consistent terminology and detailed documentation in this process. Thirdly, to assess aging-related processes and lifespan trajectories, further assessment waves need to be added to the dataset. Proper planning is required in this regard, including timely grant acquisition and keeping in touch with eligible study participants between assessment waves. Currently, yearly newsletters are sent to study participants to inform them about the progress and the future plans of the ZLS. Updated study protocols to describe the methods of future assessment waves will be published as appropriate. Similarly, updates have been published previously for long-term longitudinal birth cohort studies such as the NSHD cohort (Kuh et al., 2011) and the Lothian birth cohort study (Taylor et al., 2018).

It has been noted that it requires multiple generations of researchers to establish long-term longitudinal and ultimately lifespan studies (Mroczek et al., 2011). Accordingly, the efforts of the researchers who initiated and nurtured the ZLS over decades to improve the understanding of child health and development will now be continued by others.

Data Analysis

To embrace the multitude of data that has been assessed over the course of more than half a century will require a number of statistical methods, some of which are proposed here. Regression models will be considered to assess how factors of health and development in the first two decades of life may impact adult health outcomes. To this end, the variables measured across child and adolescent development will be considered potential predictors. Candidate variables will be specified depending on the research question. Analyses will be begun by employing classical regression models (i.e., linear regression for continuous outcomes and logistic regression for binary outcomes). When appropriately combined, the number of potential predictors available through the comprehensive ZLS-Childhood datasets may be reduced to increase their reliability and potentially also their relevance. For example, some indicators measured many times during childhood and adolescence (e.g., estimates of intellectual abilities) could be summarized by individual trajectories, typically via random intercepts and random slopes estimated from generalized linear mixed models [of which growth curve models are a particular case (Verbeke and Molenberghs, 2000)]. With this, autocorrelation can be accommodated to take into account dependencies induced by the repeated measurements in a longitudinal study (Diggle, 2002). Other examples include assessments of multidimensional domains such as motor abilities, which are measured through dozens of variables. This information may be summarized in components using the technique of simple component analysis (e.g., Rousson and Gasser, 2004). The summaries thus obtained, instead of the original variables, may subsequently be used as predictors in the regression models.

Further, neural networks may be used for data analyses because many predictors will be available even after data reduction. Neural networks are flexible models that allow the identification of complicated interactions among various predictors (Cheng and Titterington, 1994; Magoc and Magoc, 2011). Exploring the outputs of such models may provide suggestions for interactions to be included in the classical regression models.

Ultimately, which analyses are most suitable will be decided by the specific research question. Previous similar endeavors to study development from childhood into adulthood may guide these choices (e.g., McArdle et al., 2009; Friedman et al., 2010; Jones and Peskin, 2010). By employing appropriate statistical approaches, the ZLS datasets, provide a rich source of data that may help to answer questions about how health develops across the lifespan.

Ethics

Over the course of a long-term longitudinal study, ethical considerations and legal regulations are likely to change and

thus require continuous monitoring (Birmingham and Doyle, 2009). In fact, the legal basis for the ZLS has changed several times since their initiation. When they started, no approval by a formal ethical committee was required for such studies in Switzerland. Parents were informed about the aims and the procedure of the study, that participation was voluntary, and that they could withdraw at any time. They provided oral consent for participating in the study with their child. At the beginning of every visit, parents were informed about the planned assessments. All ZLS-Childhood assessments were conducted according to this procedure.

In 2008, the Ethical Committee of the University Children's Hospital Zurich classified the ZLS as unproblematic (*Unbedenklichkeitsbescheinigung*; Certificate of Clearance). This allowed the analyses of the data assessed as part of ZLS-Childhood to continue. In 2014, the new Federal Act on Research involving Human Beings, Human Research Act (HRA; The Federal Assembly of the Swiss Confederation, 2014) came in force in Switzerland. This required the ZLS to be evaluated by the Ethical Committee of the Canton of Zurich, Switzerland. The approval for the procedure with the ZLS-Childhood data (detailed below) and the assessment of the first wave of ZLS-Adulthood was granted in 2018 (Basec-Nr. 2018-00686). Amendments for future waves of assessments will be submitted as required.

As a consequence of the new HRA coming into force in 2014, the participants of ZLS-Childhood are contacted by letter and informed about what the legal basis of the ZLS had been when they were initiated and how the new HRA impacts the continued use of the ZLS-Childhood datasets for future research projects. Participants are explained the two types of ZLS-Childhood data: The vast majority of the data is considered "coded," i.e., the data is linked to the specific person via a code (HRA Art. 3; The Federal Assembly of the Swiss Confederation, 2014). This concerns for example the data on physical growth (see **Figure 6A**) or on cognitive development. In contrast, some data—most prominently the photographs—contain identifying information and are, thus, considered "uncoded." Participants are asked to return the signed consent form and indicate whether they provide (1) full consent for the continued use of their data, i.e., the use of coded and uncoded data for projects of the University Children's Hospital Zurich and the sharing of coded data with external institutions (without sharing of the respective code) or (2) partial consent, i.e., limiting the use of the data to either research projects of the University Children's Hospital Zurich (coded and uncoded data) or to external institutions (coded data). Participants are further asked to contact the principal investigator (OGJ) if they decline to consent to the continued use of their data and are sent a confirmation letter if they do. Their data is then removed from the ZLS archives, and they are not contacted any further. Participants are further informed that the Ethical Committee of the Canton of Zurich has issued a consent surrogate in case of no reply within 2 months after participants having received the information letter. This allows the continued use of the data for future research projects. The surrogate consent also covers the data of individuals who have died during the study or since or of those whose current contact details cannot be retrieved despite all efforts (HRA Art 34; The Federal Assembly of the Swiss Confederation, 2014). All future research projects with ZLS-Childhood data require individual ethical approval.

In the course of the process described above, several individuals have declined to provide consent for the continued use of their data (see "The ZLS Over the Years" for the detailed numbers). Although some have done so without further comment, others have voiced serious criticism of the way the ZLS-Childhood data was assessed and, more generally, how children were treated as study participants in the 1950s and 1960s. In particular, the naked photographs and the detailed physical examinations to assess pubertal development were described as traumatic by some individuals. The current study team takes this criticism very seriously: Former study participants are offered a personal dialogue with the principal investigator (OGJ) and their feedback is forwarded to the Ethical Committee of the Canton of Zurich, Switzerland to ensure independent documentation. A detailed discourse on the role of the children in the ZLS is needed to transparently describe the setting in which the ZLS-Childhood data, particularly of the ZLS-1 cohort, was collected and to acknowledge the distress the study has caused some of the participants. Thus, a historical and ethical reappraisal is intended of research and parenting practices and the treatment of and attitude toward children in the 1950s and 1960s.

Individuals who participate in the first wave of ZLS-Adulthood provide written informed consent for participating in the assessment, the continued use of the respective data and the linkage of the ZLS-Adulthood data with ZLS-Childhood data.

DISCUSSION

It has been argued that "longitudinal data are perhaps the most valuable type of data in adult development and aging research" (Mroczek et al., 2011, p. 128). The current article illustrates how the ZLS, previously described as among the most complete studies on child development worldwide (Tanner, 1998), are currently expanding into adulthood. The resulting long-term longitudinal dataset allows the investigation of how early life impacts adult health outcome and, ultimately, aging-related processes.

Long-term longitudinal studies provide a number of opportunities and challenges, some of which are discussed here in the context of the ZLS. This article will conclude by elaborating on some potential research questions that may be approached within the ZLS dataset.

Opportunities of Longitudinal Studies to Investigate the Impact of Childhood on Aging

Intraindividual trajectories of health and development over very long periods of time can only be studied with long-term longitudinal studies. Noteworthy examples in this regard are the Terman study, which followed individuals from mid-childhood until death (Friedman and Martin, 2011), and the NSHD, a birth cohort study whose members are currently

70 years old (Kuh et al., 2016). With the first wave of ZLS-Adulthood, comprehensive data will become available on health and development from birth into mid-adulthood (ZLS-2 and ZLS-3 cohorts) and to the transition to old age (ZLS-1 cohort). The timespan assessed will gradually increase as ZLS-participants age and additional waves are added. In addition to spanning several decades, the study design of the ZLS has other unique properties: The three cohorts include individuals born more than two decades apart, individuals with and without risk for neurodevelopmental impairments (i.e., born preterm or at term), and dyads of parents and children. This makes the ZLS highly valuable for assessing early-life antecedents of adult health and aging.

In prospective studies, data is assessed at or close to the events it relates to. For example, in the ZLS, health indicators were assessed at every visit. Parents reported on illnesses and accidents only for the period since the last visit (i.e., mostly for the past year)—a period that may be remembered well. Such prospectively collected information has been suggested to be more accurate than information that is recalled later in life (Golding, 2010). Arguably, this is particularly true when investigating the impact of early life on adult health outcomes: the period between when an event happened (i.e., childhood) and when it is remembered (i.e., old age) may span several decades.

It is important to keep in mind that long-term longitudinal studies are "semiarchival": Data assessed at earlier time-points becomes archival even if new waves of data are continuously collected (Mroczek et al., 2011). This methodological peculiarity comes with a number of challenges and limitations that are addressed next.

Challenges and Limitations of *Archived* Part of Longitudinal Datasets

Understanding the design and procedure of a study that was planned and conducted by other researchers and whose data was not originally intended to be used for what it is now is a major challenge, and accordingly, it requires time to immerse oneself into the study's dataset (Jones, 2010; Pienta and Lyle, 2018). Documentation may be spotty (Jones, 2010), and the information needs to be reviewed and inventoried carefully (Pienta and Lyle, 2018). The documentation of the three ZLS cohorts is scattered across scientific publications (e.g., Falkner, 1960), countless notebooks, and hand-written correspondence among members of the study team and external collaborators. Studying these documents has already and will continue to support the process of understanding the design of the ZLS. Furthermore, former members of the ZLS study team have previously consulted or still continue to support the current team [e.g., the former director of the Growth and Development Center of the University Children's Hospital Zurich (RHL), the former study coordinator (EK), and the former study statistician (LM)]. This likely clarifies some of the gaps in the written documentation. Nevertheless, some thoughts and decisions made by the previous study teams may not be reproducible, and the corresponding data needs to be treated accordingly in future analyses: for example, labeled as missing or to be interpreted with care.

When expanding studies on child and adolescent development into adulthood, it is important to keep in mind that the study design and the assessment instruments were decided upon with other research questions in mind than those of interest now. Consequently, information that would be desirable to have may not be available from the dataset. For example, the ZLS participants were enrolled at birth. Consequently, only a limited amount of data is available on the prenatal environment. However, a number of recent studies have reported that prenatal factors may considerably impact health in later life and aging-related processes (e.g., Franke et al., 2018). The detailed information on the medical history of pregnancy provided by the parents and the early life anthropometric variables which are available in the ZLS datasets provide some information on prenatal development. Further, the comparison of preterm and term-born individuals (ZLS-2 cohort) can inform about how the respective environment during the last part of pregnancy impacts later health outcome. However, for a detailed analysis of prenatal factors, other datasets may be better suited and should be used instead. It has been noted that prenatal enrollment in prospective cohort studies is difficult (Golding, 2009b). Thus, studies that have succeeded in doing so (e.g., the Generation R Study; Jaddoe et al., 2006) should be carefully curated, so participants can be followed into adulthood.

Alongside the enrollment criteria, which may not have been ideal to study the impact of early life on later life and aging, the instruments, likely, were not chosen with a long-term longitudinal study in mind either. For example, processing speed has been suggested as an important mediator of changes in cognition throughout the lifespan (e.g., Salthouse, 1996; Nettelbeck and Burns, 2010; Gilsoul et al., 2019). However, the ZLS-Childhood dataset contains very limited assessments of processing speed. Consequently, this cannot be studied with the ZLS datasets.

It is important to recognize and acknowledge the limitations and shortcomings that are inherent in the archived datasets. Making them transparent can help to manage the expectations about what can be investigated (e.g., Wadsworth et al., 2006).

Challenges and Limitations of *Expanding* Datasets Into Long-Term Longitudinal Data

Researchers who aim to expand a study on child development into a long-term longitudinal and eventually even lifespan study are confronted with both methodological and conceptual challenges and limitations. Methodologically, first and foremost, the study sample is predefined. The number and the type of participants (e.g., representative population sample vs. specific clinical sample) are determined by the originally enrolled sample.

Two strategies have been described when deciding upon the number of participants enrolled in a longitudinal cohort study (Golding, 2010): Including a very large number of individuals but not a large amount of data, the "huge-and-thin" strategy, or including fewer individuals of whom in-depth data is assessed, the "small-and thick" strategy. A good example of the latter is the Dunedin study, an ongoing cohort study that

has included approximately 1,000 individuals who are assessed very comprehensively and at frequent intervals (Poulton et al., 2015). Similarly, the ZLS enrolled a few hundred children in each of the three cohorts, jointly accounting for approximately 1,000 study participants. Although a high follow-up rate is always of importance in long-term longitudinal studies, this is particularly true for relatively small studies such as the ZLS to ensure a valid sample size. The first wave of ZLS-Adulthood assessments began with an event to which former participants were invited. Members of the study team informed the audience about previous results and announced the upcoming assessment wave. This social event resonated well with the study participants, and similar events are planned for the future alongside yearly newsletters. As ZLS-Childhood is expanded by ZLS-Adulthood and, ultimately, will develop into ZLS-Lifespan, it is important to build on previous experiences to maintain participation into older age and ensure optimal follow-up. The NHDS team has taught a number of valuable lessons in this regard (Kuh et al., 2016).

Besides the number of potentially eligible participants, representativeness of the study sample is important for the generalizability the findings. First, the initial sample should be representative of the population that it represents. In the ZLS-1 and ZLS-2 cohort, the research teams recruited children born into Swiss or German speaking families residing in the area of Zurich. It is reported that parents were approached at random (Falkner, 1960). Unpublished descriptive analyses of the ZLS-1 study team suggest that the cohort was representative of the population of Zurich with regard to paternal occupation; however, this remains to be evaluated in more detail. Publicly available census data (Federal Statistical Office, 2017) will be used to clarify representativeness for the ZLS-1 and ZLS-2 cohorts. The ZLS-3 cohort was recruited from the children of the ZLS-1 participants and representativeness needs to be assessed with regard to the respective population of ZLS-1 participants. Another concern in long-term longitudinal studies is whether study participants remain representative of the original sample: Selective study drop-out is a well-known issue (Young et al., 2006; Launes et al., 2014) and was discussed already in the context of the ICC studies in the 1960s (Falkner, 1960). Sociodemographic, health, and other participant characteristics determine who remains in the study over time. In the ZLS, participants to be assessed in adulthood are likely not a representative subset of those individuals enrolled at birth. This may limit generalizability of the results, and potential bias needs to be considered when interpreting future results. The data from assessments prior to drop-out help to quantify a potential shift in representativeness, for instance toward a study sample with higher socioeconomic status. Further, publicly available representative population datasets help in this regard: It has been suggested that data from household panels may serve as useful reference point for the general population (Siedler et al., 2011). Accordingly, the ZLS-Adulthood study questionnaire incorporates a number of questions from the Swiss Household Panel (Voorpostel et al., 2016) and the Swiss Health Survey (Bundesamt für Statistik, 2014), two large datasets assessed in a representative population sample in Switzerland. With these, characteristics of ZLS-Adulthood participants can be compared to characteristics of the general population. This will help to describe the ZLS cohorts in the future.

It has been noted that one of the major challenges in longitudinal studies is to balance continuity of measurements with keeping abreast of latest standards, and researchers are likely to implement what is current in their period (Hofer and Piccinin, 2009). In an attempt at appropriate balance, the study protocol of ZLS-Adulthood was designed to retain the general architecture of ZLS-Childhood: to assess parameters of physical, motor, cognitive, and social health and development and to closely describe the proximal and distal environment of participants. Some of the methods that were employed for ZLS-Childhood assessment were maintained for ZLS-Adulthood [e.g., standardized measures to assess anthropometric information, hand X-rays to quantify bone health, and the Rey-Osterrieth Complex Figure to assess visuospatial constructional abilities and memory (Rey, 1941)]. Some methods applied in ZLS-Childhood have been substituted by more modern ones, but the underlying construct remains identical. For example, duration and quality of sleep was previously assessed with a number of individual items (illustrated in **Figure 6B**). These were replaced in ZLS-Adulthood by the well-validated and clinically established Pittsburgh Sleep Quality Index (Hinz et al., 2017). Finally, a number of new methods have been added to the study protocol. For example, blood samples were collected for the first time. This allows the investigation of parameters of physical health (e.g., glycated hemoglobin, HbA1c) and, possibly even more interestingly, the quantification of biological age. Recently, a number of different approaches to do so have been published, and biological age is increasingly recognized as a reliable indicator of aging-related processes (Belsky et al., 2015; Horvath and Raj, 2018). This illustrates how advances in research may have an effect on the protocol of a long-term longitudinal study.

Importantly, not only are the ways of assessing particular constructs likely to change over the course of a long-term longitudinal study but also the constructs themselves. In the ZLS, some constructs were only formally introduced into the study protocol recently as part of the assessment of ZLS-Adulthood. Very likely, however, data touching on similar or overlapping constructs has already been assessed across childhood and adolescence without those labels. For example, the concept of executive functions was first defined only in the 1970s, but a "control mechanism" was discussed in the mid-19th century (Goldstein et al., 2014). Accordingly, the interviews with the parents of the first ZLS cohort (initiated in the 1950s) included a number of questions related to self-control and self-regulation, for instance on the frustration tolerance of the child, on strategies the child employs to self-sooth, and on the child's reactions to altered routines. Interestingly, similar questions are currently included in the preschool version of the Behavior Rating Inventory of Executive Functions (BRIEF-P; Daseking and Petermann, 2013). However, unlike in such psychometrically validated rating scales, the relevant questions in the ZLS datasets were neither labeled as relevant to executive functions nor arranged as specific scales; rather, they are scattered across many different records. If they are identified and combined

appropriately (see below for examples of how other researchers have previously approached this issue), they may reliably assess the underlying construct and serve as potential predictor of adult health outcome. In fact, child self-control has been shown to predict adult physical health, substance dependence, personal finances, and criminal offending outcomes in the Dunedin study (Moffitt et al., 2011).

Innovative methods are needed to extract psychometrically sound constructs from existing archival data of long-term longitudinal studies. A number of researchers have presented ways of accessing old data with new concepts. For example, Jack Block developed the California Q-Sort (CQS) method to derive a valid measure of personality from the diverse interview and observational data assessed in the IGS (Block, 1961). His comprehensive book provides details of the methods (Block, 2008). To date, his measure is used to study personality development across life (e.g., Jones and Meredith, 1996; Chopik and Grimm, 2019; Chopik et al., 2019) and how personality traits early in life impact health in adulthood (e.g., Peskin and Jones, 2015). Recently, one study has even derived a new construct from the CQS personality measure: a subset of items thought relevant to empathy was selected from the original measure through correlational analyses and based on theoretical considerations. Then, the newly developed measure was validated in a contemporary undergraduate student sample against an established rating scale of empathy, the Interpersonal Reactivity Index (Davis, 1983). With this, the authors were able to investigate longitudinal changes in empathy across the lifespan (Oh et al., 2020). In the comprehensive ZLS datasets, a number of such unlabeled constructs are likely buried, and they need to be identified to become available for future analyses.

Even if the underlying construct is defined, the tools to assess them were unlikely to have been constant over time (see **Figure 8** for how cognitive abilities were assessed at different ages in ZLS-Childhood). Researchers have advanced statistical methods to cope adequately with this issue. For example, in the IGS datasets, item response theory and latent curve modeling were combined in a longitudinal growth model to bring different measures of vocabulary and short-term memory into a common scale (McArdle et al., 2009). Thus, these constructs became accessible despite them being assessed with various age-appropriate tools over time. Similar approaches may be employed to investigate developmental trajectories in the ZLS datasets.

In summary, the ZLS will not constitute a final, static dataset for a while. Instead, the datasets will continue to develop as researchers engage in the tasks of working with the archived raw data to develop appropriate measures for constructs of interest and link them to adult outcome. A number of potential research questions that can be answered with the ZLS datasets in the future is outlined below.

Potential Questions to Be Addressed Within the Scope of the ZLS

Presumably, an infinite number of research questions could be addressed with the ZLS datasets. Some of those can be directly translated into a hypothesis, others may be approached in a data-driven and possibly, even exploratory, manner. The benefits of the complementary use of these strategies has been discussed previously (e.g., Kell and Oliver, 2004; Elliott et al., 2016; Matsumuro and Miwa, 2019).

Potential questions include these: How stable are motor and cognitive abilities between early childhood and late adulthood? Does a constitutional delay or an acceleration of puberty have a long-term impact on health and development in adulthood? Is the association between motor and cognitive abilities in adulthood stronger than their relatively weak association in childhood? Do physical activity and motor abilities in early childhood impact health in later adulthood? What are the childhood factors that contribute to cognitive decline in aging? How do personality traits in childhood impact adult health and development? Are early developmental milestones in social behavior predictive of adult outcome? How are emotional and social development linked to resilience over the lifespan? Are individuals with advanced biological age in childhood also biologically older in adulthood? Do better cognitive abilities in childhood protect against biological aging in adulthood? What are the long-term effects of preterm birth on health, wealth, and well-being? Are preterm-born adults biologically older than term-born adults? Is aging accelerated after preterm birth? Is the decrease in motor skill variability between individuals in childhood followed by an increase in older age? How do different parenting practices in the 1950s and 1970s impact individuals' long-term development? Is bone health in childhood predictive of bone health in adulthood? Does bone health change over generations, and if yes, why? Is the well-described Flynn effect paralleled by a secular trend in growth parameters?

CONCLUSION AND OUTLOOK

This article highlights the value of long-term longitudinal studies for aging research using the example of the ZLS. The article aims to foster the interdisciplinary discussion on how studies that were established to investigate child health and development may add to the understanding of adult health outcomes and aging-related processes in cognition, behavior, and other domains. The rich database of the ZLS includes data on physical, motor, cognitive, and social health and development and on the proximal and distal environment of individuals born with or without a risk for neurodevelopmental impairments and on parents and their children: a combination that makes these studies unique in the quest to study the importance of early life for later health and development.

In the future, the impact of the ZLS datasets may even be multiplied by collaborating with other researchers in combining datasets (see e.g., Friedman et al., 2014 for a call to integrate existing longitudinal studies to study the link between personality traits and health across the lifespan) and by the implementation of modern scientific technologies in future assessment waves (e.g., neuroimaging and genetic methods). Currently, the data of ZLS-1 participants is being aligned and compared with participants of the LifeStories study (see Lannen et al., 2021

for details). This will provide insights into how the early environment of an infant either placed in an infant care institution or growing up in a family setting affects health and development across the lifespan. Future comparison with further cohorts may help to disentangle the factors in early life that shape development into adulthood and across the lifespan.

A number of methodological and conceptual challenges remain to be addressed and solved to expand ZLS-Childhood with ZLS-Adulthood and to establish ZLS-Lifespan. However, this endeavor will make the ZLS invaluable for studying the importance of early life for health and development as individuals age.

DISSEMINATION

The findings derived from the ZLS datasets will be disseminated within the research community and beyond: Exchange and discourse with experts in the field of aging research and lifespan health development will be sought through contributions in scientific journals, choosing open access options whenever possible, and at scientific meetings. The study participants and the general public will be informed about research findings through media coverage, newsletters and information events. It is expected that the findings derived from the ZLS datasets will have a broad impact on society, as they have had in the past [e.g., with the long-term bestseller *Babyjahre* (Baby years) authored by Prof. emer. Remo H. Largo, the former Director of the Growth and Development Center of the University Children's Hospital and the former scientific leader of the ZLS and co-author of this article].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical committee of the Canton of Zurich, Switzerland (Basec-Nr. 2018-00686). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FMW, a postdoctoral researcher and the current project leader of ZLS-Adulthood, drafted and edited the manuscript. JC, a senior scientist in the ZLS since 1993, examined numerous participants of the ZLS-3 cohort, has been responsible for archiving the ZLS-Childhood datasets and provided input to the mauscript. DAE, a PhD student in the ZLS, assesses participants for ZLS-Adulthood, prepares archived data for analyses, prepared data for several figures and provided input to the manuscript. GH, a research assistant in ZLS, locates and assesses participants for ZLS-Adulthood and provided input to the manuscript. BL, the co-director of the Child Development Center, University Children's Hospital Zurich, is responsible for the at-risk population in the ZLS and provided input to the manuscript. RHL was the former principal investigator of the ZLS (from 1974 to 2005), provided input to the manuscript and approved the initially submitted manuscript. Sadly, RHL died on November 11th 2020, shortly after manuscript submission. We dedicate this work to the memory of RHL. THK, a senior scientist in the ZLS, is responsible for the motor assessments in the ZLS, prepared several figures and photographs and provided input to the manuscript. OGJ, the current principal investigator of the ZLS (2005-current), drafted parts of the manuscript, provided input to all other sections, and edited the final version. All authors contributed to the article and approved the submitted version.

FUNDING

The ZLS have previously been or are currently funded by the International Children's Center in Paris, the Department of Education of the Canton of Zurich, Switzerland, the Swiss National Science Foundation, the University of Zurich, Switzerland, the Hermann Klaus-Stiftung, the Hartmann Müller-Stiftung, the Remo Largo Stiftung für Entwicklungspädiatrie, the Foundation for Research in Science and the Humanities at the University of Zurich, the Maiores Stiftung, the Baugarten Stiftung, the Children's Research Center of the University Children's Hospital Zurich, the Velux Stiftung and the Stiftung Für das Kind Giedion Risch.

ACKNOWLEDGMENTS

We are immensely grateful to the more than 1,000 participants of the ZLS and their families for their continuous support of the studies over the past decades. Furthermore, we thank the hundreds of scientists who have contributed to the ZLS over the past 66 years; they cannot be named individually. We thank Elisabeth Kälin, the study coordinator from 1982 to 2016, who supported the location and recruitment of the former ZLS participants. Furthermore, we are thankful to Dr. Luciano Molinari, the study statistician from 1974 to 2016, who has provided valuable information in the data management process. The Stiftung Für das Kind Giedion Risch kindly provides the venue for the onsite assessments of the first wave of ZLS-Adulthood. We thank Simon Milligan for language editing of the manuscript and Susanne Staubli for her support with the illustrations. We thank Dr. Patricia Lannen and the research team at the Marie Meierhofer Children's Institute, University of Zurich, Switzerland for exchanging ideas and collaborating in their LifeStories project (see Lannen et al., 2021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.612453/full#supplementary-material

REFERENCES

Aartsen, M. J., Cheval, B., Sieber, S., Van der Linden, B. W., Gabriel, R., Courvoisier, D. S., et al. (2019). Advantaged socioeconomic conditions in childhood are associated with higher cognitive functioning but stronger cognitive decline in older age. *Proc. Natl. Acad. Sci.* 116, 5478–5486. doi: 10.1073/pnas.1807679116

Abernethy, B. (1988). Dual-task methodology and motor skills research: some applications and methodological constraints. *J. Hum. Move. Stud.* 14, 101–132.

Angermaier, M. (1974). *Psycholinguistischer Entwicklungstest: PET; Manual; Deutsche Bearbeitung des "Illinois Test of Psycholinguistic Abilities"*. Weinheim.

Aronson, L. (2020). Healthy aging across the stages of old age. *Clin. Geriatr. Med.* 36, 549–558. doi: 10.1016/j.cger.2020.06.001

Aschenbrenner, S., Tucha, O., and Lange, K. W. (2000). *Regensburger Wortflüssigkeits-Test (RWT) [Regensburger Verbal Fluency Test (RWT)]*. Göttingen: Hogrefe Verlag.

Bartholomew, K., and Horowitz, L. M. (1991). Attachment styles among young adults: a test of a four-category model. *J. Pers. Soc. Psychol.* 61:226. doi: 10.1037/0022-3514.61.2.226

Batty, G. D., Deary, I. J., and Gottfredson, L. S. (2007). Premorbid (early life) IQ and later mortality risk: systematic review. *Ann. Epidemiol.* 17, 278–288. doi: 10.1016/j.annepidem.2006.07.010

Bayley, N. (1969). *Bayley Scales of Infant Development: Birth to Two Years*. New York, NY: Psychological Corporation.

Beierlein, C., Kovaleva, A., Kemper, C. J., and Rammstedt, B. (2012). *Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU)*. Available online at: https://psycharchives.zpid.de/handle/20.500.12034/431 (accessed December 18, 2020).

Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., et al. (2015). Quantification of biological aging in young adults. *PNAS*. 112, E4104–E4110. doi: 10.1073/pnas.1506264112

Beurskens, R., and Bock, O. (2012). Age-related deficits of dual-task walking: a review. *Neural Plast.* 2012:131608. doi: 10.1155/2012/131608

Birmingham, K., and Doyle, A. (2009). Ethics and governance of a longitudinal birth cohort. *Paediatr. Perinat. Epidemiol.* 23, 39–50. doi: 10.1111/j.1365-3016.2008.00995.x

Block, J., (2008). *The Q-Sort in Character Appraisal: Encoding Subjective Impressions of Persons Quantitatively*. Washington, DC: American Psychological Association.

Block, J. (1961). *The Q-Sort Method in Personality Assessment and Psychiatric Research*. Springfield, IL.

Bonhoeffer, J., and Jenni, O. (2018). Das frühkindliche Spielverhalten–ein Spiegel der kognitiven Entwicklung. *Pädiatrie up2date*. 13, 303–321. doi: 10.1055/s-0043-115419

Brandt, J., and Benedict, R. H. (2001). *Hopkins Verbal Learning Test–Revised: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Brandt, M., Deindl, C., and Hank, K. (2012). Tracing the origins of successful aging: the role of childhood conditions and social inequality in explaining later life health. *Soc. Sci. Med.* 74, 1418–1425. doi: 10.1016/j.socscimed.2012.01.004

Brunet, O., and Lézine, I. (1951). *Le Développement Psychologique de la Première Enfance*. Paris: Presses Universitaires de France.

Bundesamt für Statistik (2014). *Schweizerische Gesundheitsbefragung 2012 [Swiss Health Survey 2012]*. Bern: Bundesamt für Statistik.

Cermakova, P., Formanek, T., Kagstrom, A., and Winkler, P. (2018). Socioeconomic position in childhood and cognitive aging in Europe. *Neurology*. 91, e1602–e1610. doi: 10.1212/WNL.0000000000006390

Cheng, B., and Titterington, D. M. (1994). Neural networks: a review from a statistical perspective. *Stat. Sci.* 9, 2–30. doi: 10.1214/ss/1177010638

Chiu, I., Piguet, O., Diehl-Schmid, J., Riedl, L., Beck, J., Leyhe, T., et al. (2018). Facial emotion recognition performance differentiates between behavioral variant frontotemporal dementia and major depressive disorder. *J Clin. Psychiatry*. 79:16m11342. doi: 10.4088/JCP.16m11342

Chopik, W. J., Edelstein, R. S., and Grimm, K. J. (2019). Longitudinal changes in attachment orientation over a 59-year period. *J. Pers. Soc. Psychol.* 116:598. doi: 10.1037/pspp0000167

Chopik, W. J., and Grimm, K. J. (2019). Longitudinal changes and historic differences in narcissism from adolescence to older adulthood. *Psychol. Aging* 34:1109. doi: 10.1037/pag0000379

Clawson, A. (1962). *The Bender Visual Motor Gestalt Test for Children: A Manual*. Los Angeles, CA: Western Psychological Services.

Concordat on Open Research Data (2016). Available online at: https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/ (accessed December 18, 2020)

Cooper, R., Hardy, R., Sayer, A. A., Ben-Shlomo, Y., Birnie, K., Cooper, C., et al. (2011). Age and gender differences in physical capability levels from mid-life onwards: the harmonisation and meta-analysis of data from eight UK cohort studies. *PLoS ONE* 6:e27899. doi: 10.1371/journal.pone.0027899

Crook, T. H., Feher, E. P., and Larrabee, G. J. (1992). Assessment of memory complaint in age-associated memory impairment: the MAC-Q. *Int. Psychogeriatr.* 4, 165–176. doi: 10.1017/S1041610292000991

Daseking, M., and Petermann, F. (2013). *BRIEF-P Verhaltensinventar zur Beurteilung exekutiver Funktionen für das Kindergartenalter; deutschsprachige Adaptation des Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P) von Gerard A. Gioia, Kimberly Andrews Espy und Peter K. Isquith; Manual*. Bern: Huber.

Davis, D. W., Turkheimer, E., Finkel, D., Beam, C., and Ryan, L. (2019). The Louisville twin study: past, present and future. *Twin Res. Hum. Genet.* 22, 735–740. doi: 10.1017/thg.2019.37

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44:113. doi: 10.1037/0022-3514.44.1.113

Deary, I. J., Gow, A. J., Pattie, A., and Starr, J. M. (2012). Cohort profile: the Lothian birth cohorts of 1921 and 1936. *Int. J. Epidemiol.* 41, 1576–1584. doi: 10.1093/ije/dyr197

Diggle, P. (2002). *Analysis of longitudinal data*. Oxford: Oxford University Press.

Dimitropoulos, A., Molinari, L., Etter, K., Torresani, T., Lang-Muritano, M., Jenni, O. G., et al. (2009). Children with congenital hypothyroidism: long-term intellectual outcome after early high-dose treatment. *Pediatr. Res.* 65, 242–248. doi: 10.1203/PDR.0b013e31818d2030

Elliott, K. C., Cheruvelil, K. S., Montgomery, G. M., and Soranno, P. A. (2016). Conceptions of good science in our data-rich world. *Bioscience* 66, 880–889. doi: 10.1093/biosci/biw115

Falkner, F. (1960). *Child Development: An International Method of Study*. New York, NY: S. Karger.

Federal Statistical Office (2017). *Census 2019*. Available online at: https://www.bfs.admin.ch/bfs/en/home/basics/census.html

Franke, K., Gaser, C., Roseboom, T. J., Schwab, M., de Rooij, S. R. (2018). Premature brain aging in humans exposed to maternal nutrient restriction during early gestation. *Neuroimage* 173, 460–471. doi: 10.1016/j.neuroimage.2017.10.047

Franke, G. H. (2017). *Mini-SCL - Mini-Symptom-Checklist. 1. Auflage ed*. Göttingen: Hogrefe.

Frey, B. B. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. Thousand Oaks, CA: Sage Publications.

Friedman, H., and Martin, L. R. (2011). *The Longevity Project: Surprising Discoveries for Health and Long Life From the Landmark Eight Decade Study*. New York, NY: Hay House, Inc.

Friedman, H. S., Kern, M. L., Hampson, S. E., and Duckworth, A. L. (2014). A new life-span approach to conscientiousness and health: combining the pieces of the causal puzzle. *Dev. Psychol.* 50:1377. doi: 10.1037/a0030373

Friedman, H. S., Kern, M. L., and Reynolds, C. A. (2010). Personality and health, subjective well-being, and longevity. *J. Pers.* 78, 179–216. doi: 10.1111/j.1467-6494.2009.00613.x

Fydrich, T., Sommer, G., and Brähler, E. (2007). *F-SOZU: Fragebogen zur sozialen Unterstützung*. Göttingen: Hogrefe.

Gale, C. R., Booth, T., Starr, J. M., and Deary, I. J. (2016). Intelligence and socioeconomic position in childhood in relation to frailty and cumulative allostatic load in later life: the Lothian Birth Cohort 1936. *J. Epidemiol. Community Health* 70, 576–582. doi: 10.1136/jech-2015-205789

Garnefski, N., Kraaij, V., and Spinhoven, P. (2001). Negative life events, cognitive emotion regulation and emotional problems. *Pers. Individ. Diff.* 30, 1311–1327. doi: 10.1016/S0191-8869(00)00113-6

Gasser, T., Köhler, W., Müller, H., Kneip, A., Largo, R., Molinari, L., et al. (1984). Velocity and acceleration of height growth using kernel estimation. *Ann. Hum. Biol.* 11, 397–411. doi: 10.1080/03014468400007311

Gierveld, J. D. J., and Tilburg, T. V. (2006). A 6-item scale for overall, emotional, and social loneliness: confirmatory tests on survey data. *Res. Aging* 28, 582–598. doi: 10.1177/0164027506289723

Gilsoul, J., Simon, J., Hogge, M., and Collette, F. (2019). Do attentional capacities and processing speed mediate the effect of age on executive functioning? *Aging Neuropsychol. Cogn.* 26, 282–317. doi: 10.1080/13825585.2018.1432746

Glaesmer, H., Grande, G., Braehler, E., and Roth, M. (2011). The German version of the satisfaction with life scale (SWLS). *Eur. J. Psychol. Assess.* 27, 127–132. doi: 10.1027/1015-5759/a000058

Golding, J. (2009a). Measuring outcomes in a longitudinal birth cohort. *Paediatr. Perinat. Epidemiol.* 23, 185–200. doi: 10.1111/j.1365-3016.2009.01016.x

Golding, J. (2009b). Who should be studied and when in a longitudinal birth cohort? *Paediatr. Perinat. Epidemiol.* 23, 15–22. doi: 10.1111/j.1365-3016.2008.00998.x

Golding, J. (2010). Are findings from large longitudinal studies of child health and development useful or just of interest? *Paediatr. Child Health* 20, 163–166. doi: 10.1016/j.paed.2009.12.003

Goldstein, S., Naglieri, J. A., Princiotta, D., and Otero, T. M. (2014). *Introduction: A History of Executive Functioning as a Theoretical and Clinical Construct. Handbook of Executive Functioning*. Springer, 3–12.

Goodenough, F. L. (1926). *Measurement of Intelligence by Drawings.* New York, NY: Harcourt Brace.

Griffiths, R. (1954). *The Abilities of Babies: A Study in Mental Measurement*. New York, NY: McGraw-Hill.

Grimm, K. J., McArdle, J., Widaman, K. F. (2011). "Family-level variance in verbal ability change in the intergenerational studies," in *Secondary Data Analysis: An Introduction for Psychologists,* eds K. H. Trzesniewski, M. B. Donnellan, and R. E. Lucas (Washington, DC: American Psychological Association), 209–229.

Grosse Holtforth, M., and Grawe, K. (2003). Der Inkongruenzfragebogen (INK). *Zeitschrift klinische Psychol. Psychother.* 32, 315–323. doi: 10.1026/0084-5345.32.4.315

Harris, D.B. (1963). *Children's Drawings as Measures of Intellectual Maturity*. New York, NY: Harcourt, Brace & World.

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., et al. (2019). The REDCap consortium: building an international community of software platform partners. *J. Biomed. Inform.* 95:103208. doi: 10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42, 377–381. doi: 10.1016/j.jbi.2008.08.010

Heim, A. (1970). *Manual for the AH4 Group Test of General Intelligence*. Windsor: NFER.

Hinz, A., Glaesmer, H., Brähler, E., Löffler, M., Engel, C., Enzenbach, C., et al. (2017). Sleep quality in the general population: psychometric properties of the Pittsburgh sleep quality index, derived from a German community sample of 9284 people. *Sleep Med.* 30, 57–63. doi: 10.1016/j.sleep.2016.03.008

Hofer, S. M., and Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychol. Methods* 14:150. doi: 10.1037/a0015566

Horvath, S., and Raj, K. (2018). DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* 19:371. doi: 10.1038/s41576-018-0004-3

Iglowstein, I., Jenni, O. G., Molinari, L., and Largo, R. H. (2003). Sleep duration from infancy to adolescence: reference values and generational trends. *Pediatrics* 111, 302–307. doi: 10.1542/peds.111.2.302

Jaddoe, V. W., Mackenbach, J. P., Moll, H. A., Steegers, E. A., Tiemeier, H., Verhulst, F. C., et al. (2006). The generation R study: design and cohort profile. *Eur. J. Epidemiol.* 21:475. doi: 10.1007/s10654-006-9022-0

Jenni, O. (2013). Wie Kinder die Welt abbilden–und was man daraus folgern kann. *Pädiatr. up2date*. 8, 227–253. doi: 10.1055/s-0032-1326475

Jenni, O. G., Chaouch, A., Caflisch, J., and Rousson, V. (2013). Infant motor milestones: poor predictive value for outcome of healthy children. *Acta Paediatr.* 102, e181–e184. doi: 10.1111/apa.12129

Jenni, O. G., Chaouch, A., Locatelli, I., Thoeni, I., Diezi, M., Werner, H., et al. (2011). Intra-individual stability of neuromotor tasks from 6 to 18 years: a longitudinal study. *Hum. Move. Sci.* 30, 1272–1282. doi: 10.1016/j.humov.2010.12.002

Jenni, O. G., Molinari, L., Caflisch, J. A., and Largo, R. H. (2007). Sleep duration from ages 1 to 10 years: variability and stability in comparison with growth. *Pediatrics* 120, e769–e776. doi: 10.1542/peds.2006-3300

Jones, C. (2010). Archival data: advantages and disadvantages for research in psychology. *Soc. Pers. Psychol. Compass* 4, 1008–1017. doi: 10.1111/j.1751-9004.2010.00317.x

Jones, C. J., and Meredith, W. (1996). Patterns of personality change across the life span. *Psychol. Aging* 11:57. doi: 10.1037/0882-7974.11.1.57

Jones, C. J., and Peskin, H. (2010). Psychological health from the teens to the 80s: multiple developmental trajectories. *J. Adult Dev.* 17:20. doi: 10.1007/s10804-009-9075-x

Jones, R. (2009). The importance of biological samples in longitudinal birth cohort studies. *Paediatr. Perinat. Epidemiol.* 23, 93–102. doi: 10.1111/j.1365-3016.2008.00999.x

Kakebeeke, T. H., Knaier, E., Chaouch, A., Caflisch, J., Rousson, V., Largo, R. H., et al. (2018). Neuromotor development in children. Part 4: new norms from 3 to 18 years. *Dev. Med. Child Neurol.* 60, 810–819 doi: 10.1111/dmcn.13793

Kell, D. B., and Oliver, S. G. (2004). Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. *Bioessays* 26, 99–105. doi: 10.1002/bies.10385

Kern, M. L., Della Porta, S. S., and Friedman, H. S. (2014). Lifelong pathways to longevity: personality, relationships, flourishing, and health. *J. Pers.* 82, 472–484. doi: 10.1111/jopy.12062

Kliem, S., Job, A.-K., Kröger, C., Bodenmann, G., Stöbel-Richter, Y., Hahlweg, K., et al. (2012). Entwicklung und Normierung einer Kurzform des Partnerschaftsfragebogens (PFB-K) an einer repräsentativen deutschen Stichprobe. *Zeitschrift Klinische Psychol. Psychother.* 41, 81–89. doi: 10.1026/1616-3443/a000135

Körner, A., Geyer, M., Roth, M., Drapeau, M., Schmutzer, G., Albani, C., et al. (2008). Persönlichkeitsdiagnostik mit dem neo-fünf-faktoren-inventar: Die 30-item-kurzversion (neo-ffi-30). *PPmP Psychother. Psychosomatik Medizinische Psychol.* 58, 238–245. doi: 10.1055/s-2007-986199

Kubinger, K. D., and Wurst, E. (1985). *Adaptives Intelligenz-Diagnostikum: AID.* Weinheim: Beltz.

Kuh, D. (2016). From paediatrics to geriatrics: a life course perspective on the MRC national survey of health and development. *Eur. J. Epidemiol.* 31, 1069–1079. doi: 10.1007/s10654-016-0214-y

Kuh, D., Pierce, M., Adams, J., Deanfield, J., Ekelund, U., Friberg, P., et al. (2011). Cohort profile: updating the cohort profile for the MRC national survey of health and development: a new clinic-based data collection for ageing research. *Int. J. Epidemiol.* 40:e1–e9. doi: 10.1093/ije/dyq231

Kuh, D., Wong, A., Shah, I., Moore, A., Popham, M., Curran, P., et al. (2016). The MRC national survey of health and development reaches age 70: maintaining participation at older ages in a birth cohort study. *Eur. J. Epidemiol.* 31, 1135–1147. doi: 10.1007/s10654-016-0217-8

Lannen, P., Sand, H., Sticca, F., Gallego, I. R., Bombach, C., Simoni, H., et al. (2021). Development and health of adults formerly placed in infant care institutions - study protocol of the life stories project. *Front. Hum. Neurosci.* doi: 10.3389/fnhum.2020.611691

Largo, R., Gasser, T., Prader, A., Stuetzle, W., and Huber, P. (1978). Analysis of the adolescent growth spurt using smoothing spline functions. *Ann. Hum. Biol.* 5, 421–434. doi: 10.1080/03014467800003071

Largo, R., Molinari, L., Pinto, L. C., Weber, M., and Due, G. (1986). Language development of term and preterm children during the first five years of life. *Dev. Med. Child Neurol.* 28, 333–350. doi: 10.1111/j.1469-8749.1986.tb03882.x

Largo, R. H., Caflisch, J. A., Hug, F., Muggli, K., Molnar, A. A., Molinari, L., et al. (2001a). Neuromotor development from 5 to 18 years. Part 1: timed performance. *Dev. Med. Child Neurol.* 43, 436–443. doi: 10.1111/j.1469-8749.2001.tb00739.x

Largo, R. H., Caflisch, J. A., Hug, F., Muggli, K., Molnar, A. A., and Molinari, L. (2001b). Neuromotor development from 5 to 18 years. Part 2: associated movements. *Dev. Med. Child Neurol.* 43, 444–453. doi: 10.1111/j.1469-8749.2001.tb00740.x

Largo, R. H., Molinari, L., Siebenthal, K., and Wolfensberger, U. (1996). Does a profound change in toilet-training affect development of bowel and bladder control? *Dev. Med. Child Neurol.* 38, 1106–1116. doi: 10.1111/j.1469-8749.1996.tb15074.x

Launes, J., Hokkanen, L., Laasonen, M., Tuulio-Henriksson, A., Virta, M., Lipsanen, J., et al. (2014). Attrition in a 30-year follow-up of a perinatal birth risk cohort: factors change with age. *PeerJ* 2:e480. doi: 10.7717/peerj.480

Lienhard, K., Schneider, D., and Maffiuletti, N. A. (2013). Validity of the optogait photoelectric system for the assessment of spatiotemporal gait parameters. *Med. Eng. Phys*. 35, 500–504. doi: 10.1016/j.medengphy.2012.06.015

Loch, N., Hiller, W., and Witthöft, M. (2011). Der cognitive emotion regulation questionnaire (CERQ). Zeitschrift für Klinische Psychologie und Psychotherapie. 40, 94–106. doi: 10.1026/1616-3443/a000079

Lückert, H.-R. (1957). *Stanford Intelligenz-Test: Handanweisung;[SIT].* Verlag f. Göttingen: Psychologie, Hogrefe.

Magoc, T., and Magoc, D. (2011). Neural network to identify individuals at health risk. *Int. J. Artif. Intel. Appl.* 2, 104–114. doi: 10.5121/ijaia.2011.2208

Matsumuro, M., and Miwa, K. (2019). "Model for data analysis process and its relationship to the hypothesis-driven and data-driven research approaches," in *International Conference on Intelligent Tutoring Systems* (Springer).

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., and Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychol. Methods* 14:126. doi: 10.1037/a0015857

McCrory, C., Dooley, C., Layte, R., and Kenny, R. A. (2015). The lasting legacy of childhood adversity for disease risk in later life. *Health Psychol*. 34:687. doi: 10.1037/hea0000147

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proc. Natl. Acad. Sci*. 108, 2693–2698. doi: 10.1073/pnas.1010076108

Moffitt, T. E., Belsky, D. W., Danese, A., Poulton, R., and Caspi, A. (2017). The longitudinal study of aging in human young adults: knowledge gaps and research agenda. *J. Gerontol. A Biol. Sci. Med. Sci*. 72, 210–215. doi: 10.1093/gerona/glw191

Morfeld, M., Kirchberger, I., and Bullinger, M. (2011). *Fragebogen zum Gesundheitszustand (SF-36)*. Bern: Hogrefe.

Mroczek, D. K., Pitzer, L., Miller, L., Turiano, N., and Fingerman, K. (2011). "The use of secondary data in adult development and aging research," in *Secondary Data Analysis: An Introduction for Psychologists, eds K. H. Trzesniewski*, M. B. Donnellan, and R. E. Lucas (Washington, DC: American Psychological Association), 121–132. doi: 10.1037/12350-007

Naef, N., Wehrle, F., Rousson, V., and Latal, B. (2019). Cohort and individual neurodevelopmental stability between 1 and 6 years of age in children with congenital heart disease. *J. Pediatr*. 215, 83–89.e2. doi: 10.1016/j.jpeds.2019.08.036

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc*. 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

Nettelbeck, T., and Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Pers. Individ. Diff*. 48, 379–384. doi: 10.1016/j.paid.2009.10.032

Neumann, E. (2002). Die Paarbeziehung Erwachsener und Erinnerungen an die Eltern-Kind-Beziehung: eine Untersuchung zur Kontinuität von Bindung. *Zeitschrift Familienforschung* 14, 234–256. Available online at: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-282823

Nishizawa, T., Morita, A., Fujiwara, T., and Kondo, K. (2019). Association between childhood socioeconomic status and subjective memory complaints among older adults: results from the Japan gerontological evaluation study 2010. *Int. Psychogeriatr*. 31, 1699–1707. doi: 10.1017/S1041610219000814

Oh, J., Chopik, W. J., Konrath, S., and Grimm, K. J. (2020). Longitudinal changes in empathy across the life span in six samples of human development. *Soc. Psychol. Pers. Sci*. 11, 244–253. doi: 10.1177/1948550619849429

Paulus, C. (2011). *Der Saarbrücker Persönlichkeitsfragebogen zur Messung von Empathie (SPF-IRI)*. Available online at: http://bildungswissenschaften.uni-saarland.de/personal/paulus/homepage/empathie.html (accessed December 18, 2020)

Peskin, H., and Jones, C. (2015). Early-and late-adolescent predictors of psychological health in adulthood: results from the intergenerational studies. *J. Adult Dev*. 22, 230–238. doi: 10.1007/s10804-015-9214-5

Petermann, F., and Petermann, U. (2012). *Wechsler Adult Intelligence Scale–Fourth Edition. Deutschsprachige Adaptation der WAIS-IV von D. Wechsler*. Frankfurt/Main: Pearson Assessment.

Pienta, A. M., and Lyle, J. (2018). Retirement in the 1950s: rebuilding a longitudinal research database. *IASSIST Q*. 42, 1–9. doi: 10.29173/iq19

Poulton, R., Moffitt, T. E., and Silva, P. A. (2015). The Dunedin multidisciplinary health and development study: overview of the first 40 years, with an eye to the future. *Soc. Psychiatry Psychiatr. Epidemiol*. 50, 679–693. doi: 10.1007/s00127-015-1048-8

Prader, A., Largo, R. H., Molinari, L., and Issler, C. (1989). Physical growth of Swiss children from birth to 20 years of age. First Zurich longitudinal study of growth and development. *Helvetica Paediatr. Acta Suppl*. 52, 1–125.

Raven, J. C. (1947). *Progressive Matrices 1947, Sets A, AB, B [and] Guide to Using Progressive Matrices (1947), Sets A, AB, B*. London: Lewis.

Raven, J. C. (1956). *Manuel D'Instruction du Matrix (Matrices Progressives, 1947)*. Paris: Editions Scientifiques et Psychotechniques..

Regard, M., Strauss, E., and Knapp, P. (1982). Children's production on verbal and non-verbal fluency tasks. *Percept. Motor Skills* 55, 839–844. doi: 10.2466/pms.1982.55.3.839

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.). *Arch. Psychol*. 28, 215–285.

Roche, A. F. (1992). *Growth, Maturation, and Body Composition: The Fels Longitudinal Study 1929-1991*. Cambridge: Cambridge University Press.

Roseboom, T., de Rooij, S., and Painter, R. (2006). The Dutch famine and its long-term consequences for adult health. *Early Hum. Dev*. 82, 485–491. doi: 10.1016/j.earlhumdev.2006.07.001

Rosenzweig, S., Fleming, E. E., Rosenzweig, L., Duhm, E., and Hansen, J. (1948). *Der Rosenzweig P-F Test: The Picture Frustration Study von Saul Rozenzweig*. Göttingen: Verlag fur Psychologie Hogrefe.

Rousson, V., and Gasser, T. (2004). Simple component analysis. *J. R. Stat. Soc. Series C*. 53, 539–555. doi: 10.1111/j.1467-9876.2004.05359.x

Ryff, C. D., and Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *J. Pers. Soc. Psychol*. 69:719. doi: 10.1037/0022-3514.69.4.719

Sabo, R., Wang, A., Deng, Y., Sabo, C., and Sun, S. (2017). Relationships between childhood growth parameters and adult blood pressure: the Fels Longitudinal Study. *J. Dev. Orig. Health Dis*. 8, 113–122. doi: 10.1017/S2040174416000520

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Rev*. 103:403. doi: 10.1037/0033-295X.103.3.403

Sanders, A. E. (2016). Shifting the focus of aging research into earlier decades of life. *Oral Dis*. 22, 166–168. doi: 10.1111/odi.12431

Schaefer, J. D., Caspi, A., Belsky, D. W., Harrington, H., Houts, R., Israel, S., et al. (2015). Early-life intelligence predicts midlife biological age. *J. Gerontol. B Psychol. Sci. Soc. Sci*. 71, 968–977. doi: 10.1093/geronb/gbv035

Schmidt, S., Muehlan, H., and Brähler, E. (2016). *AAS-R-Revised Adult Attachment Scale: Deutsche Version*. Göttingen: Hogrefe.

Schuhmacher, J., Wilz, G., Gunzelmann, T., and Brähler, E. (2000). Die sense of coherence scale von Antonovsky. Teststatistische Überprüfung in einer representativen Bevölkerungsstichprobe und Konstruktion einer Kurzskala. *PPmP Psychother Psychosom med Psychol*. 50, 472–482. doi: 10.1055/s-2000-9207

Siedler, T., Schupp, J., and Wagner, G. G. (2011). "Innovative methods within the context of secondary data: examples from household panel surveys," in *Secondary Data Analysis: An Introduction for Psychologists*, eds K. H. Trzesniewski, M. B. Donnellan, and R. E. Lucas (Washington, DC: American Psychological Association), 103–118.

Snijders-Oomen, N. (1977). *Snijders-Oomen nicht verbale Intelligenztestreihe: SON 2 1/2-7*. Groningen: Wolters-Noordhoff.

Stafford, M., Gale, C. R., Mishra, G., Richards, M., Black, S., Kuh, D. L. (2015). Childhood environment and mental wellbeing at age 60-64 years: prospective evidence from the MRC national survey of health and development. *PLoS ONE* 10:e0126683. doi: 10.1371/journal.pone.0126683

Staudinger, U. M., Fleeson, W., and Baltes, P. B. (1999). Predictors of subjective physical health and global well-being: similarities and differences between the United States and Germany. *J. Pers. Soc. Psychol*. 76:305. doi: 10.1037/0022-3514.76.2.305

Stevenson, A. J., McCartney, D. L., Hillary, R. F., Redmond, P., Taylor, A. M., Zhang, Q., et al. (2019). Childhood intelligence attenuates the association

between biological ageing and health outcomes in later life. *Trans. Psychiatry* 9, 1–8. doi: 10.1038/s41398-019-0657-5

Steyer, R., Schwenkmezger, P., Notz, P., and Eid, M. (1997). *Der mehrdimensionale Befindlichkeitsfragebogen (MDBF). [The Multidimensional Mood Questionnaire (MMQ)]*. Göttingen: Hogrefe.

Strauss, E., Sherman, E. M. S., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, 3rd Edn. New York, NY: Oxford University Press.

Sun, S. S., Liang, R., Huang, T. T.-K., Daniels, S. R., Arslanian, S., Liu, K., et al. (2008). Childhood obesity predicts adult metabolic syndrome: the Fels Longitudinal Study. *J. Pediatr*. 152, 191–200.e1. doi: 10.1016/j.jpeds.2007.07.055

Tanner, J. (1962). *Growth at Adolescence*. Oxford: Blackwell Scientific Publications.

Tanner, J. M. (1981). *A History of the Study of Human growth*. Cambridge: Cambridge University Press.

Tanner, J. M. (1998). *A Brief History of the Study of Human Growth*. Cambridge: The Cambridge Encyclopedia of Human Growth and Development University Press, 2–7.

Taylor, A. M., Pattie, A., and Deary, I. J. (2018). Cohort profile update: the Lothian birth cohorts of 1921 and 1936. *Int. J. Epidemiol*. 47, 1042–1042r. doi: 10.1093/ije/dyy022

Terman, L. M., and Merrill, M. A. (1937). *Measuring Intelligence: A Guide to the Administration of the New Revised Stanford-Binet tests of Intelligence*. Boston: Houghton Mifflin.

Terman, L. M., and Merrill, M. A. (1960). *Stanford-Binet Intelligence scale: Manual for the Third Revision, Form L-M*. Boston: Houghton Mifflin.

Terman, L. M. (1916). *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale*. Boston: Houghton Mifflin.

Terman, L. M., Merrill, M. A., and Lückert, H.-R. (1965). *Stanford-Binet Intelligenz-Test, SIT*. Handanweisung: Verlag f. Psychologie Hogrefe.

Tewes U, Titze I. (1983). Untersuchungen zur Anwendung des HAWIK in der klinischen und sonderpdagogischen Diagnostik [Studies on the application of HAWIK in clinical and special educational diagnosis]. *Zeitschrift Differentielle Diagnostische Psychol*. 4, 179-201.

The Federal Assembly of the Swiss Confederation (2014). *Federal Act on Research Involving Human Beings*. Available online at: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html (accessed December 18, 2020)

Thodberg, H. H., Jenni, O. G., Caflisch, J., Ranke, M. B., and Martin, D. D. (2009). Prediction of adult height based on automated determination of bone age. *J. Clin. Endocrinol. Metabol.* 94, 4868–4874. doi: 10.1210/jc.2009-1429

Unterrainer, J. M., Rahm, B., Kaller, C. P., Wild, P. S., Münzel, T., Blettner, M., et al. (2019). Assessing planning ability across the adult life span in a large population-representative sample: reliability estimates and normative data for the tower of London (TOL-F) task. *J. Int. Neuropsychol. Soc.* 25, 520–529. doi: 10.1017/S1355617718001248

Van den Broek, T., and Fleischmann, M. (2019). Prenatal famine exposure and mental health in later midlife. *Aging Ment. Health* 23, 166–170. doi: 10.1080/13607863.2017.1402293

Veale, J. F. (2014). Edinburgh handedness inventory–short form: a revised version based on confirmatory factor analysis. *Laterality* 19, 164–177. doi: 10.1080/1357650X.2013.783045

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer.

Voorpostel, M., Tillmann, R., Lebert, F., Kuhn, U., Lipps, O., Ryser, V.-A., et al. (2016). *Swiss Household Panel User Guide (1999–2015)*. Lausanne: FORS.

Wadsworth, M., Kuh, D., Richards, M., and Hardy, R. (2006). Cohort profile: the 1946 national birth cohort (MRC national survey of health and development). *Int. J. Epidemiol.* 35, 49–54. doi: 10.1093/ije/dyi201

Weiss, D., Job, V., Mathias, M., Grah, S., and Freund, A. M. (2016). The end is (not) near: aging, essentialism, and future time perspective. *Dev. Psychol.* 52:996. doi: 10.1037/dev0000115

Williamson, W., and Leroi, I. (2019). Thinking about dementia: is childhood too early? *Int. Psychogeriatr.* 31, 1689–1690. doi: 10.1017/S1041610219001157

Willich, O., and Friese, H. (1994). Der Hamburg-Wechsler-Intelligenztest für Kinder revision 1983 (HAWIK-R). *Diagnostica*. 40, 172–189.

Wirtz, M. A., Morfeld, M., Glaesmer, H., and Brähler, E. (2018). Konfirmatorische Prüfung der Skalenstruktur des SF-12 Version 2.0 in einer deutschen bevölkerungs-repräsentativen Stichprobe [Confirmatory analysis of the SF-12 version 2.0 scale structure in a representative German sample.]. *Diagnostica*. 64, 84–96. doi: 10.1026/0012-1924/a000194

World Health Organization (1948). *Preamble to the Constitution of WHO as adopted by the International Health Conference, New York, 19 June—22 July 1946; Signed on 22 July 1946 by the Representatives of 61 States (Official Records of WHO, no. 2. p. 100.) and Entered Into Force on 7 April 1948*. Geneva: WHO.

World Health Organization (2016). *Global Strategy and Action Plan on Ageing and Health (2016–2020)*. Geneva: World Health Organization.

Young, A. F., Powers, J. R., and Bell, S. L. (2006). Attrition in longitudinal studies: who do you lose? *Aust. N. Zeal. J. Public Health* 30, 353–361. doi: 10.1111/j.1467-842X.2006.tb00849.x

Zimmermann, P., and Fimm, B. (2009). *Testbatterie zur Aufmerksamkeitsprüfung-Version 2.2:(TAP);[Handbuch]*. Herzogenrath: Psytest.

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.

The handling Editor declared a shared affiliation, though no other collaboration, with authors BL and OGJ at time of review.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 26 February 2021
doi: 10.3389/fnhum.2021.623766

Object-Location Memory Training in Older Adults Leads to Greater Deactivation of the Dorsal Default Mode Network

Ania Mikos[1], Brigitta Malagurski[1], Franziskus Liem[1], Susan Mérillat[1] and Lutz Jäncke[1,2]*

[1] University Research Priority Program "Dynamics of Healthy Aging", University of Zurich, Zurich, Switzerland, [2] Division of Neuropsychology, Institute of Psychology, University of Zurich, Zurich, Switzerland

Substantial evidence indicates that cognitive training can be efficacious for older adults, but findings regarding training-related brain plasticity have been mixed and vary depending on the imaging modality. Recent years have seen a growth in recognition of the importance of large-scale brain networks on cognition. In particular, task-induced deactivation within the default mode network (DMN) is thought to facilitate externally directed cognition, while aging-related decrements in this neural process are related to reduced cognitive performance. It is not yet clear whether task-induced deactivation within the DMN can be enhanced by cognitive training in the elderly. We previously reported durable cognitive improvements in a sample of healthy older adults (age range = 60–75) who completed 6 weeks of process-based object-location memory training (N = 36) compared to an active control training group (N = 31). The primary aim of the current study is to evaluate whether these cognitive gains are accompanied by training-related changes in task-related DMN deactivation. Given the evidence for heterogeneity of the DMN, we examine task-related activation/deactivation within two separate DMN branches, a ventral branch related to episodic memory and a dorsal branch more closely resembling the canonical DMN. Participants underwent functional magnetic resonance imaging (fMRI) while performing an untrained object-location memory task at four time points before, during, and after the training period. Task-induced (de)activation values were extracted for the ventral and dorsal DMN branches at each time point. Relative to visual fixation baseline: (i) the dorsal DMN was deactivated during the scanner task, while the ventral DMN was activated; (ii) the object-location memory training group exhibited an increase in dorsal DMN deactivation relative to the active control group over the course of training and follow-up; (iii) changes in dorsal DMN deactivation did not correlate with task improvement. These results indicate a training-related enhancement of task-induced deactivation of the dorsal DMN, although the specificity of this improvement to the cognitive task performed in the scanner is not clear.

Keywords: healthy aging, cognitive training, task-based fMRI, default mode network, object-location memory

OPEN ACCESS

Edited by:
Torsten Wüstenberg, University of Heidelberg, Germany

Reviewed by:
Tobias Gleich, Charité – Universitätsmedizin Berlin, Germany
Stephan Heinzel, Freie Universität Berlin, Germany

****Correspondence:***
Ania Mikos
ania.mikos@uzh.ch

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

Received: *30 October 2020*
Accepted: *20 January 2021*
Published: *26 February 2021*

Citation:
Mikos A, Malagurski B, Liem F, Mérillat S and Jäncke L (2021) Object-Location Memory Training in Older Adults Leads to Greater Deactivation of the Dorsal Default Mode Network. Front. Hum. Neurosci. 15:623766. doi: 10.3389/fnhum.2021.623766

INTRODUCTION

While normal aging is accompanied by cognitive declines in processing speed, working memory, episodic memory, and reasoning (Park et al., 2002; Salthouse, 2009), promising evidence indicates that the cognitive system demonstrates plasticity across the entire life span (Hertzog et al., 2008) and that cognitive training can improve performance in many of these domains in older adults (Karbach and Verhaeghen, 2014; Chiu et al., 2017). A previously published randomized controlled cognitive training trial from our laboratory targeted an episodic memory process involving the formation of object-location associations (i.e., object-location memory; OLM), which is significantly impaired in old age (Kessels et al., 2007; Old and Naveh-Benjamin, 2008). The OLM training was designed to be process-based, targeting the efficiency of the basic cognitive processes involved through repeated practice (Lövdén et al., 2010). In comparison to an active control training, OLM training led to improvements in the trained task as well as transfer to the domains of spatial memory and reasoning that were maintained 4 months after training (Zimmermann et al., 2016). In the present study, we evaluate whether these training-related behavioral gains were accompanied by changes in neural activity in an important large-scale network (i.e., default mode network, DMN).

While the evidence for training-related improvements in brain structure (e.g., white matter integrity, cortical thickness) is rather limited in older adults (Boyke et al., 2008; Engvig et al., 2012, 2010; Lövdén et al., 2012), changes in neural activity have been more commonly observed (Park and Bischof, 2013). Functional magnetic resonance imaging (fMRI) studies have demonstrated both increases and decreases following cognitive training in task-related activity, mostly in fronto-parietal brain areas (Duda and Sweet, 2019; van Balkom et al., 2020). Reduced activations following training have been interpreted as reflecting increased neural efficiency (Brehmer et al., 2011; Heinzel et al., 2016, 2014), while increased activation can be considered facilitative in older adults who need to overcome age-related neural deficits (Goh and Park, 2009). Further neuroplastic effects have been observed in large-scale brain networks, functionally connected networks of discrete brain regions that are increasingly thought to be essential for successful cognition (Seeley et al., 2007; Wig, 2017). In healthy elderly, cognitive training has been associated with increased intra-network connectivity in the DMN, frontoparietal network (FPN), and salience network (Cao et al., 2016; De Marco et al., 2016), and with increased anti-correlation between the task-negative DMN and the task-positive FPN (Cao et al., 2016; Lebedev et al., 2018).

The DMN, the most widely studied of the large-scale brain networks, demonstrates elevated activity during undirected passive tasks and reduced activity across a wide range of external cognitive tasks (Shulman et al., 1997; Binder et al., 1999; Mazoyer et al., 2001; Buckner et al., 2008). DMN regions, including the posterior cingulate cortex, medial prefrontal cortex, inferior parietal lobule, and medial and lateral temporal lobes, form an intrinsic connectivity network at rest (Greicius et al., 2003) which is mirrored by direct structural connections (Greicius et al., 2009). DMN activation has been largely linked to self-relevant, internally directed information processing (Andrews-Hanna et al., 2010a), including autobiographical memory, self-reflective thought (Gusnard et al., 2001), envisioning future events, mind wandering (Mason et al., 2007), and considering the thoughts and perspectives of others (Raichle et al., 2001; Raichle and Snyder, 2007; Buckner et al., 2008). DMN suppression during specific goal-directed behaviors (i.e., task-induced deactivation, TID) can be interpreted as suspension of this default mode activity (Raichle et al., 2001). TID is positively associated with performance on a wide range of cognitive tasks and increases with task difficulty or cognitive load (McKiernan et al., 2003; Anticevic et al., 2012). Thus, TID is thought to facilitate performance on external cognitive tasks by diverting resources from unconstrained, distracting default mode processes (Binder et al., 1999; Andrews-Hanna, 2012; Anticevic et al., 2012).

In healthy aging, the DMN demonstrates reduced resting state activity and within-network connectivity (Andrews-Hanna et al., 2007; Damoiseaux et al., 2008; Koch et al., 2010; Vidal-Piñeiro et al., 2014; Staffaroni et al., 2018), as well as reduced task-induced deactivation. Compared to younger adults, older adults demonstrate reduced TID within the DMN in tasks of semantic classification (Lustig et al., 2003), memory (Grady et al., 2006), working memory (Sambataro et al., 2010), and visuospatial planning (Spreng and Schacter, 2012). These age-related reductions in TID have been associated with slower performance on tasks of working memory (Brown et al., 2018) and spatial judgment (Park et al., 2010), and worse performance on tests of face-name associative memory (Miller et al., 2008) and verb generation (Persson et al., 2007). Reduced TID is associated with subclinical cognitive decline even in middle-age, suggesting that it may be an early marker for subtle cognitive decline (Hansen et al., 2014).

In the present manuscript, we investigate whether the DMN demonstrates training-related plasticity in healthy older adults, particularly in terms of its capacity for TID. However, the DMN may not deactivate uniformly in response to external cognitive tasks. Growing evidence from detailed high-resolution analyses of single individuals calls into question the idea of a unitary canonical DMN that was historically defined based on group-averaged data, instead suggesting several interwoven networks (Buckner and DiNicola, 2019). Several proposed fractionations have been identified in the resting state literature (i.e., anterior vs. posterior; ventral vs. dorsal) (Damoiseaux et al., 2008; Andrews-Hanna et al., 2010b; Chen et al., 2017). In the task-based domain, Mayer et al. (2010) distinguished between "core DMN" regions that deactivate indiscriminately in response to cognitive demand and other subregions whose deactivation depends on the specific task. Several schemes point to a function involving episodic memory. Andrews-Hanna et al. (2010b) identified a medial temporal lobe subsystem involved in memory-based reconstruction, while the ventral DMN (retrosplenial cortex/medial temporal lobe intrinsic connectivity network) identified in the Shirer atlas (Shirer et al., 2012) exhibited increased functional connectivity during subject-driven episodic memory recall compared to a rest state. Here, we examine

task-related activity during performance of an untrained OLM task in the two DMN subnetworks from Shirer et al. (2012): the ventral DMN (vDMN), associated with episodic memory, and the dorsal DMN (dDMN), which more closely resembles the canonical DMN.

In summary, the current study employs task-based fMRI to further elucidate properties of two DMN branches in the context of a randomized controlled cognitive training study, which previously demonstrated improved training-related task performance and durable cognitive transfer effects. The primary aims of the present study are: (i) to investigate the patterns of activation/deactivation within the ventral and dorsal DMN networks during performance on an OLM task at baseline before initiation of training (cross-sectional analysis), (ii) to examine whether the two subnetworks respond to OLM training compared to active control training over the course of several assessments (longitudinal analysis), and (iii) to assess whether any training-related (de)activation changes correlate with improvements in scanner task performance.

MATERIALS AND METHODS

Screening and Training

Participants were recruited at lectures for senior citizens at the University of Zurich, through newspaper articles, advertisements in magazines, public talks, flyers, and word of mouth. All participants gave written informed consent. Inclusion criteria were age between 60 and 75 years, right-handedness, native German speaker or fluent in German, basic computer and internet experience, and access to a computer as well as the internet during the training period. Exclusion criteria were history of previous or current neurological and psychiatric disorders or substance use negatively affecting brain function, sensory and motor deficiencies hindering conduction of training and outcome measurements, violation of MRI safety requirements, and participation in a training study within the last 5 years. In addition, participants who scored 1.5 SD below age-, gender-, and education-specific norms in more than one subtest of the Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Assessment Battery (CERAD-NAB; Berres et al., 2000) or had a sum score >5 in the short version of the Geriatric Depression Scale (GDS; Sheikh and Yesavage, 1986) were excluded. Participants were informed about the risks of participation in MRI studies via an information form that they were asked to sign.

After screening, 67 participants were randomized to the object-location memory training condition (OLM group; $N = 36$; mean age = 66.75 $\pm$ 4.17) or a control training condition targeting visual perception (active control group, $N = 31$; mean age = 68.23 $\pm$ 3.84). There were no baseline differences between groups on demographic or cognitive screening measures (Zimmermann et al., 2016).

Object-location memory and active control training comprised two phases with 15 sessions each that participants had to complete within 3 weeks. A 1-week break separated the phases. Participants trained at home on their personal computers, with each session lasting 30–45 min. Participants were informed that the training software permitted the completion of only one session per day and that they would be contacted by e-mail or phone in case of no recorded training sessions on three consecutive days.

As described in Zimmermann et al. (2016), OLM training consisted of object-location, shape-location, and landmark-location tasks in which cued recall for associations was practiced. Each trial consisted of an encoding phase in which N associations had to be encoded, a 20-s distractor task, and a retrieval phase. Task difficulty was adapted to individual performance by increasing or decreasing N of to-be-encoded associations by one. Participants started the first session on the lowest level of difficulty with two item-location associations. The highest possible level of difficulty consisted of 21 item-location associations. Individual performance was assessed for the object-, shape-, and landmark-location tasks separately. Task difficulty was increased in the next training session if performance was greater than 70% and was decreased if performance was below 50%. Feedback was given on the percentage of correctly recalled associations and the level of difficulty achieved.

Stimuli for each of the three training tasks were as follows. For the *object-location task, N* objects, drawn from a database of 245 colored drawings of everyday objects (Snodgrass and Vanderwart, 1980; Rossion and Pourtois, 2004), were presented sequentially in a 5-×-6 grid for 4 s followed by an ISI of 0.5 s. For the *shape-location task*, N shapes, drawn from a set of 29 self-created shapes in nine different colors were presented for $N \times 3$ s. For the *landmark-location task*, stimuli were drawn from a database of 261 photographs of real-world buildings (retrieved from the internet, excluding highly salient or famous buildings) and presented on a 6-×-6 grid with a different self-created city map superimposed in each training session. During the retrieval phase, the previously presented stimuli were displayed along with the empty grid and participants had to select the cell in which the stimulus was initially presented with a mouse click.

Active control training included two phases of a visual perception task, separated by a distractor task. Stimuli and task duration were matched to those of the OLM training tasks. For the *object-perception task,* two 1-×-10 grids filled with objects were presented, and participants had to click with the mouse on the one object that differed between the two grids. For the *shape-perception task,* participants selected a target shape within a 6-×-6 grid filled with 36 shapes. For the *landmark-perception task,* participants selected a target building from within a city map filled with 21 buildings.

Five participants completed only 27–29 training sessions [OLM group: 29 sessions ($N = 1$), 27 sessions ($N = 1$); active control group: 29 sessions ($N = 2$), 28 sessions ($N = 1$)] because of technical and scheduling problems. Two participants from the OLM group completed one additional training session (in the second training phase: $N = 1$, in the follow-up period: $N = 1$).

Scanner Task

Both groups underwent fMRI while performing an object-location memory task during four sessions: baseline before initiation of either training condition (T1), after the first 3-week

training phase (T2), at the conclusion of the 6-week training period (T3), and 4 months after training completion (T4). A block design was used, consisting of 24 blocks equally divided between two runs. Each run lasted approximately 14 min. The order of the two runs was counterbalanced across participants of both groups. Before entering the scanner, participants completed five practice trials of the task on a laptop. In the MR scanner, participants lay comfortably in supine position with padded head holders restricting head movements. Stimuli were presented using the software Presentation (NeuroBehavioral Systems; NBS) which also recorded behavioral performance. The stimuli were presented on MR compatible goggles (Resonance Technology Inc., Northridge, CA, United States) which could be adjusted for poor eyesight. Object stimuli were drawn from the Bank of Standardized Stimuli (BOSS) created by Brodeur et al. (2010). For the present study, the 480 colored everyday objects were divided into two sets with comparable familiarity and object identity ratings (provided by Brodeur et al., 2010). After eliminating photo stimuli which were semantically very similar (i.e., wire, cable) and stimuli that were mostly white in color because of the white screen background, 144 stimuli for each fMRI run remained. The same number of similarly rated stimuli was used for each encoding phase. Stimuli were used only once within both runs. They were different from the object stimuli used for the OLM and active control training tasks.

A block of the scanner OLM task included four phases: encoding, distractor, recognition, and visual fixation baseline. Stimuli were presented in a 5 × 5-grid on a white background. During the encoding phase, six objects were presented serially in one of the grid cells, each for 3000 ms (ISI = 0 ms). During the distractor phase, participants were asked to solve a 1-back task. Black arrows were presented consecutively in random order and participants had to decide whether the presented arrow pointed toward the same direction as the previous one and to indicate their decisions by pressing one of two buttons on an MR compatible response box with their left or right thumbs (same direction = left, different direction = right). Each cue lasted for 1000 ms followed by an ISI of 1000 ms. The distractor phase was randomly jittered and lasted for 12000–18000 ms. During the recognition phase, participants were presented the six encoded objects sequentially in the 5 × 5-grid, each for 3000 ms (ISI = 0 ms). Three of the objects appeared in the same locations as during encoding, whereas three objects were presented in locations in which different objects had been displayed during encoding. Participants had to decide whether presented object-location associations were the encoded ones or not and indicate their decisions by pressing one of two buttons on the response box with their left or right thumbs. The subsequent visual fixation baseline phase was jittered and lasted between 9000 and 15000 ms. During this period, a black cross was presented that changed to green 2000 ms before the encoding phase of the next block started.

The scanner task is displayed in **Figure 1**. The behavioral dependent variables of interest include *hits*, defined as the number of correct responses out of 72 for each run. Hits were averaged across the two runs for each time point, yielding the variable *average hits* used in the following analyses. Additionally,

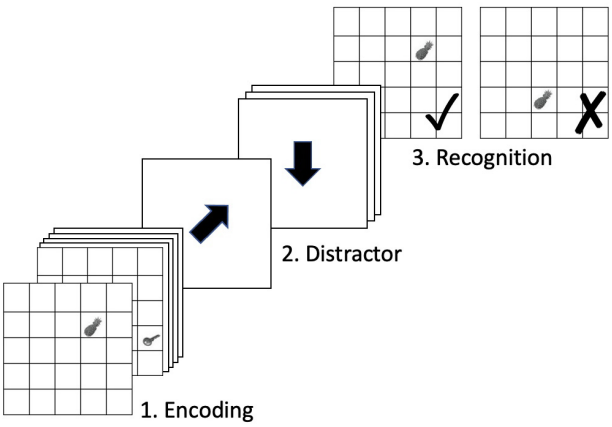


FIGURE 1 | In-scanner OLM task. Figure adapted from Zimmermann (2015).

reaction time (in milliseconds) was averaged across hit trials only and then across runs to provide *hit reaction time*.

Scanner task average hits contributed to the near-transfer spatial episodic memory composite outcome measure of this trial, reported in Zimmermann et al. (2016). In this regard, it is important to note that although the scanner task and OLM training task were somewhat similar, they differed in several important ways. Object stimuli for the scanner and training tasks were drawn from two different sources, and the training task additionally included landmark and shape stimuli. Further, the two tasks differed in the rate of stimulus presentation during encoding, duration and content of the distractor phase, and task demands during retrieval. As such, scanner task average hits is thought to represent near-transfer effects rather than training effects *per se*.

MRI Protocol

Whole brain T2-weighted EPI-BOLD data were acquired with a Philips Achieva 3T TX scanner (Philips Medical Systems, Best, Netherlands) using a 32-channel receiver head coil array. Blood-oxygen-level-dependent (BOLD) fMRI images were generated with a gradient-echo-planar-imaging (EPI) pulse sequence (TR/TE = 2500/30 ms, flip angle = 84°, matrix = 80 × 80, FOV = 240 mm × 240 mm, 44 slices, slice thickness 3 mm, 0.5 mm interslice spacing), that yielded 3 mm × 3 mm × 3 mm voxels. Slices were acquired in descending order and in transverse orientation. Each of the two runs consisted of a total of 335 volumes. Five dummy scans were performed prior to image acquisition aiming to eliminate signals arising from progressive saturation. In addition, a high-resolution T1 anatomic image (TR/TE = 8.1/3.7 ms, flip angle = 8°, matrix 240 × 240, FOV = 240 mm × 240 mm, 160 slices, slice thickness 1.0 mm, 1 mm × 1 mm × 1 mm voxels) was obtained for each subject.

Functional Image Analyses

Imaging data were first transformed into the Brain Imaging Data Structure (BIDS) format (Gorgolewski et al., 2016).

Preprocessing of each functional run was performed using fMRIPrep version 1.2.6 (Esteban et al., 2018), a Nipype (Gorgolewski et al., 2011) based tool, including the following steps: bias field correction, skull stripping, correction for head-motion parameters, slice time correction, co-registration to corresponding structural image [boundary-based registration with 9 degrees of freedom implemented in FreeSurfer v6.0.0 (Greve and Fischl, 2009)], and spatial normalization to MNI space. Motion correcting transformations, T1 weighted transformation and MNI template warp were applied in a single step using antsApplyTransformations v2.1.0 with Lanczos interpolation. Three tissue classes were extracted from T1 images using FSL FAST v5.0.9 (Zhang et al., 2001). Voxels from cerebrospinal fluid and white matter were used to create a mask in turn used to extract physiological noise regressors using aCompCor (Behzadi et al., 2007). The mask was eroded and limited to subcortical regions to limit overlap with gray matter; six principal components were estimated. Framewise displacement (Power et al., 2014) was calculated for each functional run using Nipype implementation. For more details of the pipeline see: http://fmriprep.readthedocs.io/en/latest/workflows.html.

First-level statistical analyses were performed using the general linear model approach (GLM) as implemented in SPM12[1]. Explanatory variables modeling the experimental conditions of the blocked fMRI design comprised the following four conditions: (1) encoding, (2) distractor, (3) recognition, and (4) visual fixation baseline. These four conditions were modeled for each of the two fMRI runs at each session (T1, T2, T3, and T4). In addition, the GLM included the six motion parameters, the framewise displacement, and physiological noise regressors [first principal component from aCompCor (Behzadi et al., 2007)] as obtained from fMRIPrep preprocessing in order to control for physiological and movement confounds. Contrast images were created for: (1) encoding vs. visual fixation baseline and (2) recognition vs. visual fixation baseline. The MarsBaR SPM toolbox (v0.44, marsbar.sourceforge.net) was used to extract contrast values from DMN networks for each run. Regions of interest (ROIs) for the DMN networks were selected from the atlas of Shirer et al. (2012). The atlas comprises 90 functionally derived ROIs across 14 intrinsic connectivity networks, identified via independent component analysis. Functional ROIs were shown to outperform a set of commonly used structural ROIs in classifying subject-driven cognitive states, and patterns of within- and between-network functional connectivity were related to subject-driven cognitive states (Shirer et al., 2012). For the present study, contrast values were extracted for each of the dorsal and ventral default mode networks; **Figure 2** displays the ROIs comprising each network. Each participants' extracted coefficient estimates were averaged across the two runs of each assessment time and then submitted to second-level analysis.

Missing Data

fMRI assessments were completed by all 67 participants with the exception of two participants (one from each group) who did not take part in T4 testing because of medical reasons. Three participants from the control group only completed one of the two runs at T4, two for technical reasons and one for medical reasons. For these three participants, data from the one completed run at T4 were used in statistical analyses.

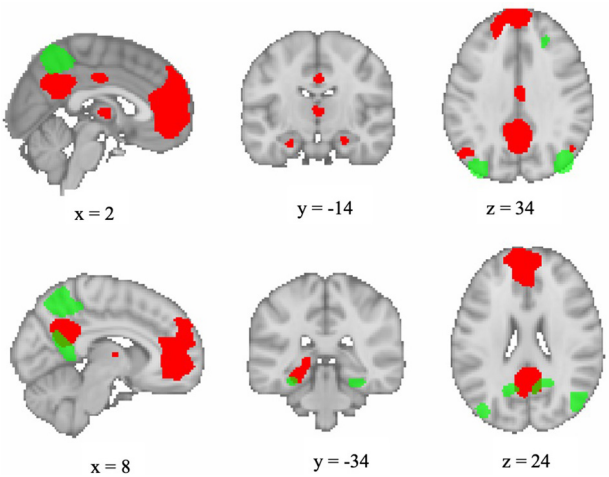


FIGURE 2 | Visualization of functional ROIs comprising dorsal DMN (red) and ventral DMN (green) networks as defined by Shirer et al. (2012). Image generated in FSLeyes version 0.27.0 (https://git.fmrib.ox.ac.uk/fsl/fsleyes/fsleyes/) using a template brain and displayed in neurological orientation.

Statistical Analyses

Scanner task data (average hits and hit reaction time) were first screened for outliers. Excessive values were defined as any value more than three median absolute deviations (MADs) above the median of the sample distribution for each group at each measurement occasion (Leys et al., 2013). Very few outliers were detected (average hits: 1 outlier from each group; hit reaction time: 1 participant from the active control group produced outlier values at T3 and T4). Removal of these outlier values did not change the overall pattern of results; therefore, analysis results from the complete data set are reported below.

For baseline (T1) analyses, we conducted independent samples *t*-tests to evaluate whether there were group differences on the scanner task or network (de)activation levels. Network (de)activation levels were also compared between the encoding and recognition conditions using independent samples *t*-tests.

For longitudinal analyses, linear mixed effects regression (LMER) was used to determine if there were statistically reliable differences for behavioral performance on the scanner task as well as for levels of network (de)activation. Three pairs of models were conducted: one pair of behavioral models with dependent variables of average hits and hit reaction time on the scanner task, and two pairs of network models (one for the dDMN and one for the vDMN), each with dependent variables of (de)activation levels in each of two task conditions (encoding and recognition). Fixed-effects predictors included group (OLM vs. active control) and time of assessment, coded as a factor including levels T2, T3, and T4. Baseline (T1) levels of the behavioral variable (for

[1]https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

the behavioral models) or network (de)activation level (for the network models) were included as covariates. Age and sex were also included as covariates in all models. Participant was modeled as a random effect to account for random variability between individuals. For all LMER models, α was set at 0.025 in order to incorporate a Bonferroni correction reflecting the fact that there were two behavioral models and two models for each of the DMN networks.

All analyses were performed in R version 3.6.0 (R Core Team., 2019). The *Routliers* package (Delacre and Klein, 2019) was used to detect outliers using the Median Absolute Deviation (Leys et al., 2013). LMER models were conducted using the *lme4* package (Bates et al., 2015) with restricted maximal likelihood estimation. Degrees of freedom for the fixed effects were estimated by the Satterthwaite approximation as implemented in the package *lmerTest* (Kuznetsova et al., 2017). The package *emmeans* (Lenth, 2020) was used to evaluate significance of marginal contrasts for training group (OLM versus active control), sex, time (each time point contrasted against the previous time point), and group X time interaction (group contrast evaluated separately at T2, T3, and T4). Simple approximations of between-subjects effect sizes reported as Cohen's *d* (small = 0.2; medium = 0.5; large = 0.8) were produced with the *effectsize* package (Ben-Shachar et al., 2020) using the estimated *t* statistic and observed degrees of freedom. The package *ggplot2* (Wickham, 2016) was used to make the figures.

RESULTS

For visualization purposes, mean scanner task performance and DMN subnetwork (de)activation levels across all four time points for each group are displayed in **Supplementary Figures 1, 2**. The following analyses first address baseline (T1) results and then examine whether training effects can be detected in a longitudinal framework.

Baseline Values and Group Comparability

Baseline Scanner Task Performance

We first verified baseline comparability of behavioral results across groups. Mean baseline values for scanner task outcomes and results of independent samples *t*-tests comparing the two groups are displayed in **Table 1**. There were no significant differences between groups for average hits or hit reaction time on the scanner task at T1.

Baseline DMN Subnetwork Results

Cross-sectional fMRI results from T1 were evaluated in order to characterize baseline patterns of activation across networks/conditions and to determine whether there were any baseline differences by group, age, or sex. As shown in **Figure 3**, the dDMN was deactivated during both task conditions and the vDMN was activated. We conducted independent samples *t*-tests to compare encoding and recognition values for each network. For the dDMN, there was more deactivation during encoding than recognition, $t(132) = -4.78$, $p < 0.001$, $d = -0.83$. For the vDMN, there was more activation during recognition than encoding, $t(132) = -7.69$, $p < 0.001$, $d = -1.34$. In summary, the dDMN was deactivated during both conditions of the memory task, more so during encoding, and the vDMN was activated during both conditions, more so during recognition.

Next, we evaluated baseline comparability of DMN subnetwork (de)activation across groups. Mean contrast values and results of independent samples *t*-tests comparing the two groups are displayed in **Table 2**. Across both networks, there were no significant group differences in contrast values for either task condition. Further regression models were conducted to test if there were effects of age or sex on baseline contrast values. Four separate models were conducted with beta value for dDMN encoding, dDMN recognition, vDMN encoding, and vDMN recognition as the dependent variables. There was no significant effects of age or sex for any of the models (all *p*-values > 0.11). Thus, results suggest that the two groups were comparable in terms of baseline levels of task-related (de)activation, and no age or sex effects were detected.

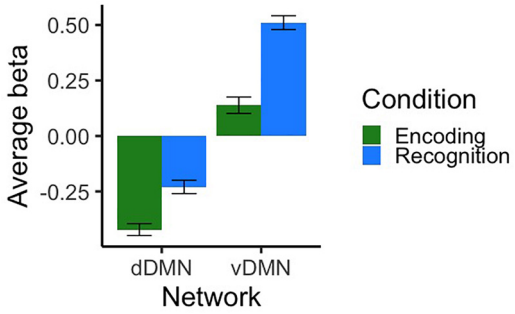


FIGURE 3 | Average beta values representing activation/deactivation at T1 for dorsal and ventral DMN networks during both conditions of the scanner task. Contrasts represent encoding and recognition conditions relative to the visual fixation baseline condition. Error bars represent SEM.

TABLE 1 | Descriptive statistics and results of independent-samples *t*-tests for baseline scanner task performance.

	OLM group		Active control group				
Dependent variable	**Mean**	***SD***	**Mean**	***SD***	***t***	***p***	***d***
Average hits[a]	57.03	4.71	57.44	5.21	−0.34	0.738	−0.08
Hit reaction time (ms)	1292.88	159.96	1240.40	148.53	1.38	0.171	0.34

[a]*Refers to the number of correct responses on the scanner task out of 72 items for each run, averaged across two runs.*

TABLE 2 | Descriptive statistics and results of independent-samples *t*-tests for network activity at baseline.

		OLM group		Active control group				
Network	**Condition**	**Mean**	**SD**	**Mean**	**SD**	***t***	***p***	***d***
Dorsal DMN	Encoding	−0.41	0.23	−0.44	0.21	0.68	0.501	0.17
	Recognition	−0.20	0.23	−0.26	0.27	1.02	0.309	0.25
Ventral DMN	Encoding	0.12	0.32	0.16	0.28	−0.57	0.571	−0.14
	Recognition	0.56	0.21	0.45	0.29	1.80	0.077	0.45

TABLE 3 | Parameter estimates for fixed effects related to the scanner task.

Outcome	**Predictor**	**Estimate**	***SE***	***t***	***p***	***d***
Average hits	Sex[a]	−0.66	0.73	−0.90	0.370	−0.23
	Age	−0.04	0.09	−0.40	0.691	
	T1 average hits	**0.60**	**0.07**	**8.12**	**<0.001**	
	Group[b]	**1.79**	**0.73**	**2.47**	**0.016**	**0.63**
	T3[c]	**1.56**	**0.44**	**3.54**	**<0.001**	
	T4[c]	0.04	0.45	0.08	0.93	
	Group[b] X T2	1.62	0.89	1.83	0.070	0.33
	Group[b] X T3	0.72	0.89	0.81	0.421	0.14
	Group[b]X T4	**3.04**	**0.90**	**3.39**	**<0.001**	**0.60**
Hit RT	Sex[a]	9.78	22.20	0.44	0.662	0.11
	Age	1.02	2.75	0.37	0.711	
	T1 RT	**0.78**	**0.07**	**10.64**	**<0.001**	
	Group[b]	−11.40	22.10	−0.52	0.608	−0.13
	T3[c]	**−32.30**	**11.80**	**−2.74**	**0.007**	
	T4[c]	1.33	12.00	0.11	0.911	
	Group[b] X T2	−0.24	25.90	−0.01	0.993	0.00
	Group[b] X T3	14.42	25.90	0.56	0.580	0.11
	Group[b] X T4	−48.46	26.30	−1.84	0.068	−0.35

[a] *Male contrasted against female.*
[b] *OLM training group contrasted against active control group.*
[c] *Contrasted against the preceding time of assessment.*
P-values are uncorrected. Cohen's d for between-subjects effects based on parameter estimates from linear mixed effects regression models. Statistically significant effects surviving Bonferroni correction appear in bold. RT, reaction time.

Longitudinal Analyses

Effect of Training Group on Scanner Task Performance

The first longitudinal analysis evaluated the impact of training on scanner task performance. **Table 3** shows the results of the LMER models evaluating fixed effects of group and time on the behavioral fMRI dependent variables (average hits and hit reaction time on the scanner task). For average hits, the effect of T3 against T2 was significant, indicating improved performance across both groups from T2 to T3. There was a significant main effect of group, whereby the OLM group performed better across T2–T4. Further, there was a significant interaction such that at T4, the OLM group performed significantly better than the control group (see **Figure 4**). These OLM training effects constituted medium effect sizes (main effect of group: $d = 0.64$, effect of group at T4: $d = 0.6$). There were no significant effects of age or sex on average hits.

For hit reaction time, the effect of T3 against T2 was significant, indicating that reaction time decreased across both groups from T2 to T3. There was no significant main effect of group and no group X time interactions. There were no significant effects of age or sex on hit reaction time.

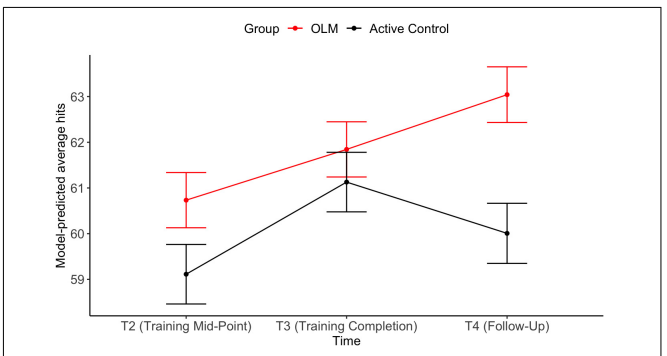


FIGURE 4 | Estimated marginal means from LMER models for average hits during the scanner task. Error bars represent SEM.

Effect of Training Group on Network Activation/Deactivation

The next longitudinal analyses evaluated whether there were statistically reliable differences in network (de)activation as a result of OLM training when controlling for baseline (T1) network activity. Performance on the scanner task (average hits) was included as an additional covariate, since behavioral analyses revealed a group difference on this variable.

Dorsal DMN

Fixed effects results of the LMER models for the dDMN are summarized in **Table 4**. For both conditions, baseline deactivation was positively and significantly related to the dependent variable (DV), indicating that T2, T3, and T4 deactivation levels are positively correlated with baseline (T1) deactivation.

The group contrast predictor was significant only in the encoding model, indicating that across T2–T4 the OLM group exhibited more deactivation compared to the active control group ($d = -0.61$). Examination of the group contrast at each time point revealed that there were significant group differences at T3 and T4 (both $d = -0.43$). Estimated marginal means from these group by time interactions are shown in **Figure 5**. Additionally, there was a significant effect of sex during the encoding condition, with men showing greater dDMN deactivation than women ($d = -0.62$). Because sex and group were both significant in the encoding model, the model was re-run allowing group and sex to interact. However, this interaction was not significant ($p = 0.85$), indicating that the stronger dDMN deactivation in men does not seem to drive the OLM training group effect.

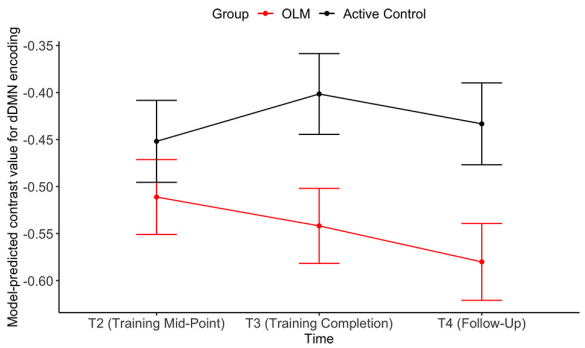


FIGURE 5 | Estimated marginal means representing group X time interactions in the dDMN encoding LMER model. Error bars represent SEM.

For the recognition model, group was not significant as a main effect, and there were no significant group X time interactions. There was no significant association between dDMN deactivation and scanner task average hits or age in either condition.

Ventral DMN

Fixed effects results of the LMER models for the vDMN are summarized in **Table 5**. For both conditions, baseline level of

TABLE 4 | Dorsal DMN: Parameter estimates for fixed effects related to activation/deactivation for both conditions of the scanner task.

Condition	Predictor	Estimate	*SE*	*t*	*p*	*d*
Encoding	**Sex[a]**	**−0.12**	**0.05**	**−2.45**	**0.017**	**−0.62**
	Age	0.00	0.01	0.24	0.808	
	T1 deactivation	**0.76**	**0.11**	**7.16**	**<0.001**	
	Average hits	0.00	0.00	0.68	0.500	
	Group[b]	**−0.12**	**0.05**	**2.42**	**0.019**	**−0.61**
	T3[c]	0.01	0.03	0.32	0.747	
	T4[c]	−0.04	0.03	−1.17	0.245	
	Group[b] X T2	−0.06	0.06	−1.01	0.315	−0.18
	Group[b] X T3	**−0.14**	**0.06**	**−2.40**	**0.018**	**−0.43**
	Group[b] X T4	**−0.15**	**0.06**	**−2.44**	**0.016**	**−0.43**
Recognition	Sex[a]	−0.04	0.04	−1.02	0.314	−0.26
	Age	0.00	0.01	0.61	0.543	
	T1 deactivation	**0.57**	**0.09**	**6.34**	**<0.001**	
	Average hits	0.01	0.00	1.49	0.137	
	Group[b]	−0.06	0.04	−1.29	0.202	−0.33
	T3[c]	0.00	0.03	−0.08	0.939	
	T4[c]	0.02	0.03	0.77	0.442	
	Group[b] X T2	−0.02	0.05	−0.34	0.738	−0.06
	Group[b] X T3	−0.10	0.05	−1.95	0.054	−0.36
	Group[b] X T4	−0.05	0.05	−0.91	0.367	−0.16

[a] *Male contrasted against female.*
[b] *OLM training group contrasted against active control group.*
[c] *Contrasted against the preceding time of assessment.*
P-values are uncorrected. Cohen's d for between-subjects effects based on parameter estimates from linear mixed effects regression models. Statistically significant effects surviving Bonferroni correction appear in bold.

TABLE 5 | Ventral DMN: Parameter estimates for fixed effects related to activation/deactivation for both conditions of the scanner task.

Condition	Predictor	Estimate	*SE*	*t*	*p*	*d*
Encoding	Sex[a]	−0.08	0.05	−1.55	0.127	−0.39
	Age	0.01	0.01	1.33	0.190	
	T1 activation	**0.67**	**0.08**	**8.02**	**<0.001**	
	Average hits	0.01	0.00	1.78	0.076	
	Group[b]	0.03	0.05	0.61	0.547	0.15
	T3[c,d]	0.07	0.03	2.11	0.036	
	T4[c]	**−0.10**	**0.03**	**−2.97**	**0.004**	
	Group[b] X T2	0.07	0.06	1.19	0.236	0.20
	Group[b] X T3	0.02	0.06	0.30	0.768	0.05
	Group[b] X T4	0.00	0.06	−0.06	0.956	−0.01
Recognition	Sex[a]	0.01	0.04	0.37	0.713	0.09
	Age	0.01	0.00	1.30	0.199	
	T1 activation	**0.77**	**0.08**	**9.60**	**<0.001**	
	Average hits	**0.01**	**0.00**	**2.63**	**0.009**	
	Group[b]	0.05	0.04	1.22	0.226	0.31
	T3[c,d]	0.06	0.03	2.20	0.030	
	T4[c]	**−0.08**	**0.03**	**−2.86**	**0.005**	
	Group[b] X T2	0.07	0.05	1.46	0.147	0.26
	Group[b] X T3	0.02	0.05	0.38	0.702	0.07
	Group[b] X T4	0.06	0.05	1.10	0.273	0.19

[a] *Male contrasted against female.*
[b] *OLM training group contrasted against active control group.*
[c] *Contrasted against the preceding time of assessment.*
[d] *Does not survive multiple comparison corrections (alpha = 0.05/2).*
P-values are uncorrected. Cohen's d for between-subjects effects based on parameter estimates from linear mixed effects regression models. Statistically significant effects surviving Bonferroni correction appear in bold.

activation was positively and significantly related to the DV, indicating that T2, T3, and T4 activation levels are positively correlated with baseline (T1) activation.

Group was not significant as a main effect for either condition. Effects of time were significant in both the encoding and recognition models. Specifically, when averaging across groups, activation increased from T2 to T3, but these differences did not survive correction for multiple comparisons. Activation then decreased significantly across groups from T3 to T4. There were no significant effects of age or sex in either condition. Finally, scanner task average hits were positively related to vDMN activation during the recognition condition.

Exploratory analyses: Dorsal DMN nodes

Post-hoc tests were conducted in order to examine whether the OLM-training-related increase in dDMN deactivation during encoding occurred across all constituent nodes or was driven by changes in only a select few. A series of LMER models was conducted with the same fixed effects, covariates, and random effects as for the longitudinal network analyses. This time, the DVs were the encoding contrast values for each of the nine functional ROIs comprising the dDMN (Shirer et al., 2012). The beta values for the group contrast (across T2–T4) from each model are shown in **Table 6**. The OLM group exhibited significantly greater deactivation in the medial prefrontal cortex, right superior frontal gyrus, and midcingulate cortex compared to the active control group. Only the medial prefrontal cortex and midcingulate cortex group differences survived Bonferroni-Holm correction for multiple comparisons. The significant group contrasts across T2–T4 demonstrated medium to large effect sizes (midcingulate cortex, $d = -0.72$; medial prefrontal cortex, $d = -0.83$).

Head Motion

We evaluated whether group differences existed in the amount of head motion, defined as the average framewise displacement (FD) at each measurement occasion; T1 median = 0.21 (range 0.08–1.04); T2 median = 0.20 (range = 0.09–0.7); T3 median = 0.20 (range = 0.07–0.77); T4 median = 0.22 (range 0.09–0.9). Mann–Whitney *U* tests were conducted given that FD values were not normally distributed. There were no significant group differences in FD at any of the time points (all p-values ≥ 0.34), suggesting that group differences in dDMN deactivation cannot be explained by group differences in head motion.

Neural Correlates of Task Improvement

The results presented above indicate that relative to the active control group, the OLM group demonstrated significantly more scanner task average hits at T4 as well as increased task-induced deactivation of the dDMN during encoding. We therefore evaluated if within-person change in dDMN deactivation levels during encoding (T4–T1) predicted scanner task average hits at T4. We calculated residuals scores for T4 average hits and for

TABLE 6 | Dorsal DMN nodes: Parameter estimates for group effect (across T2–T4) in the encoding condition.

Node	Estimate	*SE*	*t*	*p*	*d*
Medial PFC, OFC	**−0.16**	**0.05**	**−3.29**	**0.002**	**−0.83**
Left angular gyrus	−0.09	0.11	−0.84	0.402	−0.21
Right superior frontal gyrus[a]	−0.10	0.05	−2.06	0.043	−0.53
PCC, precuneus	−0.03	0.07	−0.43	0.673	−0.11
Midcingulate cortex	**−0.22**	**0.08**	**−2.82**	**0.006**	**−0.72**
Right angular gyrus	0.01	0.10	0.06	0.954	0.01
Left and right thalamus	0.00	0.07	0.02	0.986	0.00
Left hippocampus	−0.07	0.06	−1.24	0.219	−0.31
Right hippocampus	−0.13	0.06	−2.21	0.031	−0.56

[a]*Does not survive Bonferroni-Holm correction for multiple comparisons.*

OLM training group is contrasted against active control group. P-values are uncorrected. Cohen's d is based on parameter estimates from linear mixed effects regression models. Statistically significant effects surviving multiple comparisons correction appear in bold. PFC, prefrontal cortex; OFC, orbitofrontal cortex; PCC, posterior cingulate cortex.

the within-person change in deactivation by regressing out the effects of age, sex, T1 average hits, and T1 deactivation values. Next, we examined the correlation between the residualized values for deactivation change and T4 average hits for each group separately. Correlation coefficients were not significant for either group (OLM: $r = -0.27$, $p = 0.11$; active control: $r = 0.13$, $p = 0.51$). Two outliers were detected for the residualized average hits variable within the OLM group; the correlation remained non-significant after removal of these two outliers ($r = -0.15$, $p = 0.40$).

DISCUSSION

Findings highlight the functional heterogeneity of the DMN both in terms of task-based activation/deactivation during an object-location memory task and its response to cognitive training in healthy older individuals. Specifically, the ventral DMN was activated during the encoding and recognition conditions of the scanner task and the dorsal DMN was deactivated. Further, we report the novel finding that task-induced deactivation within the dorsal DMN was enhanced by process-based cognitive training compared to an active control training condition. Although the functional implications of this finding are not entirely clear, we discuss potential interpretations in terms of current theories of DMN functioning below.

Task-Related Activation/Deactivation of the Ventral and Dorsal DMN

Activation of the vDMN observed in the baseline (T1) analysis is consistent with the proposed role of a medial temporal lobe DMN subsystems in episodic memory. The vDMN as defined by Shirer et al. (2012) connects known memory regions such as the retrosplenial cortex and medial temporal lobe and demonstrates increased functional connectivity when subjects freely recall events of their day. The medial temporal lobe DMN system is thought to especially relate to associative aspects of memory, such as the retrieval of additional contextual details related to how the item was initially encountered (Andrews-Hanna et al., 2014). Associative memory was an important feature in our in-scanner OLM task paradigm, because subjects needed to retrieve the spatial location of the stimulus from the encoding condition in order to correctly answer the recognition probe. In the longitudinal model, the number of average hits from the scanner task was positively associated with vDMN activation during the recognition condition, and the baseline analysis indicated that vDMN activation was greater during recognition than encoding. These findings support the role of the vDMN particularly during the retrieval of object location associations.

The dorsal DMN, on the other hand, demonstrated task-induced deactivation. As defined in Shirer et al. (2012) on the basis of functional connectivity during subject-driven cognitive states, the dDMN consists of areas including the medial prefrontal cortex, posterior cingulate cortex, precuneus, and angular gyrus. These regions correspond to the classical DMN regions long observed to exhibit task-induced deactivation (Shulman et al., 1997; Raichle et al., 2001). In our baseline analysis, deactivation was greater during the encoding than the recognition condition. Further, the training effect of increased TID within the dorsal DMN was apparent in the encoding condition, and we discuss the potential significance of TID during encoding below.

Training-Related Increase in Dorsal DMN Deactivation

Object-location memory training was associated with significantly more dDMN deactivation during encoding relative to active control training. Analyses controlled for baseline (T1) deactivation levels, age, sex, and average hits, and thus cannot be explained by differences in these factors. Levels of head motion did not differ between the two groups at any time point, and therefore also do not seem to explain differences in deactivation.

In the following section, we interpret this task-related deactivation in the context of the results of the behavioral task that was performed concurrently in the scanner. The scanner OLM task differed from the training tasks performed by the OLM group, and it is considered to represent near-transfer within the domain of spatial episodic memory. This has implications

for interpretation of the behavioral results. For instance, it may be somewhat surprising that the active control group improved to a similar degree as the OLM group from mid-training (T2) to T3 (end of training) such that performance for both groups was equivalent at T3. However, since the scanner task differed from the OLM training task, it may be a less sensitive behavioral outcome measure than the training task itself.

While both groups demonstrated improved average hits performance at T3 relative to T2, only the OLM group showed the neural effect of increased dDMN deactivation at T3. Thus, it appears that the behavioral performance of the active control group improved at T3 but without the underlying neural change that was shown by the OLM group. This suggests that the improvement in each group relative to the previous time point may have involved different underlying mechanisms. Improvement in the active control group could have been due to practice effects and/or the mental engagement afforded by the active control perceptual training tasks without any underlying change in dDMN deactivation. The behavioral effect observed in the OLM group at T3 might reflect an increase in neural capacity, as evidenced by the concurrent increase in deactivation. Further, the OLM group maintained their improved average hits performance at T4, while the active control group did not. It is possible that the underlying neural response, evident at T3 as well as T4 explains the maintenance of the behavioral improvement in the OLM group.

Still, we did not find direct evidence for a relationship between increased dDMN suppression and improved task performance in the OLM group at T4. Lack of a significant brain-behavior relationship does not necessarily prove a lack of functional significance and could be related to several factors, including a lack of power. Further, TID may represent a task-independent phenomenon that does not directly correlate with improvement on this particular task. Indeed, a notable feature of TID is the consistency of deactivated regions across a wide range of cognitive tasks (Shulman et al., 1997; Gusnard and Raichle, 2001). Although some studies have noted a relationship between the degree of deactivation and task performance as measured by reaction time or accuracy (Persson et al., 2007; Miller et al., 2008; Park et al., 2010; Brown et al., 2018), DMN suppression may not directly relate to object-location memory performance *per se* but may instead represent a more general, task-independent phenomenon.

Relationship to Theories of DMN and TID

The mechanisms underlying TID are not fully understood, but one proposed mechanism reflects a reallocation of limited brain processing resources when attention shifts from ongoing, internal, conceptual processes to performance of an exogenous task (Binder et al., 1999; McKiernan et al., 2003). Consistent with this model, greater TID magnitude during an external task is associated with a lower frequency of task-unrelated thoughts (McKiernan et al., 2006), while reduced TID is related to attentional lapses and errors (Weissman et al., 2006; Li et al., 2007; Eichele et al., 2008). Additionally, activation of the DMN has been associated with mind wandering during a task that was highly practiced in order to elicit stimulus-independent thoughts (Mason et al., 2007) and with mind wandering during meditation (Hasenkamp et al., 2012). Finally, experience sampling during fMRI acquisition revealed that DMN activations preceded off-task thoughts and errors (Christoff et al., 2009). Drawing from this model, we interpret our finding of increased TID in the OLM group as reflecting a greater ability to reallocate processing resources from default mode areas. Behaviorally, this enhanced DMN deactivation may reflect a greater ability to suppress default cognitive processes, such as mind wandering, task-unrelated thoughts, and self-referential processing, in order to focus on the external task.

Although the role of DMN activation in internally focused mentation has been emphasized here, it is important to note that default mode activation has also been posited to reflect monitoring of the external environment (Gusnard and Raichle, 2001; Raichle et al., 2001). In the context of task fMRI, this could include waiting for upcoming task-relevant stimuli or attending to scanner noise and incidental light. Previous research indicates that older adults are vulnerable to distraction due to an inability to suppress processing of irrelevant environmental stimuli, including those related to the scanner environment (Stevens et al., 2008). Thus, the task-induced deactivation that we observed could also be related to suppression of external distractors which, as with suppression of internal distraction, would be expected to benefit their performance on an externally focused task.

The training-related increase in TID was specific to the encoding condition, underscoring the idea that the encoding phase may be more sensitive to the application of strategies and improvement. Supporting the importance of TID during encoding, Anticevic et al. (2010) reported greater DMN suppression for correct versus incorrect trials during encoding but not during later stages of a memory task (i.e., distracter and recognition probe phases). In another study, young adults showed greater DMN deactivation during encoding for scenes that were later remembered compared to those that were later forgotten (Chai et al., 2014). This suggests that DMN deactivation, and related suppression of distracting default mode processes, is especially important during initial formation of the memory traces.

Post hoc analyses revealed that the pattern of greater deactivation for the OLM group was apparent within the medial prefrontal cortex (mPFC), demonstrating a large effect size in comparison to the active control group across T2–T4. The mPFC exhibits particularly widespread connectivity within the DMN and may represent a network hub (Andrews-Hanna et al., 2014, 2010b). MPFC activity is linked to self-initiated stimulus-independent thought and emotional processing (McGuire et al., 1996; Binder et al., 1999; Gusnard et al., 2001). The mPFC may also play a unique role in suppression of task-irrelevant information. For instance, in young adults, mPFC deactivation during tests of working memory and attention was related to faster response times (Chadick and Gazzaley, 2011; Dang et al., 2013). Older adults exhibited less TID within the mPFC node of the DMN compared to young adults (Andrews-Hanna et al., 2007; Chadick et al., 2014), and among older adults

reduced mPFC suppression was related to a greater impact of distraction on a working memory task (Chadick et al., 2014). In this context, our finding of increased training-related mPFC deactivation suggests that it is possible to ameliorate age-related decline in mPFC deactivation. Modulation of mPFC activity through learning was further demonstrated in a study of healthy older adults who showed a significant mPFC deactivation in response to the explicit instruction to apply a semantic encoding strategy; in this study a greater increase in deactivation was associated with greater strategic behavior (Balardin et al., 2015).

We observed a significant effect of sex in the longitudinal analysis, whereby men showed more dDMN deactivation during encoding than women across both the OLM and active control groups. There was no interaction with training group and importantly, there were no baseline (T1) differences in dDMN deactivation by sex. Sex differences in the DMN, particularly those regarding TID, have not been extensively studied. Some reports have described a higher degree of functional connectivity in women between the posterior cingulate/precuneus and prefrontal cortex (Bluhm et al., 2008; Tomasi and Volkow, 2012), while another found no sex differences in DMN connectivity (Weissman-Fogel et al., 2010). It is difficult to interpret our incidental finding that men exhibited greater TID from T2 to T4 across both training groups, but it may suggest that future training studies should more systematically examine sex differences in TID over time.

One final point about the interpretation of reduced TID in the elderly deserves mention here. The prevailing view in the literature is that reduced TID in older adults reflects a reduced ability to suppress default mode processes and reallocate resources toward the task at hand. However, potentially contradictory results arise from studies that separate deactivation related to successfully encoded items from that related to forgotten items (i.e., subsequent memory paradigm). Specifically, some studies have shown that older adults demonstrate less deactivation for remembered versus forgotten items, and in older adults successful encoding can be associated with less deactivation (Maillet and Schacter, 2016). These patterns, which differ from those observed in younger adults, raise the possibility that older adults are more reliant on the default mode than young adults when performing attention-demanding tasks (Maillet and Schacter, 2016). That is, greater engagement (i.e., reduced suppression) of the default network in older adults might reflect increased reliance on cognitive processes mediated by the DMN, such as drawing on prior knowledge, experience, and schemas accumulated over their longer lifespan (Turner and Spreng, 2015). If this interpretation is correct, DMN suppression could potentially hamper older adult performance, and it would call into question our interpretation that the training-related increase in TID is beneficial for older adults. Clearly, more research is needed to clarify the mechanisms underlying reduced TID in older adults and its functional implications, preferably involving the subsequent memory paradigm to address the discrepant patterns observed in old versus young adults.

Limitations

Strengths of this study include the use of an active control training condition and multiple scanning sessions, but several limitations must be noted. First, it is unclear whether these results would generalize to the population of older adults, considering that our sample was relatively young (60–75 years), did not suffer from neurological or mental disorders, was highly educated, and had slightly above-average cognitive abilities as described in Zimmermann et al. (2016). Further, both groups performed the scanner task with approximately 79% accuracy at baseline, raising concerns about ceiling effects during the three subsequent sessions. There is also evidence that older adults demonstrate reduced TID relative to young adults primarily at higher levels of task demand (Persson et al., 2007; Park et al., 2010; Turner and Spreng, 2015). We cannot be sure that our task was challenging enough in this sense, but inclusion of a young adult comparison group and multiple task difficulty levels were not possible within the context of the current study.

Additionally, the atlas of Shirer et al. (2012), which we used to define the dorsal and ventral DMN subnetworks, is based on a sample of young adults. Grady et al. (2010) described a shrinkage of the extent of the DMN in terms of task-related brain activation patterns in older compared to younger adults, suggesting that a network definition based on a sample of young adults may not entirely apply to older adults. Mitigating this concern, data-driven parcellations of resting state data across different age cohorts reveal significant spatial overlap in large-scale brain systems across the adult lifespan (Han et al., 2018). In this context, we believe that the benefits of using the present atlas (i.e., standardization and comparability across studies) outweigh the concern about the young adult reference sample, but this caveat should still be considered when interpreting the results.

The block design employed in this study did not allow us to separate (de)activation related to successfully encoded items from that related to forgotten items. Known as the subsequent memory paradigm, this method is useful for elucidating neural processes related to successful learning (Daselaar et al., 2004; Maillet and Schacter, 2016). Further, the fact that our analysis calculated neural (de)activation levels across both successfully and unsuccessfully encoded items represents another possible explanation for the lack of an observed relationship with task improvement.

Finally, it is important to note that our ROI-based approach necessarily focuses on the DMN, and we cannot draw conclusions about the response to cognitive training in other regions or networks. We did not apply exploratory voxel-wise analysis because of considerations about statistical power and because of our *a priori* focus on the DMN. Nevertheless, we acknowledge that it could be useful for more exploratory types of analyses to apply the types of LMER models that we conducted here across all voxels of the brain (Madhyastha et al., 2018).

Future Directions

Several directions for future research can be considered. First, the DMN does not function in isolation, but is thought to interact with other large-scale networks. Therefore, it might also be

important to study training effects in the task-positive network that is preferentially active when individuals are engaged in attention-demanding tasks focused on the external environment (Fox et al., 2005). The task positive network may include a dorsal attention network as well as a frontoparietal control and a salience network (Seeley et al., 2007; Spreng, 2012). Some have argued that reduced DMN deactivation in the elderly is not reflective of DMN dysfunction *per se* but instead reflects a lower degree of flexible network interactivity, including greater coupling between DMN and frontoparietal/executive regions as task demands increase (Spreng and Schacter, 2012; Turner and Spreng, 2015). The salience network would also be of interest, considering that its disruption in aging is related to cognitive decline (Onoda et al., 2012) and that it exhibits increased connectivity to the DMN with increased age (Malagurski et al., 2020). Nevertheless, we found it necessary to first evaluate heterogeneity within the DMN before addressing between-network interactions. Future cognitive training studies may wish to examine dynamics across a wider range of networks, while also including multiple DMN subnetworks.

Another direction for future research involves examination of structural and functional connectivity as it relates to cognitive performance and training outcome. There is evidence that disruptions in white matter integrity may underlie aging-related decreases in DMN functional connectivity (Andrews-Hanna et al., 2007) as well as aging-related decreases in TID (Brown et al., 2018). Thus, the question arises as to whether training-related increases in TID observed in this study were related to changes in structural or functional connectivity.

Conclusion

Our findings support the heterogeneity of the DMN, with the ventral DMN being activated during a memory task and the dorsal DMN being deactivated. Further, we report the novel finding that task-induced deactivation within the dorsal DMN was enhanced by process-based cognitive training compared to an active control training condition. Given reports of reduced DMN suppression in the elderly and negative functional consequences thereof, this finding may suggest a promising mechanism through which process-based cognitive training can enhance cognitive performance in the elderly. However, this conclusion is tempered by our incomplete understanding of the mechanism underlying TID reductions in older adults, as well as by the lack of an explicit association between increased TID and task improvement in the current study. Further research, including an examination of interactions with other networks and of associated structural and functional connectivity changes, could help to elucidate relevant mechanisms underlying training-related increases in dorsal DMN deactivation.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The R analysis code and the associated data tables will be made available by the authors upon request. The raw fMRI image data underlying this article are not publicly available. Requests to access these datasets should be directed to data@dynage.uzh.ch.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Canton of Zurich. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AM: conceptualization, visualization, formal analysis, and writing – original draft preparation. BM: conceptualization, software, and writing – reviewing and editing. FL: software, data curation, and writing – reviewing and editing. SM: supervision, project administration, and writing – reviewing and editing. LJ: resources, funding acquisition, and writing – reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the Swiss National Science Foundation (1000014_146542), the Suzanne and Hans Biäsch Foundation for Applied Psychology (2012/7), the Velux Stiftung (369), and the University Research Priority Program "Dynamics of Healthy Aging" of the University of Zurich.

ACKNOWLEDGMENTS

The following researchers at the University of Zurich were involved in the design, set-up, and maintenance of the Object-Location Memory training study: Anne Eschen, Kathrin Zimmermann, LJ, and Mike Martin. We would like to thank Michael Notter for his work on earlier versions of computer code for fMRI data analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.623766/full#supplementary-material

REFERENCES

Andrews-Hanna, J. R. (2012). The brain's default network and its adaptive role in internal mentation. *Neuroscientist* 18, 251–270. doi: 10.1177/1073858411403316

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., and Buckner, R. L. (2010a). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322–335. doi: 10.1152/jn.00830.2009

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., and Buckner, R. L. (2010b). Functional-anatomic fractionation of the brain's default network. *Neuron* 65, 550–562. doi: 10.1016/j.neuron.2010.02.005

Andrews-Hanna, J. R., Smallwood, J., and Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann. N. Y. Acad. Sci.* 1316, 29–52. doi: 10.1111/nyas.12360

Andrews-Hanna, J. R., Snyder, A. Z., Vincent, J. L., Lustig, C., Head, D., Raichle, M. E., et al. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron* 56, 924–935. doi: 10.1016/j.neuron.2007.10.038

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X.-J., and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* 16, 584–592. doi: 10.1016/j.tics.2012.10.008

Anticevic, A., Repovs, G., Shulman, G. L., and Barch, D. M. (2010). When less is more: TPJ and default network deactivation during encoding predicts working memory performance. *NeuroImage* 49, 2638–2648. doi: 10.1016/j.neuroimage.2009.11.008

Balardin, J. B., Batistuzzo, M. C., Martin, M., Sato, J. R., Smid, J., Porto, C., et al. (2015). Differences in prefrontal cortex activation and deactivation during strategic episodic verbal memory encoding in mild cognitive impairment. *Front. Aging Neurosci.* 7:147. doi: 10.3389/fnagi.2015.00147

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Behzadi, Y., Restom, K., Liau, J., and Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* 37, 90–101. doi: 10.1016/j.neuroimage.2007.04.042

Ben-Shachar, M. S., Makowski, D., and Lüdecke, D. (2020). *Compute and Interpret Indices of Effect Size. CRAN*.

Berres, M., Monsch, A., Bernasconi, F., Thalmann, B., and Staehelin, H. (2000). Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. *Stud. Health Technol. Inform.* 77, 195–199. doi: 10.3233/978-1-60750-921-9-195

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Rao, S. M., and Cox, R. W. (1999). Conceptual processing during the conscious resting state: a functional MRI study. *J. Cogn. Neurosci.* 11, 80–93. doi: 10.1162/089892999563265

Bluhm, R. L., Osuch, E. A., Lanius, R. A., Boksman, K., Neufeld, R. W. J., Théberge, J., et al. (2008). Default mode network connectivity: effects of age, sex, and analytic approach. *NeuroReport* 19, 887–891. doi: 10.1097/WNR.0b013e328300ebbf

Boyke, J., Driemeyer, J., Gaser, C., Büchel, C., and May, A. (2008). Training-induced brain structure changes in the elderly. *J. Neurosci. Off. J. Soc. Neurosci.* 28, 7031–7035. doi: 10.1523/JNEUROSCI.0742-08.2008

Brehmer, Y., Rieckmann, A., Bellander, M., Westerberg, H., Fischer, H., and Bäckman, L. (2011). Neural correlates of training-related working-memory gains in old age. *NeuroImage* 58, 1110–1120. doi: 10.1016/j.neuroimage.2011.06.079

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., and Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One* 5:e10773. doi: 10.1371/journal.pone.0010773

Brown, C. A., Jiang, Y., Smith, C. D., and Gold, B. T. (2018). Age and Alzheimer's pathology disrupt default mode network functioning via alterations in white matter microstructure but not hyperintensities. *Cortex* 104, 58–74. doi: 10.1016/j.cortex.2018.04.006

Buckner, R. L., and DiNicola, L. M. (2019). The brain's default network: updated anatomy, physiology and evolving insights. *Nat. Rev. Neurosci.* 20, 593–608. doi: 10.1038/s41583-019-0212-7

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Cao, W., Cao, X., Hou, C., Li, T., Cheng, Y., Jiang, L., et al. (2016). Effects of cognitive training on resting-state functional connectivity of default mode, salience, and central executive networks. *Front. Aging Neurosci.* 8:70. doi: 10.3389/fnagi.2016.00070

Chadick, J. Z., and Gazzaley, A. (2011). Differential coupling of visual cortex with default network or frontal-parietal network based on goals. *Nat. Neurosci.* 14, 830–832. doi: 10.1038/nn.2823

Chadick, J. Z., Zanto, T. P., and Gazzaley, A. (2014). Structural and functional differences in medial prefrontal cortex underlie distractibility and suppression deficits in ageing. *Nat. Commun.* 5:4223. doi: 10.1038/ncomms5223

Chai, X. J., Ofen, N., Gabrieli, J. D. E., and Whitfield-Gabrieli, S. (2014). Development of deactivation of the default-mode network during episodic memory formation. *NeuroImage* 84, 932–938. doi: 10.1016/j.neuroimage.2013.09.032

Chen, J. E., Glover, G. H., Greicius, M. D., and Chang, C. (2017). Dissociated patterns of anti-correlations with dorsal and ventral default-mode networks at rest. *Hum. Brain Mapp.* 38, 2454–2465. doi: 10.1002/hbm.23532

Chiu, H.-L., Chu, H., Tsai, J.-C., Liu, D., Chen, Y.-R., Yang, H.-L., et al. (2017). The effect of cognitive-based training for the healthy older people: a meta-analysis of randomized controlled trials. *PLoS One* 12:e0176742. doi: 10.1371/journal.pone.0176742

Christoff, K., Gordon, A. M., Smallwood, J., Smith, R., and Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8719–8724. doi: 10.1073/pnas.0900234106

Damoiseaux, J. S., Beckmann, C. F., Arigita, E. J. S., Barkhof, F., Scheltens, P., Stam, C. J., et al. (2008). Reduced resting-state brain activity in the "default network" in normal aging. *Cereb. Cortex* 18, 1856–1864. doi: 10.1093/cercor/bhm207

Dang, L. C., O'Neil, J. P., and Jagust, W. J. (2013). Genetic effects on behavior are mediated by neurotransmitters and large-scale neural networks. *NeuroImage* 66, 203–214. doi: 10.1016/j.neuroimage.2012.10.090

Daselaar, S. M., Prince, S. E., and Cabeza, R. (2004). When less means more: deactivations during encoding that predict subsequent memory. *NeuroImage* 23, 921–927. doi: 10.1016/j.neuroimage.2004.07.031

De Marco, M., Meneghello, F., Duzzi, D., Rigon, J., Pilosio, C., and Venneri, A. (2016). Cognitive stimulation of the default-mode network modulates functional connectivity in healthy aging. *Brain Res. Bull.* 121, 26–41. doi: 10.1016/j.brainresbull.2015.12.001

Delacre, M., and Klein, O. (2019). *Routliers: Robust Outliers Detection*.

Duda, B. M., and Sweet, L. H. (2019). Functional brain changes associated with cognitive training in healthy older adults: a preliminary ALE meta-analysis. *Brain Imaging Behav.* 14, 1247–1262. doi: 10.1007/s11682-019-00080-0

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., et al. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 105, 6173–6178. doi: 10.1073/pnas.0708965105

Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø, Larsen, V. A., et al. (2012). Memory training impacts short-term changes in aging white matter: a longitudinal diffusion tensor imaging study. *Hum. Brain Mapp.* 33, 2390–2406. doi: 10.1002/hbm.21370

Engvig, A., Fjell, A. M., Westlye, L. T., Moberget, T., Sundseth, Ø, Larsen, V. A., et al. (2010). Effects of memory training on cortical thickness in the elderly. *NeuroImage* 52, 1667–1676. doi: 10.1016/j.neuroimage.2010.05.041

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2018). FMRIPrep: a robust preprocessing pipeline for functional MRI. *bioRxiv* [Preprint]. doi: 10.1101/306951 bioRxiv: 306951.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Goh, J. O., and Park, D. C. (2009). Neuroplasticity and cognitive aging: the scaffolding theory of aging and cognition. *Restor. Neurol. Neurosci.* 27, 391–403. doi: 10.3233/RNN-2009-0493

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3, 1–9. doi: 10.1038/sdata.2016.44

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: a flexible, lightweight and extensible neuroimaging

data processing framework in Python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Grady, C. L., Protzner, A. B., Kovacevic, N., Strother, S. C., Afshin-Pour, B., Wojtowicz, M., et al. (2010). A multivariate analysis of age-related differences in default mode and task-positive networks across multiple cognitive domains. *Cereb. Cortex* 20, 1432–1447. doi: 10.1093/cercor/bhp207

Grady, C. L., Springer, M. V., Hongwanishkul, D., McIntosh, A. R., and Winocur, G. (2006). Age-related changes in brain activity across the adult lifespan. *J. Cogn. Neurosci.* 18, 227–241. doi: 10.1162/jocn.2006.18.2.227

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc. Natl. Acad. Sci. U.S.A.* 100, 253–258. doi: 10.1073/pnas.0135058100

Greicius, M. D., Supekar, K., Menon, V., and Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* 19, 72–78. doi: 10.1093/cercor/bhn059

Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* 48, 63–72. doi: 10.1016/j.neuroimage.2009.06.060

Gusnard, D. A., Akbudak, E., Shulman, G. L., and Raichle, M. E. (2001). Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4259–4264. doi: 10.1073/pnas.071043098

Gusnard, D. A., and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nat. Rev. Neurosci.* 2, 685–694. doi: 10.1038/35094500

Han, L., Savalia, N. K., Chan, M. Y., Agres, P. F., Nair, A. S., and Wig, G. S. (2018). Functional parcellation of the cerebral cortex across the human adult lifespan. *Cereb. Cortex* 28, 4403–4423. doi: 10.1093/cercor/bhy218

Hansen, N. L., Lauritzen, M., Mortensen, E. L., Osler, M., Avlund, K., Fagerlund, B., et al. (2014). Subclinical cognitive decline in middle-age is associated with reduced task-induced deactivation of the brain's default mode network. *Hum. Brain Mapp.* 35, 4488–4498. doi: 10.1002/hbm.22489

Hasenkamp, W., Wilson-Mendenhall, C. D., Duncan, E., and Barsalou, L. W. (2012). Mind wandering and attention during focused meditation: a fine-grained temporal analysis of fluctuating cognitive states. *NeuroImage* 59, 750–760. doi: 10.1016/j.neuroimage.2011.07.008

Heinzel, S., Lorenz, R. C., Brockhaus, W.-R., Wüstenberg, T., Kathmann, N., Heinz, A., et al. (2014). Working memory load-dependent brain response predicts behavioral training gains in older adults. *J. Neurosci.* 34, 1224–1233. doi: 10.1523/JNEUROSCI.2463-13.2014

Heinzel, S., Lorenz, R. C., Pelz, P., Heinz, A., Walter, H., Kathmann, N., et al. (2016). Neural correlates of training and transfer effects in working memory in older adults. *NeuroImage* 134, 236–249. doi: 10.1016/j.neuroimage.2016.03.068

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? *Psychol. Sci. Public Interest J. Am. Psychol. Soc.* 9, 1–65. doi: 10.1111/j.1539-6053.2009.01034.x

Karbach, J., and Verhaeghen, P. (2014). Making working memory work: a meta-analysis of executive control and working memory training in younger and older adults. *Psychol. Sci.* 25, 2027–2037. doi: 10.1177/0956797614548725

Kessels, R. P. C., Hobbel, D., and Postma, A. (2007). Aging, context memory and binding: a comparison of "what, where and when" in young and older adults. *Int. J. Neurosci.* 117, 795–810. doi: 10.1080/00207450600910218

Koch, W., Teipel, S., Mueller, S., Buerger, K., Bokde, A. L. W., Hampel, H., et al. (2010). Effects of aging on default mode network activity in resting state fMRI: does the method of analysis matter? *NeuroImage* 51, 280–287. doi: 10.1016/j.neuroimage.2009.12.008

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lebedev, A. V., Nilsson, J., and Lövdén, M. (2018). Working memory and reasoning benefit from different modes of large-scale brain dynamics in healthy older adults. *J. Cogn. Neurosci.* 30, 1033–1046. doi: 10.1162/jocn_a_01260

Lenth, R. V. (2020). *emmeans: Estimated Marginal Means, aka LeastSquares Means.*

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013

Li, C. R., Yan, P., Bergquist, K. L., and Sinha, R. (2007). Greater activation of the "default" brain regions predicts stop signal errors. *NeuroImage* 38, 640–648. doi: 10.1016/j.neuroimage.2007.07.021

Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., and Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychol. Bull.* 136, 659–676. doi: 10.1037/a0020080

Lövdén, M., Schaefer, S., Noack, H., Bodammer, N. C., Kühn, S., Heinze, H.-J., et al. (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. *Neurobiol. Aging* 33, 620.e9–620.e22. doi: 10.1016/j.neurobiolaging.2011.02.013

Lustig, C., Snyder, A. Z., Bhakta, M., O'Brien, K. C., McAvoy, M., Raichle, M. E., et al. (2003). Functional deactivations: change with age and dementia of the Alzheimer type. *Proc. Natl. Acad. Sci. U.S.A.* 100, 14504–14509. doi: 10.1073/pnas.2235925100

Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., et al. (2018). Current methods and limitations for longitudinal fMRI analysis across development. *Dev. Cogn. Neurosci.* 33, 118–128. doi: 10.1016/j.dcn.2017.11.006

Maillet, D., and Schacter, D. L. (2016). Default network and aging: beyond the task-negative perspective. *Trends Cogn. Sci.* 20, 646–648. doi: 10.1016/j.tics.2016.05.009

Malagurski, B., Liem, F., Oschwald, J., Mérillat, S., and Jäncke, L. (2020). Longitudinal functional brain network reconfiguration in healthy aging. *Hum. Brain Mapp.* 41, 4829–4845. doi: 10.1002/hbm.25161

Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., and Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science* 315, 393–395. doi: 10.1126/science.1131295

Mayer, J. S., Roebroeck, A., Maurer, K., and Linden, D. E. J. (2010). Specialization in the default mode: task-induced brain deactivations dissociate between visual working memory and attention. *Hum. Brain Mapp.* 31, 126–139. doi: 10.1002/hbm.20850

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., et al. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.* 54, 287–298. doi: 10.1016/S0361-9230(00)00437-8

McGuire, P. K., Paulesu, E., Frackowiak, R. S., and Frith, C. D. (1996). Brain activity during stimulus independent thought. *Neuroreport* 7, 2095–2099.

McKiernan, K. A., D'Angelo, B. R., Kaufman, J. N., and Binder, J. R. (2006). Interrupting the "stream of consciousness": an fMRI investigation. *NeuroImage* 29, 1185–1191. doi: 10.1016/j.neuroimage.2005.09.030

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., and Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J. Cogn. Neurosci.* 15, 394–408. doi: 10.1162/089892903321593117

Miller, S. L., Celone, K., DePeau, K., Diamond, E., Dickerson, B. C., Rentz, D., et al. (2008). Age-related memory impairment associated with loss of parietal deactivation but preserved hippocampal activation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2181–2186. doi: 10.1073/pnas.0706818105

Old, S. R., and Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: a meta-analysis. *Psychol. Aging* 23, 104–118. doi: 10.1037/0882-7974.23.1.104

Onoda, K., Ishihara, M., and Yamaguchi, S. (2012). Decreased functional connectivity by aging is associated with cognitive decline. *J. Cogn. Neurosci.* 24, 2186–2198. doi: 10.1162/jocn_a_00269

Park, D. C., and Bischof, G. N. (2013). The aging mind: neuroplasticity in response to cognitive training. *Dialogues Clin. Neurosci.* 15, 109–119. doi: 10.31887/dcns.2013.15.1/dpark

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., and Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychol. Aging* 17, 299–320. doi: 10.1037/0882-7974.17.2.299

Park, D. C., Polk, T. A., Hebrank, A. C., and Jenkins, L. J. (2010). Age differences in default mode activity on easy and difficult spatial judgment tasks. *Front. Hum. Neurosci.* 3:75. doi: 10.3389/neuro.09.075.2009

Persson, J., Lustig, C., Nelson, J. K., and Reuter-Lorenz, P. A. (2007). Age differences in deactivation: a link to cognitive control? *J. Cogn. Neurosci.* 19, 1021–1032. doi: 10.1162/jocn.2007.19.6.1021

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048

R Core Team. (2019). *R: A language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Raichle, M. E., and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *NeuroImage* 37, 1083–1090. doi: 10.1016/j.neuroimage.2007.02.041

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682.

Rossion, B., and Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33, 217–236. doi: 10.1068/p5117

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiol. Aging* 30, 507–514. doi: 10.1016/j.neurobiolaging.2008.09.023

Sambataro, F., Murty, V. P., Callicott, J. H., Tan, H.-Y., Das, S., Weinberger, D. R., et al. (2010). Age-related alterations in default mode network: impact on working memory performance. *Neurobiol. Aging* 31, 839–852. doi: 10.1016/j.neurobiolaging.2008.05.022

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007

Sheikh, J. I., and Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. *Clin. Gerontol. J. Aging Ment. Health* 5, 165–173. doi: 10.1300/J018v05n01_09

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb. Cortex* 22, 158–165. doi: 10.1093/cercor/bhr099

Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., et al. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *J. Cogn. Neurosci.* 9, 648–663. doi: 10.1162/jocn.1997.9.5.648

Snodgrass, J. G., and Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn.* 6, 174–215. doi: 10.1037/0278-7393.6.2.174

Spreng, R. N. (2012). The fallacy of a "task-negative" network. *Front. Psychol.* 3:145. doi: 10.3389/fpsyg.2012.00145

Spreng, R. N., and Schacter, D. L. (2012). Default network modulation and large-scale network interactivity in healthy young and old adults. *Cereb. Cortex* 22, 2610–2621. doi: 10.1093/cercor/bhr339

Staffaroni, A. M., Brown, J. A., Casaletto, K. B., Elahi, F. M., Deng, J., Neuhaus, J., et al. (2018). The longitudinal trajectory of default mode network connectivity in healthy older adults varies as a function of age and is associated with changes in episodic memory and processing speed. *J. Neurosci.* 38, 2809–2817. doi: 10.1523/JNEUROSCI.3067-17.2018

Stevens, W. D., Hasher, L., Chiew, K. S., and Grady, C. (2008). A neural mechanism underlying memory failure in older adults. *J. Neurosci.* 28, 12820–12824. doi: 10.1523/JNEUROSCI.2622-08.2008

Tomasi, D., and Volkow, N. D. (2012). Gender differences in brain functional connectivity density. *Hum. Brain Mapp.* 33, 849–860. doi: 10.1002/hbm.21252

Turner, G. R., and Spreng, R. N. (2015). Prefrontal engagement and reduced default network suppression co-occur and are dynamically coupled in older adults: the default–executive coupling hypothesis of aging. *J. Cogn. Neurosci.* 27, 2462–2476. doi: 10.1162/jocn_a_00869

van Balkom, T. D., van den Heuvel, O. A., Berendse, H. W., van der Werf, Y. D., and Vriend, C. (2020). The effects of cognitive training on brain network activity and connectivity in aging and neurodegenerative diseases: a systematic review. *Neuropsychol. Rev.* 30, 267–286. doi: 10.1007/s11065-020-09440-w

Vidal-Piñeiro, D., Valls-Pedret, C., Fernández-Cabello, S., Arenaza-Urquijo, E. M., Sala-Llonch, R., Solana, E., et al. (2014). Decreased default mode network connectivity correlates with age-associated structural and cognitive changes. *Front. Aging Neurosci.* 6:256. doi: 10.3389/fnagi.2014.00256

Weissman, D. H., Roberts, K. C., Visscher, K. M., and Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nat. Neurosci.* 9, 971–978. doi: 10.1038/nn1727

Weissman-Fogel, I., Moayedi, M., Taylor, K. S., Pope, G., and Davis, K. D. (2010). Cognitive and default-mode resting state networks: do male and female brains "rest" differently? *Hum. Brain Mapp.* 31, 1713–1726. doi: 10.1002/hbm.20968

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer.

Wig, G. S. (2017). Segregated systems of human brain networks. *Trends Cogn. Sci.* 21, 981–996. doi: 10.1016/j.tics.2017.09.006

Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57. doi: 10.1109/42.906424

Zimmermann, K. (2015). *Object-Location Memory in Old Age: Brain and Cognitive Plasticity Induced by a Process-Based Training (Dissertation)*. Switzerland: University of Zurich. doi: 10.5167/uzh-113230

Zimmermann, K., von Bastian, C. C., Röcke, C., Martin, M., and Eschen, A. (2016). Transfer after process-based object-location memory training in healthy older adults. *Psychol. Aging* 31, 798–814. doi: 10.1037/pag0000123

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 05 March 2021
doi: 10.3389/fnhum.2021.635996

Diminished Feedback Evaluation and Knowledge Updating Underlying Age-Related Differences in Choice Behavior During Feedback Learning

*Tineke de Haan[1], Berry van den Berg[1], Marty G. Woldorff[2], André Aleman[3] and Monicque M. Lorist[1,3]**

[1] Department of Experimental Psychology, University of Groningen, Groningen, Netherlands, [2] Center for Cognitive Neuroscience, Duke University, Durham, NC, United States, [3] Department of Biomedical Sciences of Cells and Systems, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

In our daily lives, we continuously evaluate feedback information, update our knowledge, and adapt our behavior in order to reach desired goals. This ability to learn from feedback information, however, declines with age. Previous research has indicated that certain higher-level learning processes, such as feedback evaluation, integration of feedback information, and updating of knowledge, seem to be affected by age, and recent studies have shown how the adaption of choice behavior following feedback can differ with age. The neural mechanisms underlying this age-related change in choice behavior during learning, however, remain unclear. The aim of this study is therefore to investigate the relation between learning-related neural processes and choice behavior during feedback learning in two age groups. Behavioral and fMRI data were collected, while a group of young (age 18–30) and older (age 60–75) adults performed a probabilistic learning task consisting of 10 blocks of 20 trials each. On each trial, the participants chose between a house and a face, after which they received visual feedback (loss vs. gain). In each block, either the house or the face image had a higher probability of yielding a reward (62.5 vs. 37.5%). Participants were instructed to try to maximize their gains. Our results showed that less successful learning in older adults, as indicated by a lower learning rate, corresponded with a higher tendency to switch to the other stimulus option, and with a reduced adaptation of this switch choice behavior following positive feedback. At the neural level, activation following positive and negative feedback was found to be less distinctive in the older adults, due to a smaller feedback-evaluation response to positive feedback in this group. Furthermore, whereas young adults displayed increased levels of knowledge updating prior to adapting choice behavior, we did not find this effect in older adults. Together, our results suggest that diminished learning performance with age corresponds with diminished evaluation of positive feedback and reduced knowledge updating related to changes in choice behavior, indicating how such differences in feedback processing at the trial level in older adults might lead to reduced learning performance across trials.

Keywords: aging, fMRI, probabilistic learning, feedback, choice behavior

OPEN ACCESS

Edited by:
Christina Röcke,
University of Zurich, Switzerland

Reviewed by:
Marta Zimmer,
Budapest University of Technology and Economics, Hungary
Lihui Wang,
Shanghai Jiao Tong University, China

***Correspondence:**
Monicque M. Lorist
m.m.lorist@rug.nl

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

Received: *30 November 2020*
Accepted: *02 February 2021*
Published: *05 March 2021*

Citation:
de Haan T, van den Berg B, Woldorff MG, Aleman A and Lorist MM (2021) Diminished Feedback Evaluation and Knowledge Updating Underlying Age-Related Differences in Choice Behavior During Feedback Learning. Front. Hum. Neurosci. 15:635996. doi: 10.3389/fnhum.2021.635996

INTRODUCTION

In order to successfully interact with our environment and reach desired goals, we continuously learn from the outcomes of our actions and choices. These outcomes (feedback information) are evaluated in the light of the goals we aim to achieve and are integrated with the outcomes from prior experiences. In addition, based on choice-outcome contingencies, knowledge will be updated, enabling future behavior to be adapted accordingly. Although learning is crucial at all ages, research has shown that the ability to learn from feedback and adjust behavior accordingly declines with age (Marschner et al., 2005; Mell et al., 2005; Eppinger et al., 2011; Hämmerer et al., 2011; Ferdinand, 2019). However, the neural mechanisms underlying age-related chances in choice behavior during learning remain unclear. The present study was aimed at elucidating the relation between learning-related neural processes and changes in behavior during probabilistic learning.

Processes that occur during learning, such as feedback evaluation, integration of feedback information, and updating of knowledge, seem to be affected by age [for a review see Ferdinand and Czernochowski (2018)]. For example, neural activity related to the difference between expected and actual feedback after an action or choice (reward prediction error) is less distinct in older compared to younger adults (Mata et al., 2011; Samanez-Larkin et al., 2014; Ferdinand and Czernochowski, 2018). However, these changes do not seem to be driven by reductions in the sensitivity to reward. For instance, studies using functional magnetic resonance imaging (fMRI) have indicated that the same set of brain areas appear to be activated following positive and negative feedback in young and older adults, and the fronto-striatal response to positive feedback remains intact for older individuals (e.g., Schott et al., 2007; Cox et al., 2008; Samanez-Larkin et al., 2014). In sum, age-related learning differences do not seem to hinge on reward sensitivity, but rather on processes during feedback evaluation and knowledge updating.

Medial prefrontal cortical brain areas have been shown to play a pivotal role in learning-related processes, and the medial prefrontal anterior cingulate cortices (ACC) are thought to be especially important, as they have been shown to be active during the evaluation and integration of feedback information (Alexander and Brown, 2011; Cohen et al., 2011; Samanez-Larkin et al., 2014; Kolling et al., 2016). Furthermore, ACC activity during learning has been shown to depend on age, with older adults displaying less distinct neural patterns in these brain regions compared to young adults when evaluating feedback (Eppinger et al., 2008; Hämmerer et al., 2011; Samanez-Larkin et al., 2014).

The fact that the neural processes involved in feedback evaluation change with age can have implications for the ability to learn from feedback. For instance, behaviorally, older adults have been shown to develop a preference for processing either positive or negative feedback, depending on task demands (Eppinger et al., 2011). In a task where feedback does not have to be used for learning, for example when stimulus-reward associations are known in advance, older adults seem to focus less on negative feedback compared to young adults, as illustrated by reduced neural activity in older adults when anticipating negative but not positive feedback. In contrast, if successful performance in the future depends on feedback processing (e.g., in probabilistic learning tasks), negative feedback has a stronger impact on learning-related choice behavior of older adults compared to positive feedback. Ferdinand (2019), for example, showed that reducing the information value of negative feedback hindered older adults' learning performance more than reducing the information value of positive feedback, supporting the view that older adults rely more the negative feedback to learn in a probabilistic learning task.

Following the evaluation of feedback information, that information then needs to be used to update knowledge so it can potentially be used to guide the adjustment of subsequent behavior. Processes of knowledge updating have been linked to neural activity in fronto-parietal brain areas (Borst and Anderson, 2013). In addition, EEG studies have found neural evidence for increased knowledge updating prior to behavioral adaptation in the form of switching to a different response on the next trial (Polich, 2007; Chase et al., 2011; San Martin et al., 2013; Correa et al., 2018). The impact of age on this switch-choice behavior has been only sparsely studied, but the available studies show that both young and older adults adapt their switch choice on the next trial based on the feedback information received (Frank and Kong, 2008; Hämmerer et al., 2011). In addition, older adults displayed a bias toward negative feedback in that they modified their behavior to a lesser extent following positive feedback, a bias not found in young adults, further supporting the view that learning from positive feedback is diminished with age in probabilistic learning tasks (Frank and Kong, 2008; Hämmerer et al., 2011). Importantly, although older adults have been shown to evaluate feedback differently and adapt their choice behavior differently following feedback, the neural underpinnings of these effects remain unknown.

In order to gain more insight into the neural basis of age-related differences in feedback evaluation and knowledge updating during feedback learning, a group of younger and older participants performed a probabilistic learning task during which functional MRI measures of their brain activity were collected. The probabilistic learning task, which was based on a previous paradigm developed by our group (van den Berg et al., 2019), entailed participants choosing between houses and faces on each trial, after which they received either positive or negative feedback that was later converted into money. To gain money, participants had to actively learn which out of two stimulus types was more likely to yield a positive feedback for each of the 10 blocks of 20 trials.

In our study, older adults were expected to perform worse on the learning task compared to younger adults, and this diminished performance was expected to be paralleled by decreased neural activity related to feedback evaluation following positive feedback. In addition, we expected effects related to knowledge updating, which we investigated by contrasting neural activity related to switch behavior on the next trial. Given that we expected diminished knowledge updating in older adults, we predicted a smaller difference in neural activity related to making a switch vs. not making a switch on the next

trial, in the older adult group compared to young adults. Furthermore, these differences in feedback evaluation and knowledge updating were expected to result in different choice behavior in older adults related to learning performance. If decreased learning performance in older adults is related to different feedback evaluation and knowledge updating, as we expected, the difference in choice behavior between the age groups would be most prominent following positive feedback. In contrast, if only knowledge updating is affected, we expected the difference in choice behavior to not depend on the valence of the feedback received.

MATERIALS AND METHODS

Participants

There were 29 young and 27 older adults in this study. Participants were recruited by means of local advertisements and advertisements on social media. All participants were right-handed, had normal or correct-to-normal vision, and did not report taking any psychoactive medications. The study was conducted according to protocols approved by the Medical Ethical Committee of the University Medical Center Groningen, and all participants gave prior written informed consent. Four participants did not perform the task according to task instructions (three younger and one older adult, all female) and were therefore excluded from the final analysis. In addition, two older participants (one female) were excluded due to technical problems. Accordingly, the data from 26 young [13 female; 18–27 years; mean age (SD) = 22.2 (2.7)] and 24 older [11 female; 60–73 years; mean age (SD) = 65.5 (4.0)] participants were included in the analysis. Participants received 16 euros in compensation for their time, plus reimbursement of travel expenses and a variable monetary reward depending on their task performance [mean (SD): 5.50euro (4.40)].

Experimental Tasks and Stimuli

Materials

MRI data were acquired on a Siemens 3T scanner at the University Medical Center Groningen. The probabilistic learning task was programmed using the Presentation® software package (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) and presented on an MRI-compatible IPS LCD monitor (BOLDscreen 24, Cambridge Research Systems, resolution 1920 × 1200) that was made visible to the participant through a mirror (viewing distance ∼100 cm). Responses were made using two buttons positioned above each other on an MRI-compatible response box. The stimulus base set of 20 male face images and 20 house images (135 × 180 pixels) was identical to the one used by our group in van den Berg et al. (2019).

Probabilistic Learning Task

Participants performed a probabilistic learning task (**Figure 1**) in which they were asked to maximize their gains by learning which of two stimulus types (houses or faces) was more likely to yield a gain. The task was very similar to the task used in the EEG study of van den Berg et al. (2019), with slight adaptations of the stimulus presentation and response times. The task consisted of 10 blocks of 20 trials each. In each 20-trial block, choosing either the face images or the house images was more likely to lead to a gain (62.5 vs. 37.5%). This stimulus type will be called the "block winner." Each trial started with a pair of images, one of a face and one of a house, with one image presented to the left and the other to the right of a central fixation (**Figure 1**). The stimuli for this image-pair choice stimulus were randomly drawn from the base set (see materials section) and randomly assigned to be presented to the left and right locations. After 1,200 ms, two arrows appeared just above and below the fixation, each pointing to one of the two stimuli on the left and the right. The mapping of the arrow locations (above and below) to the direction the arrows pointed was randomly assigned on each trial. Participants indicated their choice using the response box by selecting the arrow pointing to the image they thought was more likely to yield a gain in that block. The selected arrow was highlighted for 300 ms, which was followed by a jittered interval (500–5,500 ms) during which a blank screen with fixation was shown. Following this interval, the feedback for that trial appeared on the screen for 500 ms. Feedback could either be a loss of eight points or a gain of eight points, presented as "−8" or "+8" printed in an orange or blue square (100 × 100 pixels) for a duration of 500 ms. The combination of square color and gain/loss was counterbalanced over participants. If participants did not respond within 2,000 ms following the presentation of the image pair at the beginning of the trial, the text "No response" appeared on the screen, which was then followed by a loss of eight points. The feedback stimulus was followed by a jittered presentation (1,500 – 7,000 ms) of a blank screen with a fixation cross, before the next image-pair choice stimulus appeared. Participants were instructed to try to maximize their gains by learning which of the two images was more likely to lead to a gain on the trials of a block. They were also informed about the probabilities of winning after choosing the block-winner (62.5%) vs. winning after choosing the block-loser (37.5%). In addition, participants were informed before the experiment started that their total score would be converted into a monetary reward at the end of the study. After each block, a short summary was shown that indicated the points that had been earned in that block and the total number of points accumulated up to that point in the session.

Procedure

Prior to the experimental session, participants filled out the Apathy Evaluation Scale (Marin et al., 1991; Reijnders et al., 2010), as varying degrees of apathy can affect reward processing, especially in older adults. None of the participants scored above the cut-off for clinically relevant apathetic symptoms (cut-off = 37/38; Max. AES-score$_{young}$ = 22; Max. AES-score$_{old}$ = 25), and the average apathy scores did not differ between the two age groups [average AES-score$_{young}$ = 16.1 (SD = 3.5); average AES-score$_{old}$ = 16.0 (SD = 4.6); $t_{(41.1)} = 0.1$, $p = 0.92$]. Before entering the MRI scanner, participants received task instructions and practiced one block of the probabilistic learning task on a laptop. MRI acquisition started with a T1-weighted anatomical image, after which the participants performed a short classification task (which is not relevant for this paper), followed by two sessions of

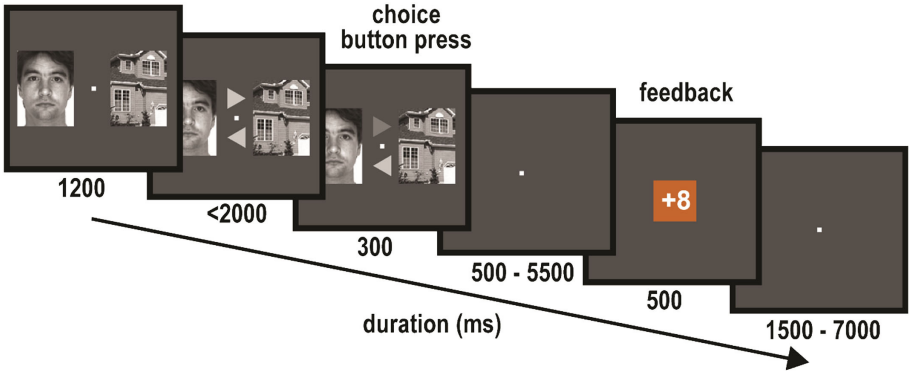


FIGURE 1 | Probabilistic learning task: on each trial an image pair of a house and a face were shown (left/right location randomized over trials), with the task to choose one or the other. After the arrow stimuli appeared, participants indicated their choice by selecting the appropriate arrow. Following a jittered interval, feedback as to whether they won or loss was presented. Each experimental block consisted of 20 of these trials, during which participants learned which of the two image types was more likely to lead to a gain in that block. The human image in this figure is one of the stimuli we used in our experiment, which is part of an image-dataset that is publicly available and free for academic use. The source of the image-dataset is: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html (Martinez and Benavente, 1998).

the probabilistic learning task, each consisting of five blocks. A short break was given between the two five-block sessions.

Behavioral Analysis

Behavioral analyses were performed using R (R Core Team, 2020). A mixed modeling approach was used to model two types of learning behavior (choosing the block winner and switching between stimulus categories) as a function of both trial position (1–20) and age group. All models included a random intercept per participant to account for individual variation in average learning performance and switch rate. A second-degree polynomial of trial number was added if it improved the model fit significantly to account for non-linear associations (which was the case for the learning rate model). A random slope for trial number per participant was found to improve the model fit of all models significantly and was therefore added to account for individual differences in learning rate and switch behavior. In addition, we also modeled switch behavior based on feedback history (effect of feedback on trial_{n-1} and trial_{n-2} on switch choice behavior on trial_n), to assess the impact of feedback integration effects on switch behavior. Again, a random intercept per participant was added to account for individual variation in average feedback integration effects, and random slopes for feedback (trial_{n-1} and trial_{n-2}) per participant were considered to account for individual differences in the impact of each feedback position. Lastly, in an exploratory analysis, switch behavior was modeled based on a combination of feedback (trial_{n-1}), trial position in the block (1–20), age group, and learning rate, to examine whether the impact of feedback on switch behavior changed across the 20-trial block. In addition to the random intercept per participant to account for individual variation in switch behavior, two random slopes were added to account for individual differences in the impact of feedback on switch behavior and the impact of trial position on switch behavior.

Models were estimated using the lme4 and AFEX package in R (Bates et al., 2015; Singmann et al., 2020), using likelihood ratio tests for fitting. Model comparison was performed using the Akaike information criterion (AIC) (Akaike, 1973), with a delta AIC-threshold of 2 (lower value indicating a better model fit). Data was pulled over all blocks, and statistical tests were considered significant at $p < 0.05$. Less than one percent of the data consisted of no responses [mean (SD): 0.9% (1.1%); min: 0% max: 5%]; these trials were not included in the analysis. No-response trials, trials preceded by a no-response trial, and the first trial of each block were excluded from the switch analyses, as these could not be classified as a switch or no-switch trial.

MRI Acquisition

A high resolution T1-weighted anatomical image was collected using a 64-channel head coil. Functional data were acquired with an echo-planar imaging (EPI) sequence (TR = 1.25 s, TE = 30 ms, 60 axial slices, 2.0 × 2.0 × 3.0 mm voxel size).

MRI Analysis

Preprocessing

All MRI analyses were performed using SPM12 toolbox (http://www.fil.ion.ucl.ac.uk/spm/). Before preprocessing the data were checked for excessive movements that could not be corrected for. No participants were excluded from further analysis. Functional data were corrected for slice acquisition time, and realignment was applied to correct for head movements. Then the images were registered to the corresponding anatomical image, and all images were spatially normalized to the ICBM space template. The data were spatially smoothed using a 6 mm Gaussian kernel.

General Linear Model Analysis

Three general linear models (GLM's) were fitted on the individual subject data to examine different learning-related processes (feedback evaluation, feedback integration, and knowledge updating). For the feedback evaluation model, gain trials and loss trials were added to the design matrix as predictors of interest. On the individual subject level, beta-estimates were calculated for gains ($ß_{gain}$) and losses ($ß_{loss}$) that were subsequently used to

compare the two age groups using second-level contrasts (e.g., Young ß$_{gain}$ – ß$_{loss}$ – Old ß$_{gain}$ – ß$_{loss}$).

For the feedback integration model, four predictors were constructed based on the possible feedback combinations on the previous and current trial (e.g., gain-loss, loss-loss, etc.). The individual beta estimates of the different feedback sequences were used in second-level analyses to assess the effects of integration across trials. More specifically, we examined whether neural activation following feedback presentation of a loss or gain depended on whether it was preceded by a gain or loss on the previous trial. For example, neural differences identified when comparing gain-loss vs. loss-loss would be indicative of previous-trial feedback (gain vs. loss) impacting the processing of current loss feedback, which would imply feedback integration over trials. Similarly, we compared neural activity following a gain on the current trial depending on whether it was preceded by a gain or a loss (loss-gain vs. gain-gain). These effects were subsequently compared between the age groups [e.g., Young (ß$_{gain-loss}$ − ß$_{loss-loss}$) − Old (ß$_{gain-loss}$ − ß$_{loss-loss}$) and Young (ß$_{loss-gain}$ − ß$_{gain-gain}$) − Old (ß$_{loss-gain}$ − ß$_{gain-gain}$)].

Finally, to examine brain activation related to knowledge updating, two predictors reflecting switch and no-switch choice behavior on the next trial were constructed and added to the GLM model. The switch predictor contained all trials where the participants would switch their choice between stimulus types (i.e., choose face vs. choose house) on the next trial. The no-switch predictor consisted of all trials where the participant would choose the same stimulus type on the next trial. On the individual subject level, beta-estimates were calculated for switch trials (ß$_{switch}$) and no-switch trials (ß$_{noswitch}$), which were subsequently used in second-level analyses to assess group differences.

All models contained six movement covariates to account for head movements during the task, and all regressors were convolved with a canonical hemodynamic response function (HRF). The GLM analyses modeled the BOLD activity for a period of 2,000 ms starting at the moment of feedback presentation, based on previous EEG-work showing that the processes of feedback evaluation and knowledge updating last approximately this long (van den Berg et al., 2019). A high-pass filter with a 128-s time constant was used to remove low frequency drifts. All second-level contrasts were evaluated by whole-brain voxel-wise *t*-tests with a *p*-value threshold of 0.001 (uncorrected) and a minimal cluster-size of five voxels. Furthermore, FWE-corrections at the cluster level were applied ($p_{FWE} < 0.05$).

RESULTS

Behavioral Results

Learning Rate

Analyses of task performance showed that the probability of choosing the block winner increased across the block for both young and older participants, indicating the ability to learn which stimulus was more likely to yield a gain within the block [$\chi^2(2) = 151.75$, $p < 0.01$] (**Figure 2A**). *Post-hoc* tests revealed that, at the start of the block (i.e., on the 1st trial), both younger and older participants chose the block winner around chance level (younger adults: $p_{(\text{choose block winner})} = 0.52$ vs. older adults: $p_{(\text{choose block winner})} = 0.53$; odds ratio = 0.06, SE = 0.21, n.s.). At the end of the block (i.e., on the 20th trial), the probability of choosing the block winner was higher in both age groups, but was higher in younger participants compared to older participants [$p_{(\text{choose block winner})} = 0.88$ vs. $p_{(\text{choose block winner})} = 0.77$, respectively; odds ratio = 0.80, SE = 2.06, $p < 0.01$], reflecting a higher learning level for the younger group.

On average, younger participants had a steeper learning curve than older adults [$\chi^2(2) = 16.78$, $p < 0.01$; average learning rate$_{young} = 0.11$ (SD = 0.07); average learning rate$_{old} = 0.07$ (SD = 0.07)]. However, we also observed large individual differences in learning rate (**Figure 2B**). Hence, to further investigate if differences in brain activity in subsequent analysis of the fMRI BOLD signal can be explained by general age-related differences or by differences in learning rate, we took into consideration learning rate as a potential confounding factor in these analyses.

Switch Behavior

Switch probability (i.e., choosing a different stimulus category [houses or faces] on the current trial than on the previous trial) decreased over the course of a 20-trial block [$\chi^2(1) = 24.16$, $p < 0.01$]. Although on average the older group switched more often than younger participants [mean number of switches (SD): young = 29 (19), old = 46 (26); $\chi^2(1) = 8.52$, $p < 0.01$], the decrease in switch probability over the block was not significantly different between the two age groups [$\chi^2(1) = 0.18$, n.s.]. When taking into account the individual differences in learning performance in the mixed model, we found that higher learning rates were associated with a lower average switch probability [$\chi^2(1) = 12.40$, $p < 0.01$]. These effects were found to be dependent on an interaction between age and trial number [Age × Trial number × Learning rate: $\chi^2(1) = 6.52$, $p = 0.01$]. *Post-hoc* testing showed that switch probability during the block was similar for young adults with low learning rates and older adults with high learning rates. Young adults with high learning rates had a lower switch probability compared to the other groups and older adults with low learning rates had significantly higher switch probabilities compared to the other three groups (**Figure 3A**).

Effects of Feedback Information

In a further analysis we examined whether switch behavior was modulated by feedback history. More specifically, we inspected the impact of feedback valence (gain or loss) on the probability to make a switch by taking into account in the model both feedback on the previous trial (feedback$_{n-1}$) and feedback on the trial before that (feedback$_{n-2}$). Switch probability was higher following a loss compared to a gain [Feedback$_{n-1}$: $\chi^2(1) = 110.68$, $p < 0.01$]. In addition, feedback on the trial two trials back also influenced the switch probability [Feedback$_{n-2}$: $\chi^2(1) = 42.68$, $p < 0.01$]; however, this effect was dependent on feedback$_{n-1}$ [Feedback$_{n-1}$ × Feedback$_{n-2}$: $\chi^2(1) = 16.57$, $p < 0.01$]. As reflected through *post-hoc* tests, the probability to make a switch was highest if both feedback$_{n-1}$ and feedback$_{n-2}$ were losses, and lowest after receiving two gains or a loss followed by a gain [$p(\text{switch})_{loss-loss} = 0.36$, $p(\text{switch})_{gain-loss} = 0.18$, $p(\text{switch})_{loss-gain} = 0.08$, $p(\text{switch})_{gain-gain} = 0.06$].

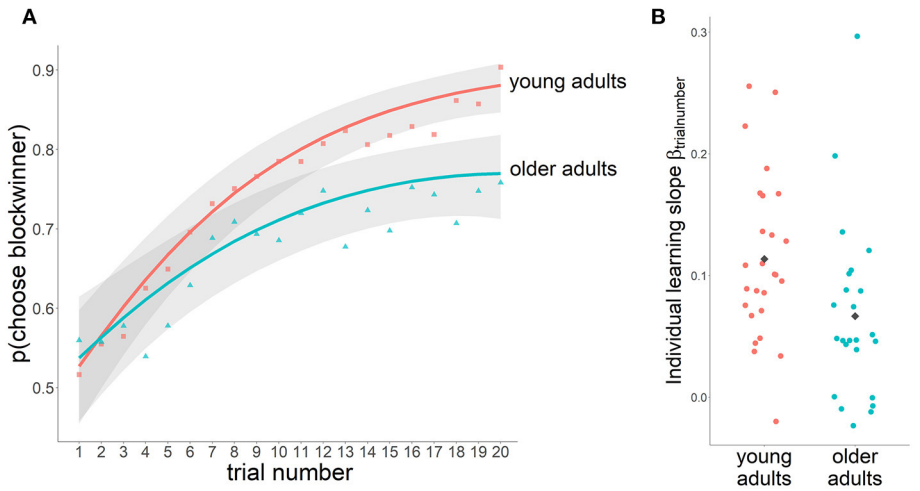


FIGURE 2 | Learning rate data (lines depict the logistic model fit with a 95% CI in gray, raw data represented as separate points). **(A)** Probability of choosing the block winner on the different trials across the block for young and older adults; **(B)** estimated individual learning rates, gray diamonds represent the average learning rate per group.

Feedback history effects were found not to differ between the two age groups in the model without the individual learning slopes [$Feedback_{n-2} \times Feedback_{n-1} \times$ Age: $\chi^2(1) = 0.02$, n.s.] (**Figure 3B**), but the higher switch rate in older compared to young adults was more pronounced following a gain as compared to a loss [Age $\times$ $Feedback_{n-1}$, $\chi^2(1) = 4.59$, $p = 0.03$]. After adding individual learning slopes to the model, the age effects were no longer significant [main effect of Age: $\chi^2(1) = 2.28$, n.s.; Age $\times$ $Feedback_{n-1}$ interaction: $\chi^2(1) = 2.98$, n.s.]. Instead, learning rate was predictive of the switch probability, with higher learning rates being related to a lower overall switch probability [$\chi^2(1) = 5.73$, $p = 0.02$]. After accounting for individual learning slopes, no interactions between feedback history and age or learning rate remained.

Lastly, in addition to our planned analyses we examined whether the relation between feedback and switch probability changed over the course of the 20-trial block. An interaction was found [$Feedback_{n-1} \times$ Trial position: $\chi^2(1) = 3.85$, $p = 0.05$], indicating that whereas the switch probability following a gain remained the same over the course of the block, switch probability following a loss decreased over the block (**Figure 3C**). This effect was found not to differ with age [$Feedback_{n-1} \times$ Trial position $\times$ Age: $\chi^2(1) = 2.95$, n.s.] or learning rate [$Feedback_{n-1} \times$ Trial position $\times$ Age: $\chi^2(1) = 1.36$, n.s.]. Hence, overall we found that older adults switched more compared to young adults, but we found no evidence for different learning behavior between the two groups.

fMRI Results

Feedback Evaluation

We found clusters with higher levels of BOLD activity following gains compared to losses throughout the brain, in frontal, parietal, temporal, occipital, and subcortical regions, in line with previous research (e.g., Cohen et al., 2008; Drueke et al., 2015; Andreou et al., 2017) (See **Supplementary Table S1** for a full overview of clusters). Conversely, we also found a trend toward higher activity following a loss compared to a gain in a small part of the frontal operculum, a part of the superior frontal gyrus, and in a part of the supplementary motor area and middle and anterior cingulate cortex (more superior compared to the gain activity effect), areas previously reported as being more active following negative feedback (Cohen et al., 2008; Bischoff-Grethe et al., 2009; Amiez et al., 2016). In addition to the general effects of feedback valence, we also found differences in feedback-related BOLD activity between the two age groups (contrast: $young_{gain>loss} > old_{gain>loss}$), especially in parts of the middle frontal gyrus, anterior cingulate cortex, and angular gyrus. The involvement of these areas in feedback evaluation (Cox et al., 2005; Cohen et al., 2008; Mies et al., 2011; Jahn et al., 2014; Lee and Kim, 2014; Drueke et al., 2015) and the diminished increase in activity following gains in older compared to younger adults observed here (**Figure 4A**), suggest that especially the processing of positive feedback is reduced with age. Although previous studies have reported a stronger ventral striatal response to positive feedback in older adults (Schott et al., 2007; Widmer et al., 2017), we did not find an effect of age in this area. In addition to the age effects, our results showed that activity following feedback was also different for participants with a low versus a high learning rate. More specifically, the difference between gains and losses was more pronounced for the participants with a high learning rate in three frontal lobe clusters covering areas related to executive functioning, such as the supplementary motor cortex, middle cingulate gyrus and the postcentral gyrus.

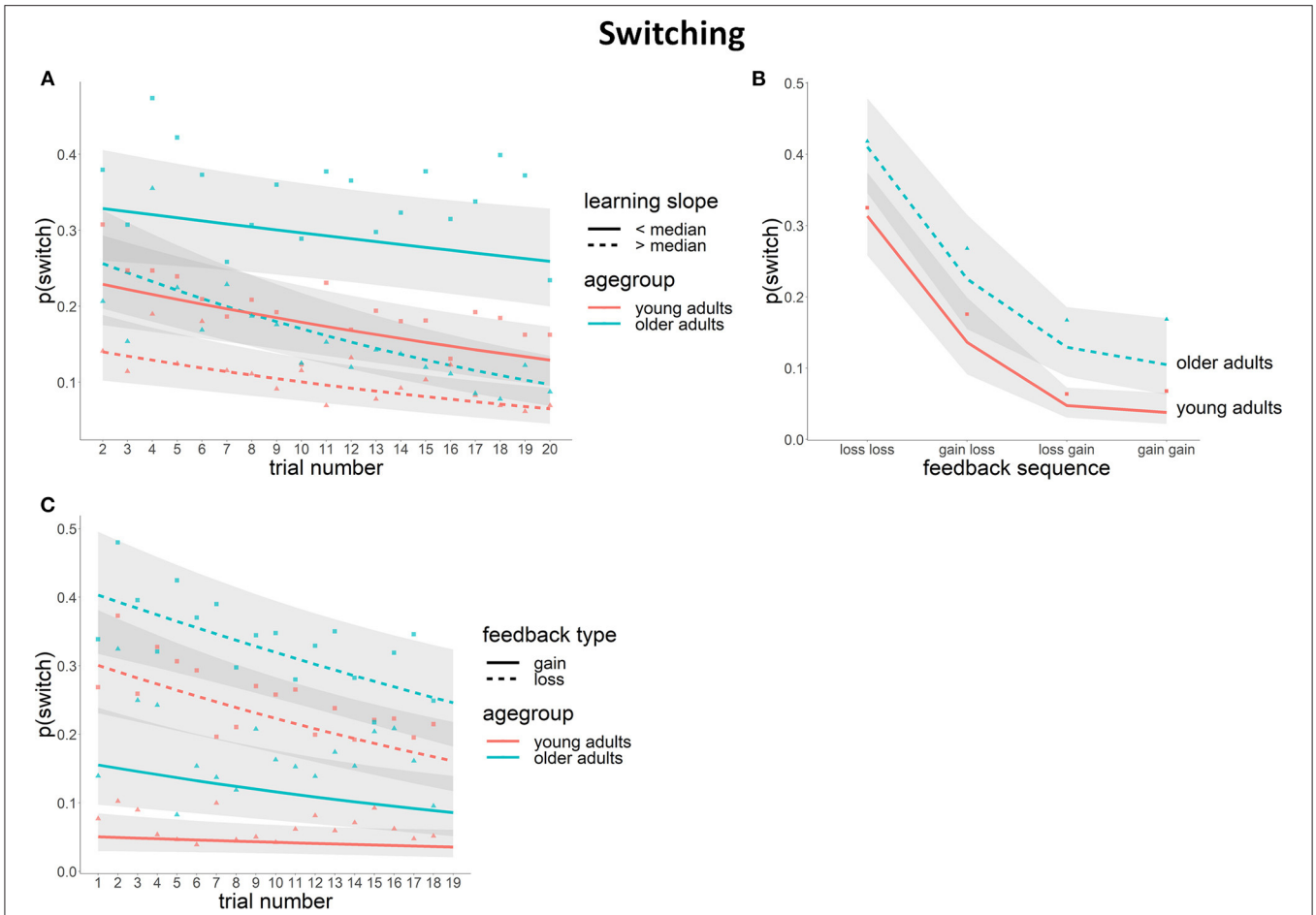


FIGURE 3 | Switch behavior [lines depict the logistic model fit with a 95% CI in gray, raw data (indicating the proportion of switch) are depicted as separate points]. **(A)** switch probability across the block for young and older participants with high and low learning rates (median split per age group); **(B)** switch probability per feedback sequence for young and older adults; **(C)** switch probability across the block per feedback type and age group.

Switch-Related Activity

In order to examine knowledge-updating effects, we compared levels of BOLD activity following feedback presentation on trials preceding a trial on which participants switched to the other stimulus category (switch trials), relative to trials preceding a no-switch trial. Of particular interest were the frontal and centro-pariatal brain areas, given that they are thought to be the origin of the LPC component linked to switch behavior by EEG research (Chase et al., 2011; San Martin et al., 2013). In the visualization (**Figure 4B**) we focused on the anterior cingulate cortex given its role in the processing of feedback and the updating of beliefs (Shenhav et al., 2013).

We found higher levels of activity preceding switch trials in frontal and parietal brain areas, the cerebellum, and the insular cortices (for a full overview of significant clusters see **Supplementary Table S2**). A direct comparison between the two age groups showed that switch-specific activity was more pronounced in young compared to older participants in three clusters, one covering parts of the anterior and middle cingulate cortex, the superior frontal gyrus, and the supplementary motor cortex, one covering parts of the precentral and postcentral gyrus, and one covering parts of the middle and superior frontal gyrus (**Figure 4B**). These results suggest enhanced knowledge updating in young adults prior to making a switch. Switch-related enhancement of activity was not observed in the older group, suggesting a less pronounced link between knowledge updating and choice behavior in this group. Besides the age-related effects, the increase in activity preceding switch compared to no-switch trials was found to be more pronounced in the group with high learning rates compared to low learning rates in parts of the middle frontal gyrus, medial superior frontal gyrus, and anterior cingulate cortex related to executive functioning.

Feedback History

The integration of feedback information over trials was examined by comparing neural activity following different feedback sequences to investigate if activity following current feedback depended on feedback on the previous trial (e.g., comparing a gain-loss sequence to a loss-loss one). Our results showed that neural activity after receiving gains and losses depended

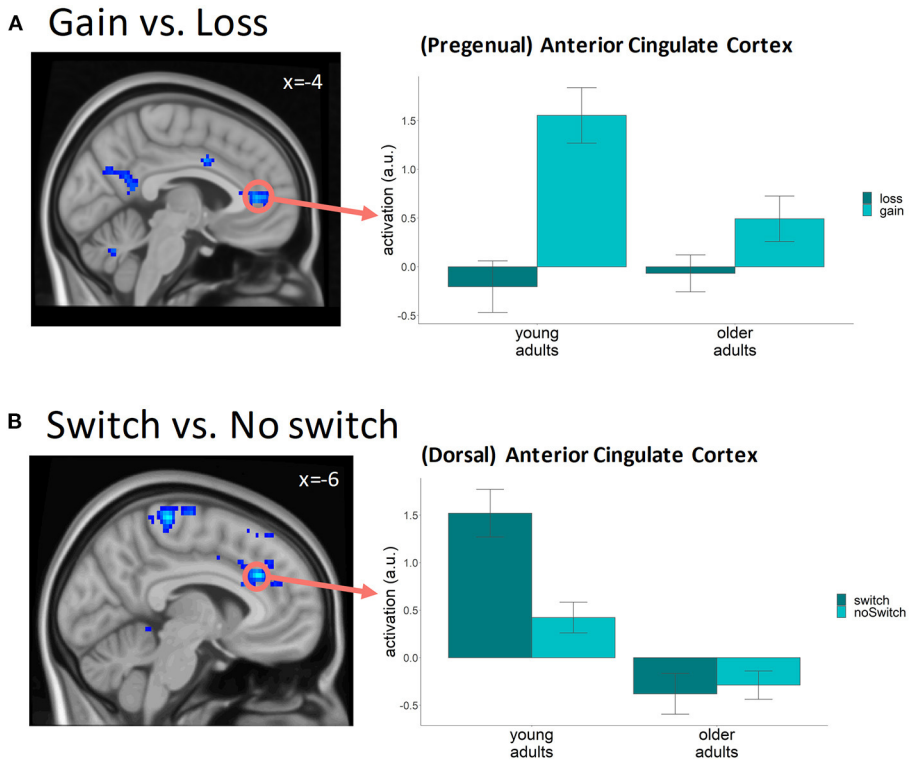


FIGURE 4 | fMRI results ($p < 0.001$ uncorrected voxel level, $p_{FWE-cluster} < 0.05$) **(A)** Cluster covering the anterior cingulate cortex where the difference in brain activity between gains and losses was bigger in young compared to older participants. Bar graph was based on a 5 mm sphere ROI centered around the approximate center of the ACC cluster (MNI −4, 40, 8); **(B)** Clusters where the difference in BOLD activity preceding a switch on the next trial compared to a no-switch is bigger in young as compared to older participants. Bar graph is based on a 5 mm sphere ROI centered around the peak voxel in the left anterior cingulate gyrus (MNI −6, 28, 29).

on the feedback received on the previous trial. When a gain was preceded by a loss as compared to a gain, more activity was found in areas related to reward processing, including the superior frontal gyrus, the supramarginal gyrus, and the anterior insula, suggesting that receiving negative feedback enhances reward processing of positive feedback on the next trial (see **Supplementary Table S3** for full results). No significant clusters were found where activity was higher when the gain was preceded by a gain as compared to when it was preceded by a loss. With regard to receiving negative feedback, we found enhanced activity in the inferior temporal gyrus and fusiform gyrus when a loss was preceded by a gain compared to a loss. When a loss was preceded by another loss we found higher levels of activity in areas involved in feedback evaluation and decision making, including like the medial superior frontal gyrus and anterior insula, suggesting that receiving negative feedback enhances processing of the subsequent loss.

Taken together, these results suggest that feedback processing is impacted by preceding feedback, and especially by negative feedback. Direct group comparisons did not reveal any significant differences in the effect of feedback history between young and older participants. In participants with a high learning rate compared to a low learning rate, however, we did find a larger enhancement of activity in the right anterior insula and frontal operculum when a loss was preceded by another loss compared to a gain, suggesting a more prominent effect of previous feedback in participants with high learning rates when receiving negative feedback.

DISCUSSION

In this study, we aimed to elucidate age-related differences in the neural mechanisms underlying learning. Older and younger adults performed a probabilistic learning task in which they were asked to learn which one out of two stimulus types was more likely to lead to a gain. The behavioral results indicate that over the course of a block of 20 trials, participants increasingly chose the option that was more likely to yield a gain, showing that they were able to learn from the feedback. However, in line with previous research, we also found that older participants had more difficulty learning from the probabilistic feedback compared to young participants, as indicated by a lower learning rate (Mell et al., 2005; Hämmerer et al., 2011). This diminished performance in older adults coincided with altered choice behavior, as well as in differences in feedback evaluation and knowledge updating at the neural level.

In order to gain more insight in how differences in learning performance arise, it is valuable to examine the choice behavior leading up to this performance in more detail. During our task, participants were given feedback on each trial, and based on this

information they could decide whether or not to switch their choice from one stimulus to the other on the next trial. Our results show that participants indeed used feedback information to adapt their choice behavior, switching more often following a loss compared to a gain. However, because of the probabilistic nature of the feedback, participants knew to expect a loss occasionally even when choosing the correct stimulus (i.e., block winner). Therefore, feedback information should not be used only at a trial level, but it should also be integrated over trials in order to learn which of the two choice options was more likely to be the block winner and to choose that option more and more often. In line with this, the probability to switch to the other stimulus was found to not only depend on the most recent feedback, but also on feedback history. Furthermore, switch behavior was shown to be related to learning performance, with the making of fewer switches in general being associated with a higher learning rate. Together these results indicate that switch choice behavior was guided by feedback information and linked to learning performance, underlining its relevance in the study of feedback related learning.

As feedback information was accumulated over the course of each 20-trial block, participants were expected to become more certain about which stimulus was the block winner and, as a result, become more consistent in their choices. This was reflected in the decreasing number of switches across the block, in combination with the increasing chance of choosing the block winner. At the same time, this growing bias toward choosing the block winner across the block can be expected to coincide with requiring more conflicting evidence before switching to the other stimulus. In other words, because of the probabilistic feedback, participants will be more likely to accept more losses when they are more convinced their choice is right. As certainty about the identity of the block winner is higher at the end of the block compared to the beginning, participants are expected to accept more losses at the end of the block before adapting their choice behavior. We indeed showed that switch probability following a loss decreased across the block, indicating that the information value of the feedback decreased as participants became more certain of the block winner. These results show how choice behavior was not only adapted based on feedback information received, but also as participants learned the right choice (e.g., with time on task).

Importantly, switch choice behavior and its modulation depended on age. First of all, older adults were found to switch more in general compared to younger adults, regardless of feedback valence or trial number within the block. One of the factors that might be related to this difference in switch tendency is the uncertainty of probabilistic feedback. It could be argued that older adults are less able to use relative uncertainty to guide learning (Nassar et al., 2016). Given that the value of feedback at the start of the 20-trial block is more uncertain compared to the end of the block when the participant has learned the stimulus reward associations, optimal use of uncertainty in the learning process would implicate a decreasing switch tendency following negative feedback across the block associated with the declining level of uncertainty. Suboptimal use of uncertainty in older adults would therefore imply a reduced adaptation of switch tendency across the block, which is not in line with our findings of a comparable modulation of the probability that participants shift between the choice options across the block in the two age groups following negative feedback, suggesting that the higher switch rate in older adults is most likely not driven by a less efficient use of uncertainty during learning. However, the higher general switch rate in older adults in combination with the fact that this higher tendency was more pronounced following positive compared to negative feedback, might also suggest that older adults adhere to a different learning strategy that results in alternative choice behavior and less successful learning. Our findings are in line with findings of Hämmerer et al. (2011), who also reported a larger age-related difference in switch rate following positive feedback. They argued that this suggests that switch choice behavior is affected less strongly by positive feedback in older compared to younger adults, meaning that that older adults learn less from positive feedback compared to negative feedback. Taken together, our results suggest that age impacts learning-related behavior at the trial level, both with regard to the magnitude of switch choice tendency and how it is affected by feedback valence.

In line with the fact that the influence of positive feedback on subsequent choice behavior was different between the two age groups, our fMRI analysis showed age-related differences related to the processing of positive feedback. More specifically, we found that even though neural activity levels following gains were higher compared to losses in both age groups, older adults demonstrated a less differential BOLD response to gains and losses in brain areas that have been associated with, amongst other things, feedback evaluation (e.g., ACC, Alexander and Brown, 2011; Neubert et al., 2015; Kolling et al., 2016; Bradley et al., 2017; Wiseman et al., 2018). ROI visualizations indicate that this smaller differential response to gains and losses was due to a diminished neural response to positive feedback (i.e., because BOLD activity increased less compared to baseline following positive feedback in the older adults, the activity levels following gains and losses were more similar in this group). A less differential neural response to feedback valence in older adults has also been reported in EEG-studies in which smaller differences in the feedback related negativity (FRN) component were found (Mathewson et al., 2008; Hämmerer et al., 2011), which has been argued to reflect a reduced ability to discriminate feedback valence in light of the task goal (Hämmerer et al., 2011). Our results show that age effects were specifically driven by a smaller increase in activity in brain areas following positive feedback in older compared to younger adults. In combination with our behavioral results, this suggests that positive feedback is processed differently in the older group, potentially altering its contribution to the learning process, and resulting in a diminished ability to make optimal behavioral changes. We did not replicate previous findings of a stronger ventral striatal response to the rewarding feedback in older as compared to younger adults (Schott et al., 2007; Widmer et al., 2017), which might be related to differences in task characteristics between our study and previous research. In the probabilistic learning task that we used, participants had to continuously learn which stimulus type was associated with a higher chance

of winning, while the stimulus-reward associations in these previous studies were known beforehand, thereby reducing the degree of uncertainty in a task. This difference might have an impact on the information value of feedback information and the influences of feedback on subsequent behavior.

When comparing neural activity related to the different feedback sequences, higher levels of activity were found in the inferior temporal gyrus and fusiform gyrus when negative feedback was preceded by positive versus negative feedback. As these brain areas have been associated with the processing of faces and houses, it could be that these results reflect cortical (re)activation of stimulus processing areas as part of establishing, updating and storage of stimulus-reward associations, as we reported earlier in van den Berg et al. (2019). However, our dataset does not lend itself for a more detailed analysis of this effect, and thus this interpretation remains speculative and would need to be confirmed by future research.

In addition to feedback evaluation, we examined the neural processes related to switch choice behavior in order to investigate potential age-related differences in knowledge updating. Previous EEG-research showed that making a switch on the next trial was accompanied by a larger Late Positive Complex (LPC) component from 300 to 600 ms post feedback stimulus, which was thought to reflect enhanced brain activity in frontal and centro-parietal brain areas prior to a switch (Fischer and Ullsperger, 2013; San Martin et al., 2013). One of the potentially contributing brain areas of particular interest is the anterior cingulate cortex (ACC), which is thought to be serving as a central hub in the allocation of cognitive control and is involved in processing of feedback and updating of beliefs (Shenhav et al., 2013).

In line with the EEG-studies (Chase et al., 2011; San Martin et al., 2013; Correa et al., 2018), our results indeed showed increased activity in the ACC following feedback presentation on the trial prior to a switch in young adults, which might reflect increased levels of knowledge updating prior to behavioral adjustment. In older adults, however, no evidence was found for increased ACC activity on the trial prior to a switch, suggesting that updating of knowledge was less effective in this group. Based on the results of the present study, we cannot be certain whether the equivalent levels of switch- and no-switch related neural activity in older adults specifically reflected relatively diminished knowledge updating prior to a behavioral switch or relatively enhanced knowledge updating prior to not switching in comparison to young adults. In addition, as switch choice behavior is inherently linked to feedback valence (higher switch tendencies following negative compared to positive feedback), and fMRI has limited temporal resolution, distinguishing between feedback evaluation and knowledge updating processes is limited with this method. The low temporal resolution of fMRI does not allow us to temporally distinguish processes related to feedback evaluation feedback integration and knowledge updating, thus restricting us to interpreting our results based on the specific contrasts showing the differences in activation. This however does not guarantee that the activation patterns are exclusive for these processes. Nevertheless, the distinct spatial locations of the reported feedback-evaluation and knowledge-updating effects, corresponding to diminished brain activity in feedback-evaluation areas following positive feedback in older adults, and a less prominent increase of activity in areas involved in the updating of beliefs prior to adaptations in choice behavior in this group, seem to indicate that age impacts both feedback processing and knowledge updating during learning.

In conclusion, successful learning requires the evaluation of feedback, updating of knowledge, and adaptation of behavior. We provided evidence showing that diminished learning performance in older adults corresponded with making more switch choices. In addition, older adults were shown to process positive feedback to a lesser extend compared to young adults, which might have led to the reduced knowledge updating we found in this group. Together our findings showed how learning-related processes are impacted by age both at the behavioral and neural levels, and how different feedback-evaluation processes could lead to reduced learning performance across trials.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on reasonable request. Requests to access the datasets should be directed to Berry van den Berg, berry.van.den.berg@rug.nl.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University Medical Center Groningen medical ethics committee. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

BB, MW, and ML: study design. TH and BB: data acquisition and data processing. TH, BB, MW, AA, and ML: manuscript writing. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.635996/full#supplementary-material

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. Second Int. Symp. Inform. Theory*. 1973, 267–281.

Alexander, W. H., and Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* 14, 1338–1344. doi: 10.1038/nn.2921

Amiez, C., Wutte, M. G., Faillenot, I., Petrides, M., Burle, B., and Procyk, E. (2016). Single subject analyses reveal consistent recruitment of frontal

operculum in performance monitoring. *NeuroImage* 133, 266–278. doi: 10.1016/j.neuroimage.2016.03.003

Andreou, C., Frielinghaus, H., Rauh, J., Mußmann, M., Vauth, S., Braun, P., et al. (2017). Theta and high-beta networks for feedback processing: a simultaneous EEG-fMRI study in healthy male subjects. *Transl. Psychiatry* 7:e1016. doi: 10.1038/tp.2016.287

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R., and Grafton, S. (2009). The influence of feedback valence in associative learning. *NeuroImage* 44, 243–251. doi: 10.1016/j.neuroimage.2008.08.038

Borst, J. P., and Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proc. Natl. Acad. Sci. U. S. A.* 110, 1628–1633. doi: 10.1073/pnas.1221572110

Bradley, K. A. L., Case, J. A. C., Freed, R. D., Stern, E. R., and Gabbay, V. (2017). Neural correlates of RDoC reward constructs in adolescents with diverse psychiatric symptoms: a Reward Flanker Task pilot study. *J. Affect. Disord.* 216, 36–45. doi: 10.1016/j.jad.2016.11.042

Chase, H. W., Swainson, R., Durham, L., Benham, L., and Cools, R. (2011). Feedback-related negativity codes prediction error but not behavioral adjustment during probabilistic reversal learning. *J. Cogn. Neurosci.* 23, 936–946. doi: 10.1162/jocn.2010.21456

Cohen, M. X., Elger, C. E., and Weber, B. (2008). Amygdala tractography predicts functional connectivity and learning during feedback-guided decision-making. *NeuroImage* 39, 1396–1407. doi: 10.1016/j.neuroimage.2007.10.004

Cohen, M. X., Wilmes, K. A., and van de Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends Cogn. Sci.* 15, 558–566. doi: 10.1016/j.tics.2011.10.004

Correa, C. M. C., Noorman, S., Jiang, J., Palminteri, S., Cohen, M. X., Lebreton, M., et al. (2018). How the level of reward awareness changes the computational and electrophysiological signatures of reinforcement learning. *J. Neurosci.* 38, 10338–10348. doi: 10.1523/JNEUROSCI.0457-18.2018

Cox, K. M., Aizenstein, H. J., and Fiez, J. A. (2008). Striatal outcome processing in healthy aging. *Cogn. Affect. Behav. Neurosci.* 8, 304–317. doi: 10.3758/CABN.8.3.304

Cox, S. M. L., Andrade, A., and Johnsrude, I. S. (2005). Learning to like: a role for human orbitofrontal cortex in conditioned reward. *J. Neurosci.* 25, 2733–2740. doi: 10.1523/JNEUROSCI.3360-04.2005

Drueke, B., Weichert, L., Forkmann, T., Mainz, V., Gauggel, S., and Boecker, M. (2015). Neural correlates of positive and negative performance feedback in younger and older adults. *Behav. Brain Funct.* 11:17. doi: 10.1186/s12993-015-0062-z

Eppinger, B., Hämmerer, D., and Li, S. C. (2011). Neuromodulation of reward-based learning and decision making in human aging. *Ann. N. Y. Acad. Sci.* 1235, 1–17. doi: 10.1111/j.1749-6632.2011.06230.x

Eppinger, B., Kray, J., Mock, B., and Mecklinger, A. (2008). Better or worse than expected? *Aging Learn. ERN Neuropsychol.* 46, 521–539. doi: 10.1016/j.neuropsychologia.2007.09.001

Ferdinand, N. K. (2019). The influence of task complexity and information value on feedback processing in younger and older adults: no evidence for a positivity bias during feedback-induced learning in older adults. *Brain Res.* 1717, 74–85. doi: 10.1016/j.brainres.2019.04.011

Ferdinand, N. K., and Czernochowski, D. (2018). Motivational influences on performance monitoring and cognitive control across the adult lifespan. *Front. Psychol.* 9:1018. doi: 10.3389/fpsyg.2018.01018

Fischer, A. G., and Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron* 79, 1243–1255. doi: 10.1016/j.neuron.2013.07.006

Frank, M. J., and Kong, L. (2008). Learning to avoid in older age. *Psychol. Aging* 23, 392–398. doi: 10.1037/0882-7974.23.2.392

Hämmerer, D., Li, S.-C., Müller, V., and Lindenberger, U. (2011). Life span differences in electrophysiological correlates of monitoring gains and losses during probabilistic reinforcement learning. *J. Cogn. Neurosci.* 23, 579–592. doi: 10.1162/jocn.2010.21475

Jahn, A., Nee, D. E., Alexander, W. H., and Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. *NeuroImage* 95, 80–89. doi: 10.1016/j.neuroimage.2014.03.050

Kolling, N., Behrens, T. E. J., Wittmann, M. K., and Rushworth, M. F. S. (2016). Multiple signals in anterior cingulate cortex. *Curr. Opin. Neurobiol.* 37, 36–43. doi: 10.1016/j.conb.2015.12.007

Lee, W., and Kim, S. Il. (2014). Effects of achievement goals on challenge seeking and feedback processing: behavioral and fMRI evidence. *PLoS ONE* 14, 1–12. doi: 10.1371/journal.pone.0107254

Marin, R. S., Biedrzycki, R. C., and Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiatry Res.* 38, 143–162. doi: 10.1016/0165-1781(91)90040-V

Marschner, A., Mell, T., Wartenburger, I., Villringer, A., Reischies, F. M., and Heekeren, H. R. (2005). Reward-based decision-making and aging. *Brain Res. Bull.* 67, 382–390. doi: 10.1016/j.brainresbull.2005.06.010

Martinez, A., and Benavente, R. (1998). *The AR Face Database.* Technical Report-24. Barcelona: Computer Vision Center (CVC).

Mata, R., Josef, A. K., Samanez-Larkin, G. R., and Hertwig, R. (2011). Age differences in risky choice: a meta-analysis. *Ann. N. Y. Acad. Sci.* 1235, 18–29. doi: 10.1111/j.1749-6632.2011.06200.x

Mathewson, K. J., Dywan, J., Snyder, P. J., Tays, W. J., and Segalowitz, S. J. (2008). Aging and electrocortical response to error feedback during a spatial learning task. *Psychophysiology* 45, 936–948. doi: 10.1111/j.1469-8986.2008.00699.x

Mell, T., Heekeren, H. R., Marschner, A., Wartenburger, I., Villringer, A., and Reischies, F. M. (2005). Effect of aging on stimulus-reward association learning. *Neuropsychologia* 43, 554–563. doi: 10.1016/j.neuropsychologia.2004.07.010

Mies, G. W., van der Molen, M. W., Smits, M., Hengeveld, M. W., and Van der Veen, F. M. (2011). The anterior cingulate cortex responds differently to the validity and valence of feedback in a time-estimation task. *NeuroImage* 56, 2321–2328. doi: 10.1016/j.neuroimage.2011.04.015

Nassar, M. R., Bruckner, R., Gold, J. I., Li, S. C., Heekeren, H. R., and Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nat. Commun.* 7, 1–13. doi: 10.1038/ncomms11609

Neubert, F. X., Mars, R. B., Sallet, J., and Rushworth, M. F. S. (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 112, 2695–2704. doi: 10.1073/pnas.1410767112

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019

R Core Team (2020). *R: a language and environment for statistical computing.* R Foundation for Statistical Computing. Available online at: https://www.r-project.org/ (accessed November 28, 2020).

Reijnders, J. S. A. M., Scholtissen, B., Weber, W. E. J., Aalten, P., Verhey, F. R. J., and Leentjens, A. F. G. (2010). Neuroanatomical correlates of apathy in Parkinson's disease: a magnetic resonance imaging study using voxel-based morphometry. *Mov. Disord.* 25, 2318–2325. doi: 10.1002/mds.23268

Samanez-Larkin, G. R., Worthy, D. A., Mata, R., McClure, S. M., and Knutson, B. (2014). Adult age differences in frontostriatal representation of prediction error but not reward outcome. *Cogn. Affect. Behav. Neurosci.* 14, 672–682. doi: 10.3758/s13415-014-0297-4

San Martin, R., Appelbaum, L. G., Pearson, J. M., Huettel, S. A., and Woldorff, M. G. (2013). Rapid brain responses independently predict gain maximization and loss minimization during economic decision making. *J. Neurosci.* 33, 7011–7019. doi: 10.1523/JNEUROSCI.4242-12.2013

Schott, B. H., Niehaus, L., Wittmann, B. C., Schütze, H., Seidenbecher, C. I., Heinze, H. J., et al. (2007). Ageing and early-stage Parkinson's disease affect separable neural mechanisms of mesolimbic reward processing. *Brain* 130, 2412–2424. doi: 10.1093/brain/awm147

Shenhav, A., Botvinick, M. M., and Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* 79, 217–240. doi: 10.1016/j.neuron.2013.07.007

Singmann, H., Bolker, B., Westfall, J., Aust, F., and Ben-Shachar, M. S. (2020). *afex: Analysis of Factorial Experiments (R package version 0.28-0).* Available online at: https://cran.r-project.org/package=afex (accessed November 28, 2020).

van den Berg, B., Geib, B. R., Martín, R. S., and Woldorff, M. G. (2019). A key role for stimulus-specific updating of the sensory cortices in the learning of stimulus–reward associations. *Social Cogn. Affect. Neurosci.* 14, 173–187. doi: 10.1093/scan/nsy116

Widmer, M., Stulz, S., Luft, A. R., and Lutz, K. (2017). Elderly adults show higher ventral striatal activation in response to motor performance related rewards than young adults. *Neurosci. Lett.* 661, 18–22. doi: 10.1016/j.neulet.2017.09.038

Wiseman, S. J., Booth, T., Ritchie, S. J., Cox, S. R., Muñoz Maniega, S., Valdés Hernández, M., et al. (2018). Cognitive abilities, brain white matter hyperintensity volume, and structural network connectivity in older age. *Hum. Brain Map.* 39, 622–632. doi: 10.1002/hbm.23857

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.

frontiers in Human Neuroscience

SYSTEMATIC REVIEW
published: 07 April 2021
doi: 10.3389/fnhum.2021.623315

Effects of Transcranial Direct Current Stimulation (tDCS) on Cognitive Performance and Cerebral Oxygen Hemodynamics: A Systematic Review

Mathieu Figeys[1]*, *Michael Zeeman*[2] and *Esther Sung Kim*[1,3]

[1] Faculty of Rehabilitation Medicine, University of Alberta, Edmonton, AB, Canada, [2] Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada, [3] Department of Communication Sciences and Disorders, University of Alberta, Edmonton, AB, Canada

Background: There is increasing evidence to support the efficacy of transcranial direct current stimulation (tDCS) applications in cognitive augmentation and rehabilitation. Neuromodulation achieved with tDCS may further regulate regional cerebral perfusion affiliated through the neurovascular unit; however, components of cerebral perfusion decrease across aging. A novel neuroimaging approach, functional near-infrared spectroscopy (fNIRS), can aid in quantifying these regional perfusional changes. To date, the interaction of the effects of tDCS on cognitive performance across the lifespan and obtained fNIRS hemodynamic responses remain unknown.

Objective: This review aims to examine the effects of tDCS on cognitive performance and fNIRS hemodynamic responses within the context of cognitive aging.

Methods: Six databases were searched for studies. Quality appraisal and data extraction were conducted by two independent reviewers. Meta-analysis was carried out to determine overall and subgroup effect sizes.

Results: Eight studies met inclusion criteria. The overall effect size demonstrates that tDCS can alter cognitive performance and fNIRS signals, with aging being a potential intermediary in tDCS efficacy.

Conclusion: From the studies included, the effects of tDCS on cognitive performance and fNIRS metrics are most prominent in young healthy adults and appear to become less robust with increasing age. Given the small number of studies included in this review further investigation is recommended.

Keywords: cognition, transcranial direct current stimulation (tDCS), functional near infrared spectroscopy (fNIRS), aging, cerebral perfusion, cerebral hemodynamic functional response

OPEN ACCESS

Edited by:
Susan Mérillat, University of Zurich, Switzerland

Reviewed by:
Adrian B. Curtin, Drexel University, United States
Jaclyn Anne Stephens, Colorado State University, United States

***Correspondence:**
Mathieu Figeys
figeys@ualberta.ca

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

Received: *29 October 2020*
Accepted: *11 March 2021*
Published: *07 April 2021*

Citation:
Figeys M, Zeeman M and Kim ES (2021) Effects of Transcranial Direct Current Stimulation (tDCS) on Cognitive Performance and Cerebral Oxygen Hemodynamics: A Systematic Review. Front. Hum. Neurosci. 15:623315. doi: 10.3389/fnhum.2021.623315

INTRODUCTION

Interventions to enhance cognitive functioning are increasingly being used as a potential avenue to combat the effects of dementia and age-related cognitive decline. These range from behavioral training programs to non-invasive brain stimulation (Butler et al., 2018; Song et al., 2018; Zhang et al., 2019; Chou et al., 2020). Transcranial direct current stimulation (tDCS), one type of

non-invasive brain stimulation, involves the application of a low-dose electrical current across the brain. tDCS is often paired with behavioral training protocols and is hypothesized to alter the efficacy of training-induced cognitive performance. Increasing evidence suggests that tDCS acts beyond neuronal structures and may modulate cerebral perfusion (Stagg et al., 2013). The relationships between the mechanisms of cognition, cerebral perfusion, and neuronal activity remain poorly understood, especially when considering healthy and pathological cognitive aging. With the use of functional near-infrared spectroscopy (fNIRS) to measure key factors in perfusion, as well as cognitive performance metrics, the impact of aging on these mechanisms can be explored. The purpose of this systematic review is to begin to explore the effects of tDCS on cognitive performance and fNIRS signals, with an emphasis on how these may differ across age.

Non-invasive Electrical Brain Stimulation

Among available transcranial electrical current stimulation modalities, tDCS, and transcranial alternating current stimulation (tACS) are the most commonly reported techniques within the literature (Polanía et al., 2018). Direct current (DC) stimulation is utilized in tDCS, compared to an oscillating sinusoidal-current at a set frequency used in tACS. The physiological effects of tACS neuromodulation are thought to target specific neuronal frequency bands (Polanía et al., 2018), compared to neural polarity modulation involving voltage-dependent ion channels in tDCS (Nitsche et al., 2003). These differences in electrical properties may result in different neurophysiological responses. In this review, we focus on the cognitive and cerebral perfusion effects of tDCS, in combined tDCS and fNIRS protocols.

Transcranial Direct Current Stimulation

Transcranial direct current stimulation (tDCS) is one form of non-invasive brain stimulation that has been used in numerous healthy and clinical populations (Meinzer et al., 2015; Prehn and Flöel, 2015; Smith et al., 2015; Cappon et al., 2016; Berryhill and Martin, 2018; Ke et al., 2019; Martinotti et al., 2019; Matar et al., 2020). Low-dose direct current applied to the brain is thought to modulate resting membrane threshold with application-dependent stimulation montages producing a differential increase or decrease in neuronal excitability (Nitsche and Paulus, 2000, 2001; Paulus, 2011; Stagg and Nitsche, 2011). The effects of tDCS are often examined using behavioral task metrics but reported results have been variable (Prehn and Flöel, 2015; Cappon et al., 2016; Woods et al., 2018; Rose et al., 2020).

Neuronal modulation induced by tDCS works in a summative fashion across neurons. Anodal tDCS is believed to invoke hypopolarization without reaching the depolarization threshold, whereas cathodal stimulation is thought to further shift the neuron into a hyperpolarized state (Stagg and Nitsche, 2011). These effects have proven beneficial in cognitive studies across aging and clinical populations; anodal tDCS has been demonstrated to increase performance on working memory (Ohn et al., 2008), cognitive control (Boudewyn et al., 2020), and language (Flöel et al., 2008). In contrast, cathodal stimulation has been demonstrated to decrease cognitive control (Wolkenstein et al., 2014). Thus, the potential clinical utility of tDCS targeting cognitive augmentation in aging and in cognitive disorders such as Mild Cognitive Impairment (MCI) may be of significant value.

tDCS can be easily paired with other treatment modalities, including cognitive rehabilitation protocols. For instance, researchers have reported that anodal-tDCS paired with cognitive training in young adults resulted in higher performance on a working memory task compared to the sham condition (Ke et al., 2019). Although these findings are promising, wide variability in terms of results and effect sizes exists within the tDCS literature. Numerous methodological variables including tDCS dosage, location, and length of stimulation, as well as population parameters such as age, education, and health status, may impact reported results. Overall, a consensus seems to be emerging that there is no clear advantage of adding tDCS to cognitive protocols (Cruz Gonzalez et al., 2018). Even with this uncertainty, the use of tDCS has been demonstrated to increase regional blood flow in those receiving tDCS paired with cognitive training (Das et al., 2019). Therefore, tDCS may potentially evoke other physiological and neurological mechanisms beyond behavioral responses.

Working memory is a cognitive function, which has been shown to be affected by age-related changes (Dickstein et al., 2013). In turn, aging may impact the efficacy of tDCS during working memory tasks. In a meta-analysis specifically examining the effects of tDCS on working memory in healthy young adults, no significant differences in performance were reported (Mancuso et al., 2016). However, when tDCS was paired with cognitive training, a small yet significant effect size was observed on working memory performance (Mancuso et al., 2016). A separate study investigating the effects of tDCS on working memory in older adults reported increased functional connectivity in the group receiving active anodal stimulation compared to the sham stimulation group during an *n*-back task (Nissim et al., 2019). Despite the increase in functional connectivity in the anodal group, no significant differences in performance were noted on the *n*-back task (Nissim et al., 2019).

Age and disease status may play a pivotal role in tDCS outcomes, including aging-related cognitive disorders. A meta-analysis conducted by Hsu et al. (2015) examined the effects of non-invasive brain stimulation, including tDCS, on cognitive function in healthy older adults and those with Alzheimer's dementia. A small effect size was reported in healthy older adults, and a large effect size was found in older adults with Alzheimer's (Hsu et al., 2015). Similar results in healthy older adults were reported by Summers et al. (2016) with a moderate effect size. When examining effect sizes obtained across studies, there appears to be a trend of tDCS augmenting performance to a greater degree in those with lower cognitive functioning. That is, older adults with cognitive impairment seem to receive a greater benefit than healthy older adults, who in turn receive a greater benefit than young healthy adults (Hsu et al., 2015; Mancuso et al., 2016; Summers et al., 2016; Nissim et al., 2019). This finding should be interpreted with caution, however, as methodological and population variability is present across studies included within the published literature.

Cerebrovascular Perfusion Changes Across Aging and tDCS Considerations

In addition to neuro-cognitive modulation, tDCS may invoke cerebroperfusional modulation associated with cortical hemodynamic functions (Zheng et al., 2011; Takai et al., 2016; Quinn et al., 2020). However, the interaction between tDCS induced effects on cognition and cerebral perfusion across aging remains widely unknown. Post-tDCS cerebral perfusion changes have been measured using neuroimaging techniques such as functional magnetic resonance imaging (fMRI) (Antal et al., 2011) and functional near-infrared spectroscopy (fNIRS) (Patel et al., 2020). Widespread decreases in cerebral perfusion after cathodal and anodal tDCS have been reported using arterial spin labeling (Stagg et al., 2013). Furthermore, regional decreases in blood-oxygen-level-dependent signals have been reported beyond, but not within, the region of stimulation (Antal et al., 2011). Regarding fNIRS, significant interindividual and methodological variability on reported tDCS effects exists in tDCS-fNIRS study designs (Patel et al., 2020). However, increases in cortical activation are reported during resting state; interestingly, a decreased level of cortical activation has also been reported during online tasks (Patel et al., 2020).

Changes in cerebral blood flow and cerebrovascular structure such as plaque formation, rarefaction, and vascular-wall connectivity appear to be aging dependent [see Sonntag et al. (2019) for an overview]. Moreover, disorders impacting both systemic and cerebral vasculature are associated with pathological age-related cognitive decline (Gasecki et al., 2013; Hardigan et al., 2016; Iadecola and Gottesman, 2019). Current evidence suggests a decrease in cerebral blood flow occurs in individuals with MCI beyond the extent of normal cognitive aging (De Eulate et al., 2017; Leeuwis et al., 2018; McKetton et al., 2019; Kim et al., 2020), yet it remains unclear whether this is an accompanying or a causal factor. Consequently, normal and pathological vascular changes may impact tDCS-evoked neuromodulation and cerebral perfusion modulation in older adults relative to young adults. Ultimately, when considering the potential effects of tDCS on cognitive performance and cerebral perfusion, different responses may occur across age and disease status.

It is important to consider structures and mechanisms beyond the neuron and their potential impacts on cognition, such as the neurovascular unit. The neurovascular unit comprises a dynamic interaction between the neuron, vasculature, and glial cells (Iadecola, 2004); the mechanism in which tDCS directly acts upon the neurovascular unit beyond the neuron itself remains unclear. Applied stimulation appears to alter vessel diameter to accommodate for the regional increase in neuronal metabolism (Iadecola et al., 1997). tDCS may also alter astrocytic mediated responses resulting in downstream vascular responses (LeMaistre Stobart et al., 2013). tDCS induced perfusional modulation occurs across cortical and subcortical structures (Stagg et al., 2013). Thus, perfusion changes may underlie behavioral-induced tDCS effects (Stagg et al., 2013), potentially through neurovascular coupling.

Investigating the interaction of cerebral perfusion and cognition, total cerebral blood flow appears to decrease across healthy aging. In an investigation of cerebral perfusion and cognitive aging, Catchlove et al. (2018) report a cerebral blood flow difference of roughly 84.15 mL min^{-1} between the younger and older adult groups. Interestingly, the investigators reported an interaction between total cerebral blood flow and attention in older adults, but not in younger adults. This interaction between cognitive performance and cerebral blood flow in older adults demonstrates an unexpected inverse relationship, with increased performance associated with a decrease in cerebral blood flow, potentially suggesting higher neural efficiency mechanisms (Catchlove et al., 2018).

There appears to be a trend toward declining cerebral blood flow in older adults with pathological cognitive impairment. Kitagawa et al. (2009) report a statistically significant lower cerebral blood volume in older adults with cognitive impairment compared to cognitively healthy age-matched controls. In addition to certain subcortical structures, significant differences in frontal, temporal, parietal, and occipital cortices were all present between groups differing in cognitive status (Kitagawa et al., 2009). Similarly, significantly lower cerebral blood flow was reported in older adults with Alzheimer's dementia compared to those with subjective cognitive impairment (Leijenaar et al., 2017).

Again, a general trend may be arising from the literature, suggesting that the greatest tDCS modulation of cerebral blood flow occurs in healthy young adults, followed by healthy older adults, and finally older adults with cognitive impairment. Note, this is in the opposite direction of the previously hypothesized trend of tDCS impacting behavioral performance to a greater degree in those with cognitive impairments. To summarize, the neurophysiological mechanisms of tDCS may act downstream on the neurovascular unit. When tDCS is applied, both neuronal and perfusional modulation occurs. As vasodilation results in a localized influx of blood, these perfusional changes may be quantified using fNIRS.

Functional Near-Infrared Spectroscopy

fNIRS is a novel functional neuroimaging technique that utilizes near-infrared light to measure hemoglobin chromophores (oxyhemoglobin; HbO, deoxyhemoglobin; HbR, and total hemoglobin; HbT) (Wilcox and Biondi, 2015). Concentrations of each chromophore can be calculated by applying the measured optical properties in a modified Beer-Lambert equation (Wilcox and Biondi, 2015). Under normal circumstances, cortical activation increases oxyhemoglobin concentration with an associated decrease in deoxyhemoglobin concentration (Wilcox and Biondi, 2015). These concentrations can quantify local perfusion changes within the first few centimeters of the brain cortex and has been previously correlated with fMRI BOLD signals (Huppert et al., 2006). fNIRS has been used increasingly within cognitive neuroscience research, and signal responses are sensitive to both cognitive load and cognitive state (Fishburn et al., 2014). As fNIRS primarily measures the superficial cerebral structures composed of gray matter (Quaresima et al., 2005;

Bigio and Fantini, 2016) it can be a useful neuroimaging tool for examining the effects of tDCS.

fNIRS has several advantages over other neuroimaging methods. fNIRS devices tend to be more cost-efficient than an fMRI or EEG, user-friendly, and increasingly portable (with lightweight wireless options that can pair over Bluetooth). fNIRS is advantageous in that it can control for movement and be applied to individuals who have contraindications for MRI (Obrig, 2014; Almajidy et al., 2020), and may be better tolerated by older adults (Stephens and Berryhill, 2016). While the temporal resolution is significantly higher than fMRI, spatial resolution is limited to the superficial layers of the cortex (Obrig, 2014; Almajidy et al., 2020). Given this expanding area of research, further discussion regarding the utility of fNIRS in cognitive paradigms as a function of aging is required.

Purpose

Previous studies have successfully utilized fMRI with tDCS during cognitive tasks, though only a handful have implemented fNIRS with tDCS [see Patel et al. (2020) for a review]. As methodological and perfusional considerations differ between fNIRS protocols and other types of neuroimaging, this study will solely review tDCS-fNIRS protocols targeting cognition. Specifically, the purpose of this systematic review is to explore the neuromodulatory effects of tDCS delivery on cognitive performance and oxygen hemodynamics. Furthermore, the variable of age will be explored across reported metrics. The proposed research questions are as follows:

1. *Does tDCS alter cognitive performance and regional oxygenation during cognitive tasks as measured by fNIRS?*
2. *Does aging impact the efficacy of tDCS on cognitive performance and fNIRS signals?*

Based on the literature, it is hypothesized that tDCS effects on cognitive performance will be greater in older adults compared to younger adults. Regarding fNIRS metrics, we hypothesize young adults will experience greater perfusional change than older adults due to decreasing cerebral blood flow rates in aging.

METHODS

Search Strategy

Electronic searches were conducted using the following databases: CINAHL, Embase, Medline, PsychInfo, Pubmed, Scopus, and Web of Science using Boolean operators in consultation with a research librarian. Search terms included (transcranial direct current stimulation OR tDCS) AND (near-infrared spectroscopy OR functional near-infrared spectroscopy OR fNIRS). This search method resulted in all available tDCS and fNIRS articles; cognitive-orientated studies were then manually extracted. Database searches were conducted on February 19, 2020, and updated on December 27, 2020. No date restrictions were placed on the literature search. Compiled results were imported into Covidence (Covidence Systematic Review Software, Veritas Health Innovation, Melbourne, Australia), where inclusion and exclusion criteria were applied.

Inclusion and Exclusion Criteria

Full-text journal articles published in English were included if they applied tDCS (either concurrent or sequential) and fNIRS to a cognitive paradigm. Non-cognitive study protocols (such as motor function) and review articles were excluded. Further, articles were included if they reported baseline and post-tDCS stimulation metrics on both cognitive performance and recorded fNIRS signals. To compare the efficacy of tDCS, studies were included if they reported a control (sham) and treatment group, or a crossover design study. No restrictions were placed on tDCS type, duration, current intensity, or time of stimulation. Other non-invasive brain stimulation methods such as transcranial magnetic stimulation and transcranial alternating current stimulation were excluded as physiological effects may differ from tDCS. Within this review focusing on cognition, articles reporting healthy adults, or older adults with MCI or dementia were included, with no boundaries on age limits. All other medical diagnoses and mental health disorders were excluded. If studies reported additional metrics in addition to a cognitive paradigm, only the reported interaction between tDCS on performance and fNIRS recordings within the context of the cognitive domain was included within the analysis.

Quality Assessment

Each article was reviewed and underwent quality appraisal by two independent reviewers. Six articles were found in the initial search, and two additional articles were included in the updated literature search. Appraisal checklists were selected according to study design using The Joanna Briggs Institute Critical Appraisal Checklist for Randomized Controlled Trials (Joanna Briggs Institute, 2017) or the Ding et al. (2015) checklist for crossover design. Traditional quality appraisal tools may bias crossover research designs, hence to minimize bias, the proposed checklist outlined in Ding et al. (2015) was applied. Quality assessment tools for other study designs were not required for the final selection of articles due to a relative homogeneity in study designs. Discrepancies in the quality assessment were discussed and resolved. Scores were assigned to each study according to checklist criteria to allow for comparison. Fleiss's kappa was calculated in SPSS Version 26 (IBM Corporation, Armonk, NY, USA) to determine the initial inter-reliability between the reviewers.

Meta-Analysis

Appropriate statistical values for effect size calculations (including: means, medians, standard deviations, standard errors, *p*-values, *F*-Values, and regression coefficients) in addition to sample sizes were extracted from the identified articles. Data was extrapolated from reported figures when necessary. Cohen's *d* effect sizes were calculated for the changes in cognitive performance and fNIRS signals reported within each study. If regression-based beta-estimates were reported without an r value, an estimated r value was calculated using the criteria outlined by Peterson and Brown (Peterson and Brown, 2005). This imputed r value was then utilized within the conventional effect size analysis outlined by Cohen (Cohen, 1988). Effect sizes

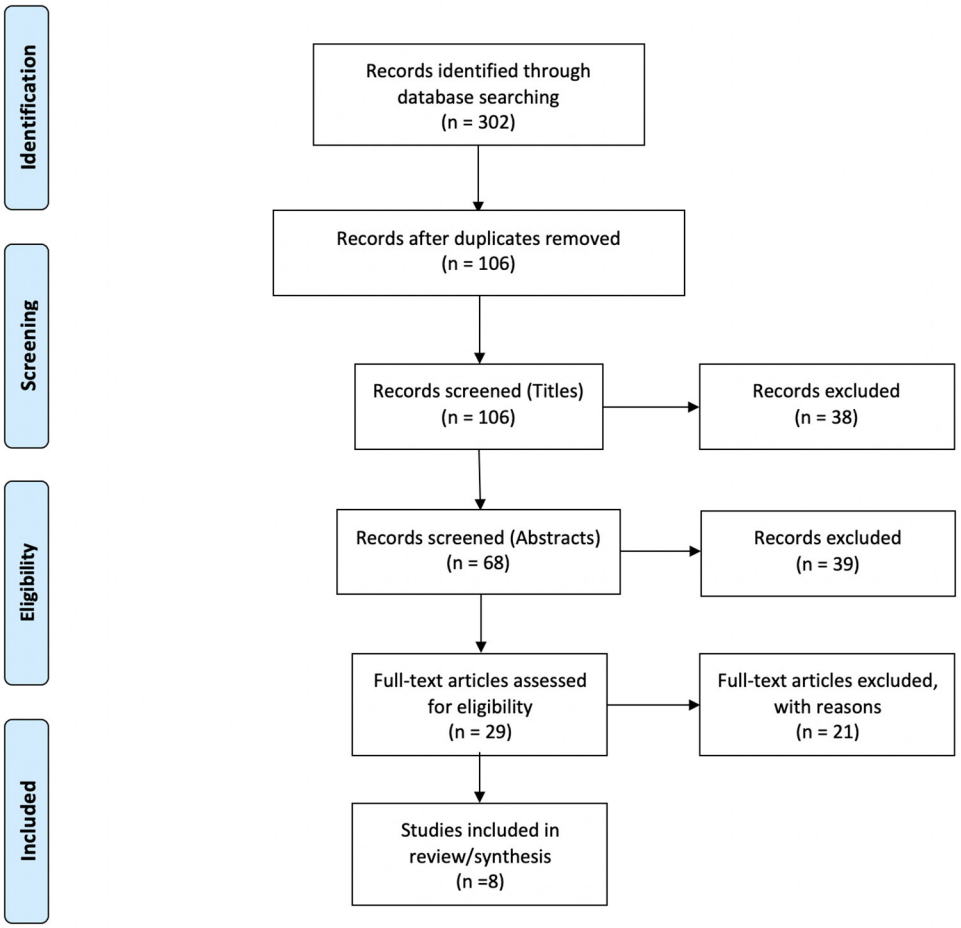


FIGURE 1 | PRISMA flow diagram.

were interpreted as: small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$).

These effect sizes were then imported into Stata (Version 16; StataCorp, College Station, TX, USA) to further process and run the meta-analysis. To investigate the variable of age, a subgroup meta-analysis was performed. A random-effects model using restricted maximum likelihood was utilized to conduct the meta-analysis. REML minimizes bias while reducing mean squared error compared to other meta-analysis approaches (Langan et al., 2019). It should be noted that with the small number of studies present with varying protocols, a high level of heterogeneity is suspected. We will report overall heterogeneity I^2 statistics, however, REML derived point-heterogeneity in limited meta-analysis sample sizes should be interpreted with caution and reported with confidence intervals (Von Hippel, 2015; Langan et al., 2019).

RESULTS

Study Selection

Of the 302 references identified during the initial database search, 196 duplicates were removed. 106 studies were screened, 29 of which underwent full-text review. Twenty-one articles were excluded for the following reasons: lacking a cognitive protocol ($n = 9$), wrong patient population of interest ($n = 4$), not an empirical research study ($n = 4$), lacking a fNIRS protocol ($n = 2$), lacking application of tDCS ($n = 1$), and lacking cognitive task measures with fNIRS ($n = 1$) resulting in eight studies suitable to be included within the review (Jones et al., 2015; Choe et al., 2016; Ehlis et al., 2016; Stephens and Berryhill, 2016; Herrmann et al., 2017; Borragán et al., 2018; Di Rosa et al., 2019; McKendrick et al., 2020). Please refer to the PRISMA diagram in **Figure 1** for details. **Table 1** describes the participant demographics across all included studies.

Quality Assessment

Quality scores ranged widely depending on the appraisal tool used. Four articles were appraised using the Ding et al. (2015) crossover study checklist, and each had a total score of 3/9, though the scoring of individual items varied (see **Table 2**) (Jones et al., 2015; Ehlis et al., 2016; Borragán et al., 2018; Di Rosa et al., 2019). Four articles were appraised using the JBI Critical Appraisal Checklist for Randomized Controlled Trials (Joanna Briggs Institute, 2017) with a mean score of 10/13 (Choe

TABLE 1 | Characteristics of participants.

Reference	Population	Exclusion criteria	Sex	Mean age (SD if reported)	Mean education (SD)
Borragán et al., 2018	Healthy young adults ($n = 22$; 20 in final sample)	Poor sleep quality; moderate-high usual CF (cognitive fatigue), excessive sleepiness, excessive anxiety/depression	8M, 14F	23 (2.28)	NR
Di Rosa et al., 2019	Healthy older adults (60–80 years) ($n = 24$; 21 in final sample) (*Experiment 1)	History of neurological/psychiatric illness, contraindications to tDCS, left-handed	9M, 12F	69.7 (5.1)	14.1 (3.3) years
Ehlis et al., 2016	Healthy young adults; (Group 1 $n = 23$; Group 2 $n = 23$)	Left-handed, history of mental/neurologic disorders, contraindications to tDCS	1: 9M, 14F 2: 12M, 11F	1: 32.1 (10.5) 2: 24.3 (2.4)	NR
Jones et al., 2015	Healthy young adults ($n = 24$) (*Experiment 1)	Neurological/psychiatric symptoms or head injuries; medications	12M, 12F	23.8 (3.7)	NR; University students
Herrmann et al., 2017	Healthy young adults ($n = 61$)	Mental, neurological, or psychiatric illness; current use of psychopharmaceuticals, contraindications to tDCS	31M, 30F	24.3	NR; 55 College students; 6 with 10 years of education
Stephens and Berryhill, 2016	Healthy older adults ($n = 90$; 30 in each group Sham, Active1–1 mA, Active2−2 mA)	Neurologic/psychiatric diseases, contraindications to tDCS, seizure disorders, medications, MMSE < 22	Sham: 14M, 16F Active1: 14M, 16F Active2: 13M, 17F	Sham: 69.9 Active1: 68.6 Active2: 68.6	Sham: 15.2 years Active1: 15.8 years Active2: 15.7 years
Choe et al., 2016	Healthy adults ($n = 32$) DLPFC Active: $n = 7$ DLPFC Sham: $n = 7$ M1 Active: $n = 10$ M1 Sham: $n = 11$	Poor visual acuity, history of epipetic seizures, history of known neurological disorders, pregnancy (or likely to become pregnant during the study)	M: 31 F: 1	DLPFC Stim: 35 (11) DLPFC Sham: 42 (13) M1Stim: 41 (16) M1Sham: 31 (5)	NR
McKendrick et al., 2020	Cognitively healthy young adults (Sham: $n = 10$; Active: $n = 11$)	Current use of psychopharmaceutical agents	M: 10 F: 11	20.3	NR; University students

NR, Not Reported; M, Male; F, Female; MMSE, Mini-Mental State Examination; M, Male; F, Female.

TABLE 2 | Quality assessment—crossover studies.

References	Checklist from Ding et al. (2015) for cross-over studies									Total score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
Borragán et al., 2018	1	−1	0	0	0	1	1	1	0	3/9
Di Rosa et al., 2019	1	−1	0	0	0	1	1	1	0	3/9
Ehlis et al., 2016	1	0	0	0	1	1	0	0	0	3/9
Jones et al., 2015	1	−1	0	0	0	1	1	1	0	3/9
Total item score	4/4	−3/4	0/4	0/4	1/4	4/4	3/4	3/4	0/4	

Each item was scored according to the risk of bias: 1 low risk, 0 unclear, −1 high risk.
(1), Appropriate crossover design; (2), Randomized treatment order; (3), Carryover effect; (4), Unbiased data; (5), Allocation concealment; (6), Blinding; (7), Incomplete outcome data; (8), Selective outcome reporting; (9), Other bias.

TABLE 3 | Quality assessment—randomized control trials.

Reference	JBI critical appraisal checklist for randomized controlled trials (Joanna Briggs Institute, 2017)													Total score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	
Herrmann et al., 2017	1	0	0	1	1	1	1	1	1	1	1	1	1	11/13
Stephens and Berryhill, 2016	1	0	1	1	0	0	1	1	1	1	1	1	1	10/13
Choe et al., 2016	1	0	0	1	1	0	1	1	1	1	1	1	1	10/13
McKendrick et al., 2020	1	0	0	1	0	0	1	1	1	1	1	1	1	9/13
Total item score	4/4	0/4	1/4	4/4	2/4	1/4	4/4	4/4	4/4	4/4	4/4	4/4	4/4	

Each item was scored according to answer: 1 yes, 0 unclear or N/A, −1 no.
(1), Was true randomization used for assignment of participants to treatment groups? (2), Was allocation to treatment groups concealed? (3), Were treatment groups similar at baseline? (4), Were participants blind to treatment assignment? (5), Were those delivering treatment blind to treatment assignment? (6), Were outcomes assessors blind to treatment assignment? (7), Were treatment groups treated identically other than the intervention of interest? (8), Was follow up complete and if not, were differences between groups in terms of their follow up adequately described and analyzed? (9), Were participants analyzed in the groups to which they were randomized? (10), Were outcomes measured in the same way for treatment groups? (11), Were outcomes measured in a reliable way? (12), Was appropriate statistical analysis used? (13), Was the trial design appropriate, and any deviations from the standard RCT design (individual randomization, parallel groups) accounted for in the conduct and analysis of the trial?

et al., 2016; Stephens and Berryhill, 2016; Herrmann et al., 2017; McKendrick et al., 2020). The mean quality percent score of all articles was 57.5% with a range of 33.3–84.6%. Descriptions of the individual items and corresponding scores are described in **Tables 2**, **3**. Inter-rater reliability was considered strong with a Fleiss' κ of 0.851.

Impact of tDCS on Cognitive Task Outcomes

All eight studies reviewed investigated anodal tDCS compared to sham stimulation, with two of these studies also including a cathodal tDCS stimulation condition (Ehlis et al., 2016; Herrmann et al., 2017). Only two articles reported an increase in immediate cognitive performance (Di Rosa et al., 2019; McKendrick et al., 2020). A third study reported no increase in cognitive performance, however, an increase in an untrained task at 1-month follow-up was evident, dependent on dose (i.e., the greatest increase in those receiving 2 mA, followed by 1 mA, compared to sham) (Stephens and Berryhill, 2016). There were no reported effects of tDCS on verbal fluency task performance. The two studies which included older adult participants (Stephens and Berryhill, 2016; Di Rosa et al., 2019) both reported improvements in cognitive performance. Only one of the six studies with young adult participants reported an increase in accuracy and precision on a spatial memory task (McKendrick et al., 2020). tDCS parameters and cognitive effects are presented in **Table 4**.

All eight studies were eligible to be included in the cognitive performance meta-analysis. A moderate level of overall heterogeneity was observed [$I^2 = 50.43\%$, $\chi^2_{(8)} = 19.06$, $p = 0.01$]. An overall effect size for tDCS effects on cognitive performance of $d = 0.26$ (95% CI: −0.03 to 0.55, $p = 0.077$) was obtained. A non-significant trend-wise decrease in the effects of tDCS on cognition was seen in the pooled effect sizes of tDCS as age increased. **Figure 2** provides a summary of the calculated tDCS effect sizes on cognitive performance.

Impact of tDCS on fNIRS Outcomes

Studies differed in reported fNIRS measures (HbO, HbR, HbT, and calculated oxygenation metrics). Within the context of a cognitive task, three studies reported no effects of anodal tDCS on HbO (Choe et al., 2016; Stephens and Berryhill, 2016; Herrmann et al., 2017). Three studies reported an increase in HbO signals following anodal tDCS (Jones et al., 2015; Ehlis et al., 2016; Di Rosa et al., 2019), one study reported a trend-wise decrease in HbO signals following cathodal stimulation (Ehlis et al., 2016), and another study reported no cathodal tDCS effects (Herrmann et al., 2017). When considering HbR, one study reported an increase in HbR concentration within the frontotemporal cortex following anodal stimulation (Herrmann et al., 2017). Hemispheric differences were reported in two studies (Borragán et al., 2018; Di Rosa et al., 2019). Lastly, when examining oxygenation-derived values from HbO and HbR signals, two articles report decreases in regional oxygenation hemodynamic responses with anodal stimulation compared to sham (Borragán et al., 2018; McKendrick et al., 2020), fNIRS parameters are highlighted in **Table 5** below.

All eight articles were eligible for inclusion within the fNIRS meta-analysis. A moderate level of heterogeneity remained present when examining the overall effects of tDCS on obtained fNIRS signals [$I^2 = 44.63\%$, $\chi^2_{(8)} = 13.71$, $p = 0.09$]. Effect sizes were calculated, however consideration of the signal directionality (i.e., if the effect size corresponds to an increase or decrease of an fNIRS signal) in the overall meta-analysis model was not taken into account. An overall effect size of $d = 0.63$ (95% CI: 0.32–0.94, $p < 0.001$) was obtained. Further, a statistically significant effect size of $d = 0.82$ (95% CI: 0.48–1.16, $p < 0.001$) was present in young adults, whereas non-significant effect sizes of 0.48 (95% CI: −0.47 to 1.43) and 0.53 (95% CI: −0.28 to 1.34) were determined in the middle-aged adult and older-aged adult groups, respectively. **Figure 3** provides a forest plot of the included studies and their respective calculated effect sizes.

TABLE 4 | tDCS parameters and effects on cognition.

Reference	Montage	Participant grouping	Age (SD)	# tDCS sessions	Active tDCS parameters	Region stimulated	tDCS administration (Online/offline to cognitive task)	Significant changes in cognitive performance?
Borragán et al., 2018	Anodal/Sham	Within Subject	23 (2.28)	1 Active/1 Sham	1.5 mA for 25 min	Anode: left dorsolateral prefrontal cortex (F3) cathode: right forearm	Online	No
Di Rosa et al., 2019	Anodal/Sham	Within Subject	69.7 (5.1)	1 Active/1 Sham	1.5 mA for 26 min	Left PFC between F3 and F7; reference on contralateral shoulder	Online	Yes: Anodal tDCS with reward motivation increased WM performance (Baseline WM as a modulator)
Ehlis et al., 2016	Anodal/Sham, Cathodal/Sham	Within Subject	(1) 32.1 (10.5) (2) 24.3 (2.4)	1 Active/1 Sham	1 mA for 20 min	Broca's area (between C3, F3, F7); reference on contralateral supraorbital region	Offline to VFT	No
Jones et al., 2015	Anodal/Sham	Within Subject	23.8 (3.7)	1 Active/1 Sham	1.5 mA for 10 min	Anode over left prefrontal cortex (between F3 and F7); cathode over the contralateral cheek	Offline	No
Herrmann et al., 2017	Anodal, Sham, Cathodal	Between Group	24.3 (NR)	1	1.5 mA for 26 min	Bilateral Prefrontal Cortex	Online	No
Stephens and Berryhill, 2016	Anodal/Sham	Between Group	Sham: 69.9 (NR) Active2: 68.6 (NR) Active2: 68.6 (NR)	5	1 or 2 mA (two separate groups) for 15 min	Anode over F4; reference on the contralateral cheek	Offline (tDCS was paired with WM training)	*n*-back, No significant differences, however, a trend was seen in the Active2 group of increased benefit *2 mA tDCS did significantly increase far transfer tasks after 1 month
Choe et al., 2016	Anodal/Sham	Between Group	DLPFC Stim: 35 (11) DLPFC Sham: 42 (13) M1Stim: 41 (16) M1Sham: 31 (5)	4	2 mA for 60 min	Right dorsolateral prefrontal cortex: Anodes: F6 and FC6; Cathodes: Fp2, AF4, and AF8 Left Motor Cortex: Anodes: CP1 and CP3; Cathodes: Fp1, F9, F8	Online (motor finger-tapping task done prior)	*n*-Back, No significant differences between DLPFC stimulation condition as well as M1 stimulation conditions on accuracy *Reduced variability within individual learning rates with DLPFC stimulation, however, the trend appears to be minimal with M1 stimulation.
McKendrick et al., 2020	Anodal/Sham	Within Subject and Between Group	20.3 (NR)	2 Control: Sham & Sham Active: Sham & Active	1 mA for 15 min	Right ventrolateral prefrontal cortex: Anode over F10; cathode over F2	Online	Yes: Anodal tDCS increased spatial memory task performance

NR, Not Reported; WM, Working Memory; VFT, Verbal Fluency Task; DLPFC, Dorsolateral prefrontal cortex.

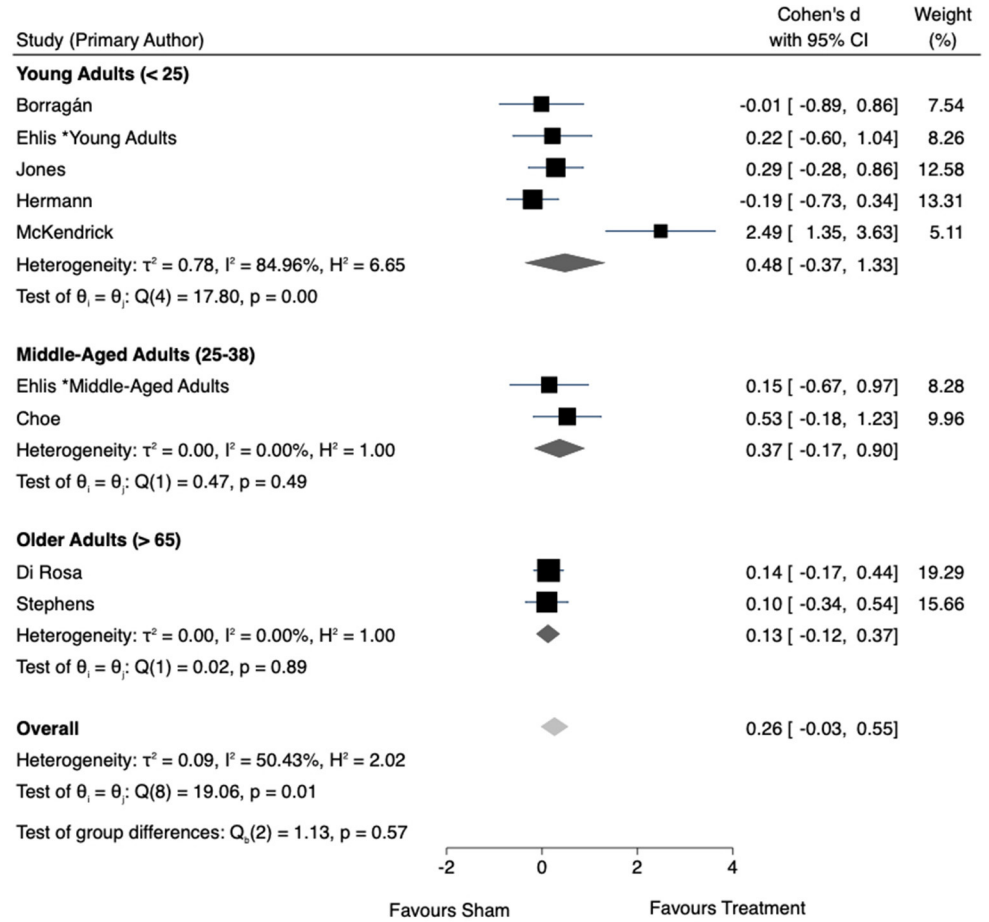


FIGURE 2 | tDCS effects on cognitive performance by age.

DISCUSSION

In this systematic review, we explored the effects of tDCS on cognitive performance and fNIRS-based hemodynamics. A secondary question explored how these measures are affected by aging. The studies reviewed included RCTs ($n = 4$) and within-subject crossover designs ($n = 4$). Four studies included young adults (mean age < 25) (Jones et al., 2015; Herrmann et al., 2017; Borragán et al., 2018; McKendrick et al., 2020), two included older adults (mean age > 65 years old) (Stephens and Berryhill, 2016; Di Rosa et al., 2019), one included middle-aged adults (mean age between 25 and 38 years old) (Choe et al., 2016), and one study had both a young-adult and middle-adult group as participants (Ehlis et al., 2016). Based on the studies included in this review, tDCS does have an impact on cognitive performance and cerebral hemodynamics, as measured by fNIRS metrics. Further, as expected, aging processes appeared to alter the effectiveness of tDCS applications.

Five studies, all of which included young adults, reported no cognitive performance gains following anodal stimulation when compared to sham. Interestingly, in the subgroup meta-analysis, the pooled effect size was greatest in young adults under the age of 25 ($d = 0.48$), followed by middle-aged adults aged 25–38 ($d = 0.37$), and older adults over 65 ($d = 0.13$). This trend was in the opposite direction from our initial hypothesis, which was based on previous reports of tDCS effects being greater in studies with older or cognitively impaired participants (Hsu et al., 2015; Summers et al., 2016; Ke et al., 2019; Nissim et al., 2019). Nonetheless, there are other reports of aging-related resistance to tDCS effects. For instance, Leach et al. (2019) reported tDCS-evoked cognitive gains in associative memory in young adults, which was absent in older adults in the same study. This is further in line with a previous tDCS meta-analysis specific to older adults, where no significant gains were reported in any cognitive domain (Horvath et al., 2015). Yet others have proposed that factors such as baseline performance or education level, as opposed to age, may modulate tDCS efficacy in older adults (Berryhill and Jones, 2012; Learmonth et al., 2015). Clearly, this is an area that warrants further study, and may even require tDCS protocols that are adapted to address the structural and

TABLE 5 | fNIRS parameters.

Reference	fNIRS optode placement	Concurrent /sequential to tDCS	Signals reported	Recording parameters	Signal processing and analysis	Cognitive task measured with fNIRS
Borragán et al., 2018	Bilateral Superior Frontal Cortex	Concurrent	COE (HbR-HbO)	Channels: 24 channels SDD: 3 cm Λ: 685 and 830 nm Sampling rate: 20 Hz Other: Triggered to event onset/offset of TloadDback task	Software: HomER Filter: Low pass (0.009–0.08 Hz) Analysis: Grand averaging of COE by 4 min blocks, ANOVA	TLoadDBack
Di Rosa et al., 2019	Inferior and Midfrontal Gyri, Supplementary motor area, intraparietal sulcus	Concurrent	HbO, HbR	Channels: 4 laser diodes and 8 photo-multiplier tubes. 38 channels, 2 short channels Λ: 690 nm and 83 nm SDD: 3 cm, Short channels: 0.8 cm Sampling rate: 7.8 Hz	Software: HomER2 Filter: Band pass filter (0.01 and 3 Hz); Corrections: Removal of signal-noise ratio <2 and motion artifacts. Age-dependent DPF. Consolidation: GLM approach of hemodynamic modeling with Gaussian functions. Mean HbO, mean HbR, mean hemodynamic responses in interval 5–11 s after stimulus onset. Analysis: ROI, ANOVA	Visuospatial WM task, reward incentives
Ehlis et al., 2016	Bilateral frontotemporal regions	Sequential	HbO, HbR	Channels: 44 channels (2 × 22) in two 3×5 optode arrays. Λ: 695 ± 20 nm and 830 ± 20 nm Sampling rate: 10 Hz	Software: MATLAB Filter: Low pass (0.3 Hz) Corrections: Linear fit function (10 s baseline, last 10 s of rest), noise correction by interpolation of mean adjacent channel signals Analysis: Means of the last 20 s of individual averaged activation was calculated (across each individual, condition, tDCS stimulation session, and channel). Channel wise t-maps, ROI Analysis, ANOVA	Verbal Fluency Test
Jones et al., 2015	Left prefrontal cortex	Sequential	HbO	Channels: 3 channels Λ: 690 and 830 nm SDD: 2.6 cm Sampling rate: 50 Hz	Software: HomeER2 Filter: Low pass filter (0.5 Hz) Corrections: Removal of first 5 s of each 25 s block and motion artifacts. Consolidation: Mean HbO per condition; recorded over final 20 s of each 25 s block. Normalization of HbO difference scores. Analysis: ANOVA	WM Change Detection Task

(Continued)

TABLE 5 | Continued

Reference	fNIRS optode placement	Concurrent /sequential to tDCS	Signals reported	Recording parameters	Signal processing and analysis	Cognitive task measured with fNIRS
Herrmann et al., 2017	Bilateral prefrontal cortices	Concurrent	HbO, HbR	Channels: 52; Three rows (each with 11 optodes, SSD 3 cm). 33 optodes (17 laser diodes and 16 photodetectors) SDD: 3 cm Sampling rate: 10 Hz	Software: MATLAB Filter: Low pass (0.5 Hz) and discrete cosine filters Corrections: Removal of high-frequency artifacts using 5 s moving average, a common average reference to removing physiological noise, DPF. Analysis: Effect size (baseline to task performance), t-maps, ROI, ANOVA	Verbal Fluency Test
Stephens and Berryhill, 2016	Bilateral prefrontal cortices	Sequential	HbO	Channels: 14 Sampling rate: 50 Hz	Software: HomER2 Filter: Low pass filter (0.5 Hz); Corrections: Removal of motion artifacts. Normalization of each channel Analysis: Peak HbO amplitude per channel standardized per participant across time, transformed into an overall percentage of channels with decrease activation across time.	*n*-Back Task
Choe et al., 2016	M1, Right dorsolateral prefrontal cortex	Concurrent	HbO, HbR, HbT	Channels: 20 channels (10 channels over M1; 10 channels over the right dorsolateral prefrontal cortex). SDD: < 3.5 cm Sampling rate: 8 Hz	Software: nirsLab, SPM Filter: Band-pass filter (0.01 Hz – 0.2 Hz) Corrections: Inter-trial signals removed from time-series. Average baseline concentration subtracted from task-evoked concentration changes Analysis: HbO, HbR, HbT average concentrations ran for each channel, participant, task, and time. Concentrations were averages within time (days) across all *n*-back trials. Concentrations were further region and grouped averaged across the total time difference. General linear model-based SPM was performed, multiple comparison correction of channels.	*n*-Back Task
McKendrick et al., 2020	Bilateral prefrontal cortices (Anterior and dorsolateral prefrontal cortices, Pars Triangularis, Pars Opercularis)	Concurrent	HbO, HbR, Oxygenation (HbO–HbR)	Channels: 16 Λ: 730 and 850 nm SDD: 2.5 cm Sampling rate: 2 Hz	Software: COBI Studio software Filter: Low pass filtered (0.1 Hz) Corrections: Motion artifact assessment Analysis: Temporal hemodynamic function temporally group averaged. Linear mixed effect modeling with restricted maximum likelihood. Bayesian information criterion to determine random and fixed effects. False discover rate corrections.	Spatial memory task

COE, Cerebral Oxygen Exchange; HbO, Oxyhemoglobin; HbR, Deoxyhemoglobin; HbT, Total Hemoglobin; ROI, Region of Interest; ANOVA, Analysis of Variance; SDD, Source-Detector Distance; DPF, Differential Pathlength Factor; Λ, Wavelength.

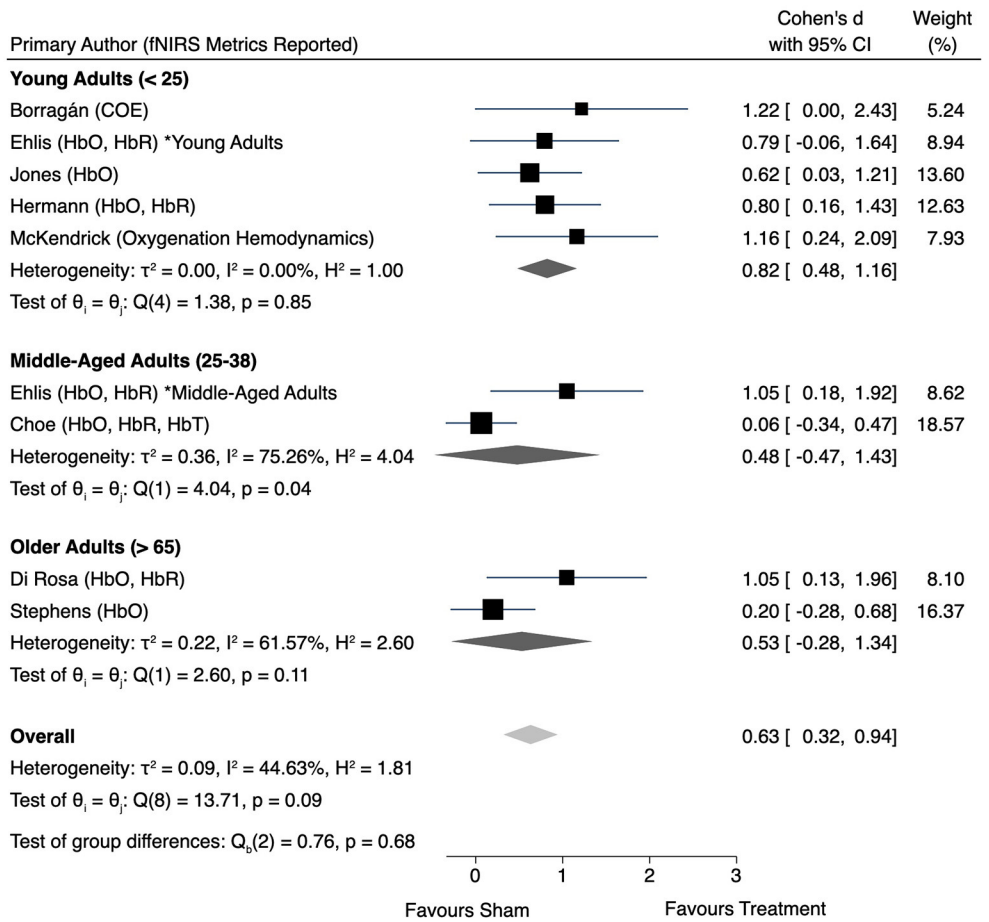


FIGURE 3 | Effect of tDCS on cortical activation measured with fNIRS.

neuroanatomical changes associated with aging brains (Habich et al., 2020).

For the purposes of the specific questions in this review, we included studies that explored the effect of tDCS on some aspect of cognition. Undoubtedly, there was much variability in the cognitive tasks used in the studies, including verbal fluency tasks ($n = 2$) (Ehlis et al., 2016; Herrmann et al., 2017), spatial memory tasks ($n = 1$) (McKendrick et al., 2020), and working memory tasks ($n = 5$) (Jones et al., 2015; Choe et al., 2016; Stephens and Berryhill, 2016; Borragán et al., 2018; Di Rosa et al., 2019). Within this latter category, there was a large amount of procedural variability. One WM task was a modification of the n-back task called T-load D-back, which incorporates both the n-back and a number decision task into one process (Borragán et al., 2018). Another was a novel visuospatial task that required both identification and location memory of pictures and letters (Di Rosa et al., 2019). A third study utilized an operation span task while another conducted a battery of n-back and letter span tasks (Jones et al., 2015). This heterogeneity in the behavioral assessment of WM introduces a potential reason/confound for the variability of tDCS effects. Though not within the scope of this review, two of the included studies further assessed the role of motivation on tDCS efficacy, both of which found that higher motivation via financial incentive augmented behavioral performance to a greater extent in anodal tDCS groups (Jones et al., 2015; Di Rosa et al., 2019). Further, one tDCS and fNIRS study examined additional variables related to flight simulation, however only the cognitive component was included in this review (Choe et al., 2016). It is possible the variation in effect sizes reported in this review is reflective of the differences in cognitive tasks used across the various studies.

The majority of articles reviewed utilized a working memory paradigm as the cognitive measure. The impact of tDCS on enhancing working memory task performance in younger adults has previously been reported (Katsoulaki et al., 2017). However, tDCS effect sizes within the cognitive domain of working memory also appear to differ across adulthood, and in older adults with mild cognitive impairment or dementia (Hsu et al., 2015; Mancuso et al., 2016; Stephens and Berryhill, 2016; Summers et al., 2016; Di Rosa et al., 2019). Although

our search did not yield any studies of tDCS and fNIRS in individuals with cognitive impairments, this is a population in which further study could be illuminative of the impact of tDCS on cognitive performance and cerebral perfusion. Future investigations based on theoretical models of cognitive aging, such as the Hemispheric-Asymmetry Reduction in Older Adults (HAROLD) (Cabeza, 2002), Compensation-Related Utilization of Neural Circuits Hypothesis (CRUNCH) (Reuter-Lorenz and Cappell, 2008), and the Scaffolding Theory of Aging and Cognition (STAC) (Park and Reuter-Lorenz, 2009) may provide useful frameworks for further inquiry.

The studies reviewed lacked a standardized metric of cerebral oxygenation, reflected in the variety of fNIRS signals reported (**Table 5**). Even within studies reporting the same metric however, effects of tDCS on cerebral oxygenation were mixed. For instance, three studies reported increases in HbO following tDCS stimulation (Jones et al., 2015; Ehlis et al., 2016; Di Rosa et al., 2019) while three studies reported no significant changes (Choe et al., 2016; Stephens and Berryhill, 2016; Herrmann et al., 2017). One study reported an increase in HbR following anodal stimulation (Herrmann et al., 2017) and another two studies reported decreases in oxygenation when estimated as a function of HbO and HbR (Borragán et al., 2018; McKendrick et al., 2020). As there is little consensus on the downstream cognitive effects of changes in HbO and HbR concentration, this is an area where future studies may help to further elucidate the mechanisms underlying tDCS-induced cognitive enhancement.

In the studies reviewed, tDCS was found to impact cerebral perfusion as measured by fNIRS, demonstrated by our overall statistically significant moderate effect size of $d = 0.63$. We hypothesized that young adults would exhibit greater perfusional change relative to older adults following tDCS, as measured by fNIRS metrics. This hypothesis was supported by our subgroup analysis. A statistically significant effect size of $d = 0.82$ was present within the younger adults, whereas non-significant effect sizes were reported for middle-aged and older adults. It is possible that the large effect sizes calculated for the studies reporting decreased oxygenation (Borragán et al., 2018; McKendrick et al., 2020) may have skewed the overall effect size, therefore these results should be interpreted with caution due to the limited number of studies and level of heterogeneity present.

The theoretical grounding of this review is based on the premise that the interaction between the neuron (when modulated by tDCS) and associated cerebral perfusion at the neurovascular unit impacts cognitive performance. However, other changes beyond the level of the neurovascular unit, such as cerebral atrophy should be considered. In a study investigating cerebral blood flow changes across aging, Meltzer et al. (2000) noted there were no age-related cerebral perfusion differences using positron emission tomography (PET), after correcting for brain volume. This suggests that cerebral atrophy, not cerebral blood flow, may underlie functional deterioration, Conversely, another study using arterial spin labeling found that cerebral perfusion was significantly correlated with cortical thickness and total brain volume, as well as performance on executive function tasks (Alosco et al., 2013). However, there was no direct association between brain volume and cortical thickness with cognitive function. Another study using PET in participants with hypertension and lacunar infarcts or white matter lesions reported that lower cerebral blood flow precedes cognitive decline 3 years later, measured using the Mini-Mental State Examination tool (Kitagawa et al., 2009). From these findings, it appears possible that cerebral blood flow underlies a common mechanism present in both cognitive decline and cerebral atrophy.

No articles with individuals with MCI or dementia were identified in our search, demonstrating the need for cognitive-based tDCS and fNIRS research protocols with these populations. Significant effects of tDCS on cognitive performance have previously been reported in the literature (Cruz Gonzalez et al., 2018), and there is evidence the effectiveness of non-invasive brain stimulation may vary among older adults with MCI (Chu et al., 2021). Further research is needed to investigate potential age-related changes in cognitive mechanisms to explain this variability.

Limitations and Future Directions

With the limited number of articles suitable for review, studies were grouped by age despite having varying cognitive tasks. Although spatial memory and working memory may represent similar cognitive mechanisms, verbal fluency tasks may be grounded in an alternative cognitive domain altogether. The studies using verbal fluency tasks were conducted in younger adults, which potentially impacted the effect sizes reported (**Figure 2**). Nonetheless, studies employing working memory tasks were included across all subgroups included in effect size calculations. With ongoing research in the field, it is recommended that future reviews conduct an analysis accounting for the different cognitive tasks utilized in addition to age.

Research investigating aging-related differences in tDCS and cognition as it relates to cerebral perfusion yields meaningful insight into the current understanding of these cognitive processes and the ability for neuromodulation. Future directions should also aim to investigate populations with microvascular changes (such as diabetes and chronic hypertension) in addition to larger vascular changes (such as aortic and carotid stenosis), using a cognitive-orientated tDCS and fNIRS paradigm to further assess the role of cerebral blood flow and vascular health in cognitive task performance.

There exists a possibility in which repeated tDCS sessions might induce different physiological changes within and beyond the stimulated brain region, and this should further be assessed within the context of cognitive aging. In addition to tDCS stimulation frequency, the effects of current intensity, time, regions of stimulation, and montage (anodal or cathodal) require further investigation regarding cognitive performance across normal and pathological cognitive aging. tDCS effects and direction of change (i.e., increases or decreases) of specific chromophores (HbO, HbR, HbT) or oxygenation is yet to be determined. With the limited number of cognitive-oriented

tDCS and fNIRS studies, it is recommended that additional studies be conducted before establishing the directionality of these unknown variables in meta-analysis. With interindividual differences, it is recommended to perform electric field modeling using structural neuroimaging of each participant if available to assist in optimizing tDCS parameters and regions of stimulation. Lastly, to the author's knowledge, no widely available graphical user interface or software is available to model the effects of tDCS current on cerebral perfusion, which is an avenue to explore in the future using perfusional neuroimaging methods including fNIRS.

CONCLUSION

With the eight included tDCS and fNIRS studies on cognition, we report significant overall effect sizes on cognitive performance and fNIRS signals due to tDCS-evoked neuromodulation. Further, age-related differences appear to alter the efficacy of tDCS effects. With the limited number of studies and heterogeneity in combined tDCS, fNIRS, and cognitive testing parameters, further research is required to test the efficacy and directionality of fNIRS signals. Confounding variables such as baseline performance, education, health status, and factors impacting cerebral blood flow should further be investigated and included in future study designs. In conclusion, tDCS may be a promising tool for neuromodulation and cerebral perfusion modulation, however, significant research is still needed to determine which groups are more susceptible to tDCS-evoked effects.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

MF: conceptualization, methodology, investigation, validation, formal analysis, resources, data curation, writing–original draft, writing-review and editing, visualization, and project administration. MZ: conceptualization, methodology, validation, formal analysis, resources, data curation, writing–original draft, writing–review and editing, visualization, and project administration. EK: conceptualization, methodology, validation, formal analysis, resources, data curation, writing–original draft, writing–review and editing, visualization, project administration, and supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Liz Dennett for her assistance and consultation during the database search.

REFERENCES

Almajidy, R. K., Mankodiya, K., Abtahi, M., and Hofmann, U. G. (2020). A newcomer's guide to functional near infrared spectroscopy experiments. *IEEE Rev. Biomed. Eng.* 13, 292–308. doi: 10.1109/RBME.2019.2944351

Alosco, M. L., Gunstad, J., Jerskey, B. A., Xu, X., Clark, U. S., Hassenstab, J., et al. (2013). The adverse effects of reduced cerebral perfusion on cognition and brain structure in older adults with cardiovascular disease. *Brain Behav.* 3, 626–636. doi: 10.1002/brb3.171

Antal, A., Polania, R., Schmidt-Samoa, C., Dechent, P., and Paulus, W. (2011). Transcranial direct current stimulation over the primary motor cortex during fMRI. *Neuroimage* 55, 590–596. doi: 10.1016/j.neuroimage.2010.11.085

Berryhill, M. E., and Jones, K. T. (2012). tDCS selectively improves working memory in older adults with more education. *Neurosci. Lett. J.* 521, 148–151. doi: 10.1016/j.neulet.2012.05.074

Berryhill, M. E., and Martin, D. (2018). Cognitive effects of transcranial direct current stimulation in healthy and clinical populations: an overview. *J. ECT* 34, e25–35. doi: 10.1097/YCT.0000000000000534

Bigio, I. J., and Fantini, S. (2016). *Quantitative Biomedical Optics.* Cambridge: Cambridge University Press.

Borragán, G., Gilson, M., Guerrero-Mosquera, C., Di Ricci, E., Slama, H., and Peigneux, P. (2018). Transcranial direct current stimulation does not counteract cognitive fatigue, but induces sleepiness and an inter-hemispheric shift in brain oxygenation. *Front. Psychol.* 9:2351. doi: 10.3389/fpsyg.2018.02351

Boudewyn, M. A., Scangos, K., Ranganath, C., and Carter, C. S. (2020). Using prefrontal transcranial direct current stimulation (tDCS) to enhance proactive cognitive control in schizophrenia. *Neuropsychopharmacology* 45, 1877–1883. doi: 10.1038/s41386-020-0750-8

Butler, M., McCreedy, E., Nelson, V. A., Desai, P., Ratner, E., Fink, H. A., et al. (2018). Does cognitive training prevent cognitive decline? *Ann. Intern. Med.* 168, 63–68. doi: 10.7326/M17-1531

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychol. Aging* 17, 85–100. doi: 10.1037/0882-7974.17.1.85

Cappon, D., Jahanshahi, M., and Bisiacchi, P. (2016). Value and efficacy of transcranial direct current stimulation in the cognitive rehabilitation: a critical review since 2000. *Front. Neurosci.* 10:157. doi: 10.3389/fnins.2016.00157

Catchlove, S. J., Macpherson, H., Hughes, M. E., Chen, Y., Parrish, T. B., and Pipingas, A. (2018). An investigation of cerebral oxygen utilization, blood flow and cognition in healthy aging. *PLoS One* 13:e0197055. doi: 10.1371/journal.pone.0197055

Choe, J., Coffman, B. A., Bergstedt, D. T., Ziegler, M. D., Phillips, M. E., J. C., et al. (2016). Transcranial direct current stimulation modulates neuronal activity and learning in pilot training. *Front. Hum. Neurosci.* 10:34. doi: 10.3389/fnhum.2016.00034

Chou, Y., hui, Ton That, V., and Sundman, M. (2020). A systematic review and meta-analysis of rTMS effects on cognitive enhancement in mild cognitive impairment and Alzheimer's disease. *Neurobiol. Aging* 86, 1–10. doi: 10.1016/j.neurobiolaging.2019.08.020

Chu, C.-S., Li, C.-T., Brunoni, A. R., Yang, F.-C., Tseng, P.-T., Tu, Y.-K., et al. (2021). Cognitive effects and acceptability of non-invasive brain stimulation on Alzheimer's disease and mild cognitive impairment: a component network meta-analysis Neurodegeneration. *J. Neurol. Neurosurg. Psychiatry* 92, 195–203. doi: 10.1136/jnnp-2020-323870

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences Second Edition*. New York, NY: Lawrence Erlbaum Associates.

Cruz Gonzalez, P., Fong, K. N. K., Chung, R. C. K., Ting, K. H., Law, L. L. F., and Brown, T. (2018). Can transcranial direct-current stimulation alone or combined with cognitive training be used as a clinical intervention to

improve cognitive functioning in persons with mild cognitive impairment and dementia? A systematic review and meta-analysis. *Front. Hum. Neurosci.* 12:416. doi: 10.3389/fnhum.2018.00416

Das, N., Spence, J. S., Aslan, S., Vanneste, S., Mudar, R., Rackley, A., et al. (2019). Cognitive training and transcranial direct current stimulation in mild cognitive impairment: a randomized pilot trial. *Front. Neurosci.* 13:307. doi: 10.3389/fnins.2019.00307

De Eulate, R. G., Goñi, I., Galiano, A., Vidorreta, M., Recio, M., Riverol, M., et al. (2017). Reduced cerebral blood flow in mild cognitive impairment assessed using phase-contrast MRI. *J. Alzheimers Dis.* 58, 585–595. doi: 10.3233/JAD-161222

Di Rosa, E., Brigadoi, S., Cutini, S., Tarantino, V., Dell'Acqua, R., Mapelli, D., et al. (2019). Reward motivation and neurostimulation interact to improve working memory performance in healthy older adults: a simultaneous tDCS-fNIRS study. *Neuroimage* 202:116062. doi: 10.1016/j.neuroimage.2019.116062

Dickstein, D. L., Weaver, C. M., Luebke, J. I., and Hof, P. R. (2013). Dendritic spine changes associated with normal aging. *Neuroscience* 251, 21–32. doi: 10.1016/j.neuroscience.2012.09.077

Ding, H., Hu, G. L., Zheng, X. Y., Chen, Q., Threapleton, D. E., and Zhou, Z. H. (2015). The method quality of cross-over studies involved in Cochrane Systematic Reviews. *PLoS One* 10:e120519. doi: 10.1371/journal.pone.0120519

Ehlis, A.-C., Haeussinger, F. B., Gastel, A., Fallgatter, A. J., and Plewnia, C. (2016). Task-dependent and polarity-specific effects of prefrontal transcranial direct current stimulation on cortical activation during word fluency. *Neuroimage* 140, 134–140. doi: 10.1016/j.neuroimage.2015.12.047

Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Front. Hum. Neurosci.* 8:76. doi: 10.3389/fnhum.2014.00076

Flöel, A., Rösser, N., Michka, O., Knecht, S., and Breitenstein, C. (2008). Noninvasive brain stimulation improves language learning. *J. Cogn. Neurosci.* 20, 1415–1422. doi: 10.1162/jocn.2008.20098

Gasecki, D., Kwarciany, M., Nyka, W., and Narkiewicz, K. (2013). Hypertension, brain damage and cognitive decline. *Curr. Hypertens. Rep.* 15, 547–558. doi: 10.1007/s11906-013-0398-4

Habich, A., Slotboom, J., Peter, J., Wiest, R., and Klöppel, S. (2020). No effect of anodal tDCS on verbal episodic memory performance and neurotransmitter levels in young and elderly participants. *Neural Plast.* 2020:8896791. doi: 10.1155/2020/8896791

Hardigan, T., Ward, R., and Ergul, A. (2016). Cerebrovascular complications of diabetes: focus on cognitive dysfunction. *Clin. Sci.* 30, 1807–1822. doi: 10.1042/CS20160397

Herrmann, M. J., Horst, A. K., Loble, S., Moll, M. T., Katzorke, A., and Polak, T. (2017). Relevance of dorsolateral and frontotemporal cortex on the phonemic verbal fluency - a fNIRS-study. *Neuroscience* 367, 169–177. doi: 10.1016/j.neuroscience.2017.10.028

Horvath, J. C., Forte, J. D., and Carter, O. (2015). Quantitative review finds no evidence of cognitive effects in healthy populations from single-session transcranial direct current stimulation (tDCS). *Brain Stimul.* 8, 535–550. doi: 10.1016/j.brs.2015.01.400

Hsu, W. Y., Ku, Y., Zanto, T. P., and Gazzaley, A. (2015). Effects of noninvasive brain stimulation on cognitive function in healthy aging and Alzheimer's disease: a systematic review and meta-analysis. *Neurobiol. Aging* 36, 2348–2359. doi: 10.1016/j.neurobiolaging.2015.04.016

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 368–382. doi: 10.1016/j.neuroimage.2005.08.065

Iadecola, C. (2004). Neurovascular regulation in the normal brain and in Alzheimer's disease. *Nat. Rev. Neurosci.* 5, 347–360. doi: 10.1038/nrn1387

Iadecola, C., and Gottesman, R. F. (2019). Neurovascular and cognitive dysfunction in hypertension: epidemiology, pathobiology, and treatment. *Circ. Res.* 124, 1025–1044. doi: 10.1161/CIRCRESAHA.118.313260

Iadecola, C., Yang, G., Ebner, T. J., and Chen, G. (1997). Local and propagated vascular responses evoked by focal synaptic activity in cerebellar cortex. *J. Neurophysiol.* 78, 651–659. doi: 10.1152/jn.1997.78.2.651

Joanna Briggs Institute (2017). *The Joanna Briggs Institute Critical Appraisal Tools for Use in JBI Systematic Reviews*.

Jones, K. T., Gözenman, F., and Berryhill, M. E. (2015). The strategy and motivational influences on the beneficial effect of neurostimulation: a tDCS and fNIRS study. *Neuroimage* 105, 238–247. doi: 10.1016/j.neuroimage.2014.11.012

Katsoulaki, M., Kastrinis, A., and Tsekoura, M. (2017). The effects of anodal transcranial direct current stimulation on working memory. *Adv. Exp. Med. Biol.* 987, 283–289. doi: 10.1007/978-3-319-57379-3_25

Ke, Y., Wang, N., Du, J., Kong, L., Liu, S., Xu, M., et al. (2019). The effects of transcranial direct current stimulation (tDCS) on working memory training in healthy young adults. *Front. Hum. Neurosci.* 13:19. doi: 10.3389/fnhum.2019.00019

Kim, C. M., Alvarado, R. L., Stephens, K., Wey, H. Y., Wang, D. J. J., Leritz, E. C., et al. (2020). Associations between cerebral blood flow and structural and functional brain imaging measures in individuals with neuropsychologically defined mild cognitive impairment. *Neurobiol. Aging* 86, 64–74. doi: 10.1016/j.neurobiolaging.2019.10.023

Kitagawa, K., Oku, N., Kimura, Y., Yagita, Y., Sakaguchi, M., Hatazawa, J., et al. (2009). Relationship between cerebral blood flow and later cognitive decline in hypertensive patients with cerebral small vessel disease. *Hypertens. Res.* 32, 816–820. doi: 10.1038/hr.2009.100

Langan, D., Higgins, J. P. T., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., et al. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res. Synth. Methods* 10, 83–98. doi: 10.1002/jrsm.1316

Leach, R. C., McCurdy, M. P., Trumbo, M. C., Matzen, L. E., and Leshikar, E. D. (2019). Differential age effects of transcranial direct current stimulation on associative memory. *J. Gerontol.* 74, 1163–1173. doi: 10.1093/geronb/gby003

Learmonth, G., Thut, G., Benwell, C. S. Y., and Harvey, M. (2015). The implications of state-dependent tDCS effects in aging: behavioural response is determined by baseline performance. *Neuropsychologia* 74, 108–119. doi: 10.1016/j.neuropsychologia.2015.01.037

Leeuwis, A. E., Smith, L. A., Melbourne, A., Hughes, A. D., Richards, M., Prins, N. D., et al. (2018). Cerebral blood flow and cognitive functioning in a community-based, multi-ethnic cohort: the SABRE study. *Front. Aging Neurosci.* 10:279. doi: 10.3389/fnagi.2018.00279

Leijenaar, J. F., Maurik, I. S., Kuijer, J. P. A., Flier, W. M., Scheltens, P., Barkhof, F., et al. (2017). Lower cerebral blood flow in subjects with Alzheimer's dementia, mild cognitive impairment, and subjective cognitive decline using two-dimensional phase-contrast magnetic resonance imaging. *Alzheimers Dement. (Amst).* 9, 76–83. doi: 10.1016/j.dadm.2017.10.001

LeMaistre Stobart, J. L., Lu, L., Anderson, H. D. I., Mori, H., and Anderson, C. M. (2013). Astrocyte-induced cortical vasodilation is mediated by D-serine and endothelial nitric oxide synthase. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3149–3154. doi: 10.1073/pnas.1215929110

Mancuso, L. E., Ilieva, I. P., Hamilton, R., and Farah, M. J. (2016). Does transcranial direct current stimulation improve healthy working memory?: a meta-analytics review. *J. Cogn. Neurosci.* 28, 1063–1089. doi: 10.1162/jocn_a_00956

Martinotti, G., Lupi, M., Montemitro, C., Miuli, A., Di Natale, C., Spano, M. C., et al. (2019). Transcranial direct current stimulation reduces craving in substance use disorders: a double-blind, placebo-controlled study. *J. ECT.* 35, 207–211. doi: 10.1097/YCT.0000000000000580

Matar, S. J., Sorinola, I. O., Newton, C., and Pavlou, M. (2020). Transcranial direct-current stimulation may improve discourse production in healthy older adults. *Front. Neurol.* 11:935. doi: 10.3389/fneur.2020.00935

McKendrick, R., Falcone, B., Scheldrup, M., and Ayaz, H. (2020). Effects of transcranial direct current stimulation on baseline and slope of prefrontal cortex hemodynamics during a spatial working memory task. *Front. Hum. Neurosci.* 14:64. doi: 10.3389/fnhum.2020.00064

McKetton, L., Cohn, M., Tang-Wai, D. F., Sobczyk, O., Duffin, J., Holmes, K. R., et al. (2019). Cerebrovascular resistance in healthy aging and mild cognitive impairment. *Front. Aging Neurosci.* 11:79. doi: 10.3389/fnagi.2019.00079

Meinzer, M., Lindenberg, R., Phan, M. T., Ulm, L., Volk, C., and Flöel, A. (2015). Transcranial direct current stimulation in mild cognitive impairment: behavioral effects and neural mechanisms. *Alzheimers Dement.* 11, 1032–1040. doi: 10.1016/j.jalz.2014.07.159

Meltzer, C. C., Cantwell, M. N., Greer, P. J., Ben-Eliezer, D., Smith, G., Frank, G., et al. (2000). Does cerebral blood flow decline in healthy aging? A PET study with partial-volume correction. *J. Nucl. Med.* 41, 1842–1848.

Nissim, N. R., O'Shea, A., Indahlastari, A., Telles, R., Richards, L., Porges, E., et al. (2019). Effects of in-scanner bilateral frontal tDCS on functional connectivity of the working memory network in older adults. *Front. Aging Neurosci*. 11:51. doi: 10.3389/fnagi.2019.00051

Nitsche, M. A., Fricke, K., Henschke, U., Schlitterlau, A., Liebetanz, D., Lang, N., et al. (2003). Pharmacological modulation of cortical excitability shifts induced by transcranial direct current stimulation in humans. *J. Physiol*. 553, 293–301. doi: 10.1113/jphysiol.2003.049916

Nitsche, M. A., and Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *J. Physiol*. 527, 633–639. doi: 10.1111/j.1469-7793.2000.t01-1-00633.x

Nitsche, M. A., and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* 57, 1899–1901. doi: 10.1212/WNL.57.10.1899

Obrig, H. (2014). NIRS in clinical neurology - a "promising" tool? *Neuroimage* 85, 535–546. doi: 10.1016/j.neuroimage.2013.03.045

Ohn, S. H., Park, C., Il Yoo, W. K., Ko, M. H., Choi, K. P., Kim, G. M., et al. (2008). Time-dependent effect of transcranial direct current stimulation on the enhancement of working memory. *Neuroreport* 19, 43–47. doi: 10.1097/WNR.0b013e3282f2adfd

Park, D. C., and Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. *Annu. Rev. Psychol*. 60, 173–196. doi: 10.1146/annurev.psych.59.103006.093656

Patel, R., Dawidziuk, A., Darzi, A., Singh, H., and Leff, D. R. (2020). Systematic review of combined functional near-infrared spectroscopy and transcranial direct-current stimulation studies. *Neurophotonics* 7:1. doi: 10.1117/1.NPh.7.2.020901

Paulus, W. (2011). Transcranial electrical stimulation (tES - tDCS tRNS, tACS) methods. *Neuropsychol. Rehabil*. 21, 602–617. doi: 10.1080/09602011.2011.557292

Peterson, R. A., and Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *J. Appl. Psychol*. 90, 175–181. doi: 10.1037/0021-9010.90.1.175

Polanía, R., Nitsche, M. A., and Ruff, C. C. (2018). Studying and modifying brain function with non-invasive brain stimulation. *Nat. Neurosci*. 21, 174–187. doi: 10.1038/s41593-017-0054-4

Prehn, K., and Flöel, A. (2015). Potentials and limits to enhance cognitive functions in healthy and pathological aging by tDCS. *Front. Cell. Neurosci*. 9:355. doi: 10.3389/fncel.2015.00355

Quaresima, V., Ferrari, M., Torricelli, A., Spinelli, L., Pifferi, A., and Cubeddu, R. (2005). Bilateral prefrontal cortex oxygenation responses to a verbal fluency task: a multichannel time-resolved near-infrared topography study. *J. Biomed. Opt*. 10:11012. doi: 10.1117/1.1851512

Quinn, D. K., Upston, J., Jones, T., Brandt, E., Story-Remer, J., Fratzke, V., et al. (2020). Cerebral perfusion effects of cognitive training and transcranial direct current stimulation in mild-moderate TBI. *Front. Neurol*. 11:545174. doi: 10.3389/fneur.2020.545174

Reuter-Lorenz, P. A., and Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Curr. Dir. Psychol. Sci*. 17, 177–182. doi: 10.1111/j.1467-8721.2008.00570.x

Rose, N. S., Thomson, H., and Kliegel, M. (2020). No effect of transcranial direct-current stimulation to dorsolateral prefrontal cortex on naturalistic prospective memory in healthy young and older adults. *J. Cogn. Enhanc*. 4, 211–218. doi: 10.1007/s41465-019-00155-2

Smith, R. C., Boules, S., Mattiuz, S., Youssef, M., Tobe, R. H., Sershen, H., et al. (2015). Effects of transcranial direct current stimulation (tDCS) on cognition, symptoms, and smoking in schizophrenia: a randomized controlled study. *Schizophr. Res*. 168, 260–266. doi: 10.1016/j.schres.2015.06.011

Song, D., Yu, D. S. F., Li, P. W. C., and Lei, Y. (2018). The effectiveness of physical exercise on cognitive and psychological outcomes in individuals with mild cognitive impairment: a systematic review and meta-analysis. *Int. J. Nurs. Stud*. 79, 155–164. doi: 10.1016/j.ijnurstu.2018.01.002

Sonntag, W. E., Eckman, D. M., Ingraham, J., and Riddle, D. R. (2019). "Regulation of Cerebrovascular Aging," in *Brain Aging: Models, Methods, and Mechanisms*, ed D. R. Riddle (Boca Raton, FL: CRC Press; Taylor & Francis).

Stagg, C. J., Lin, R. L., Mezue, M., Segerdahl, A., Kong, Y., Xie, J., et al. (2013). Widespread modulation of cerebral perfusion induced during and after transcranial direct current stimulation applied to the left dorsolateral prefrontal cortex. *J. Neurosci*. 33, 11425–11431. doi: 10.1523/JNEUROSCI.3887-12.2013

Stagg, C. J., and Nitsche, M. A. (2011). Physiological basis of transcranial direct current stimulation. *Neuroscientist* 17, 37–53. doi: 10.1177/1073858410386614

Stephens, J. A., and Berryhill, M. E. (2016). Older adults improve on everyday tasks after working memory training and neurostimulation. *Brain Stimul*. 9, 553–559. doi: 10.1016/j.brs.2016.04.001

Summers, J. J., Kang, N., and Cauraugh, J. H. (2016). Does transcranial direct current stimulation enhance cognitive and motor functions in the ageing brain? A systematic review and meta- analysis. *Ageing Res. Rev*. 25, 42–54 doi: 10.1016/j.arr.2015.11.004

Takai, H., Tsubaki, A., Sugawara, K., Miyaguchi, S., Oyanagi, K., Matsumoto, T., et al. (2016). Effect of transcranial direct current stimulation over the primary motor cortex on cerebral blood flow: a time course study using near-infrared spectroscopy. *Adv. Exp. Med. Biol*. 876, 335–341. doi: 10.1007/978-1-4939-3023-4_42

Von Hippel, P. T. (2015). The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Med. Res. Methodol*. 15:35. doi: 10.1186/s12874-015-0024-z

Wilcox, T., and Biondi, M. (2015). fNIRS in the developmental sciences. *Wiley Interdiscipl. Rev. Cogn. Sci*. 6, 263–283. doi: 10.1002/wcs.1343

Wolkenstein, L., Zeiller, M., Kanske, P., and Plewnia, C. (2014). Induction of a depression-like negativity bias by cathodal transcranial direct current stimulation. *Cortex* 59, 103–112. doi: 10.1016/j.cortex.2014.07.011

Woods, A. J., Cohen, R., Marsiske, M., Alexander, G. E., Czaja, S. J., and Wu, S. (2018). Augmenting cognitive training in older adults (The ACT Study): design and methods of a phase III tDCS and cognitive training trial. *Contemp. Clin. Trials* 65, 19–32. doi: 10.1016/j.cct.2017.11.017

Zhang, H., Huntley, J., Bhome, R., Holmes, B., Cahill, J., Gould, R. L., et al. (2019). Effect of computerised cognitive training on cognitive outcomes in mild cognitive impairment: a systematic review and meta-analysis. *BMJ* 9:e027062. doi: 10.1136/bmjopen-2018-027062

Zheng, X., Alsop, D. C., and Schlaug, G. (2011). Effects of transcranial direct current stimulation (tDCS) on human regional cerebral blood flow. *Neuroimage* 58, 26–33. doi: 10.1016/j.neuroimage.2011.06.018

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 14 April 2021
doi: 10.3389/fnhum.2021.605213

Working Memory Training Effects on White Matter Integrity in Young and Older Adults

Sabine Dziemian[1,5,6,7], Sarah Appenzeller[1], Claudia C. von Bastian[2,3], Lutz Jäncke[4,5] and Nicolas Langer[1,5,6,7]**

[1] Department of Methods of Plasticity Research, Institute of Psychology, University of Zurich, Zurich, Switzerland, [2] Department of Psychology, University of Sheffield, Sheffield, United Kingdom, [3] Neuroscience Institute, University of Sheffield, Sheffield, United Kingdom, [4] Institute of Psychology, Department of Neuropsychology, University of Zurich, Zurich, Switzerland, [5] University Research Priority Program "Dynamic of Healthy Aging", University of Zurich, Zurich, Switzerland, [6] Center for Reproducible Science, University of Zurich, Zurich, Switzerland, [7] Neuroscience Center Zurich (ZNZ), Zurich, Switzerland

OPEN ACCESS

Edited by:
Xiaobo Li,
New Jersey Institute of Technology, United States

Reviewed by:
Nico Bunzeck,
University of Lübeck, Germany
José Manuel Reales,
National University of Distance Education (UNED), Spain

***Correspondence:**
Sabine Dziemian
sabine.dziemian.academic@gmail.com
Nicolas Langer
n.langer@psychologie.uzh.ch

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

Received: 11 September 2020
Accepted: 15 March 2021
Published: 14 April 2021

Citation:
Dziemian S, Appenzeller S, von Bastian CC, Jäncke L and Langer N (2021) Working Memory Training Effects on White Matter Integrity in Young and Older Adults. Front. Hum. Neurosci. 15:605213. doi: 10.3389/fnhum.2021.605213

Objectives: Working memory is essential for daily life skills like reading comprehension, reasoning, and problem-solving. Healthy aging of the brain goes along with working memory decline that can affect older people's independence in everyday life. Interventions in the form of cognitive training are a promising tool for delaying age-related working memory decline, yet the underlying structural plasticity of white matter is hardly studied.

Methods: We conducted a longitudinal diffusion tensor imaging study to investigate the effects of an intensive four-week adaptive working memory training on white matter integrity quantified by global and tract-wise mean diffusivity. We compared diffusivity measures of fiber tracts that are associated with working memory of 32 young and 20 older participants that were randomly assigned to a working memory training group or an active control group.

Results: The behavioral analysis showed an increase in working memory performance after the four-week adaptive working memory training. The neuroanatomical analysis revealed a decrease in mean diffusivity in the working memory training group after the training intervention in the right inferior longitudinal fasciculus for the older adults. There was also a decrease in mean diffusivity in the working memory training group in the right superior longitudinal fasciculus for the older and young participants after the intervention.

Conclusion: This study shows that older people can benefit from working memory training by improving their working memory performance that is also reflected in terms of improved white matter integrity in the superior longitudinal fasciculus and the inferior longitudinal fasciculus, where the first is an essential component of the frontoparietal network known to be essential in working memory.

Keywords: working memory training, working memory, healthy aging, diffusion tensor imaging, white matter integrity, tractography

INTRODUCTION

Working memory (WM) is defined as the ability to maintain and manipulate goal-relevant information in the face of interference (Baddeley, 2002, 2003; Jonides et al., 2008). WM is essential for daily life skills including reading comprehension (de Jonge and de Jong, 1996), reasoning, and problem-solving (Shah and Miyake, 1999; Pickering, 2006). Healthy aging of the brain is associated with WM decline that can affect the capability of older people to live independently. Therefore, research on how to prevent WM decline with aging is highly demanded, especially given the current demographic change the world is facing (World Health Organization, 2015).

While advances of pharmacological studies did not find adequate therapies so far, recent studies aimed at non-pharmacological intervention, such as cognitive training and physical exercise (Lautenschlager et al., 2008; Geda et al., 2010; Pieramico et al., 2014). Cognitive training interventions are a promising tool for delaying age-related cognitive decline (Kelly et al., 2014) or even improving cognitive functions (Mahncke et al., 2006; Wolinsky et al., 2006; Cheng et al., 2012; Anguera et al., 2013). The vast majority of cognitive interventions target WM (Kelly et al., 2014).

However, the neural mechanisms underlying the beneficial effects of WM training on aging brains remain unclear and are subject to debate. Neuroimaging techniques facilitate the investigation of the impact of WM training on aging individuals and the exploration of its mechanisms. Several neuroimaging meta-analyses demonstrated that frontoparietal regions are consistently activated during different WM tasks and modalities (Wager and Smith, 2003; Owen et al., 2005; Rottschy et al., 2012; Salmi et al., 2018; Emch et al., 2019). The frontoparietal network typically consists of the lateral prefrontal cortex (PFC), which is associated with encoding, manipulation, and response selection (D'Esposito et al., 2000), and the posterior parietal cortex (PPC), which is relevant for storing, maintaining, and retrieving information (Jonides et al., 1998; Guerin and Miller, 2011; Nielsen et al., 2017). In addition, modality-dependent higher-level sensory areas are also shown to be essential for WM functions (Jung and Haier, 2007).

Thus, efficient communication between these brain areas appears crucial for WM performance. Evidence from functional neuroimaging has revealed increased functional connectivity between PFC and PPC as WM load increases (Ma et al., 2012; Dima et al., 2014). Moreover, coupling between visual areas and the parietal and inferior temporal regions, known as the ventral and dorsal visual stream (Goodale and Milner, 1992; Hebart and Hesselmann, 2012), have been linked to WM performance (Hampson et al., 2006; Cocchi et al., 2013; Langer et al., 2013; Shen et al., 2015). Various WM training studies in healthy young adults revealed alterations in brain characteristics of the frontoparietal network (Langer et al., 2013; Finc et al., 2020). A recent meta-analysis has further substantiated induced functional brain changes in the frontoparietal network following WM training (Duda and Sweet, 2019).

Considerable evidence indicates that in addition to the transformation of functional brain characteristics, WM training can even lead to underlying structural brain plasticity in gray and white matter. Prior longitudinal studies identified gray matter changes in brain areas of the frontoparietal network after cognitive (including WM) training in healthy young and older adults (Lampit et al., 2015; Jiang et al., 2016; Metzler-Baddeley et al., 2016; Román et al., 2016).

The relevant cortical brain regions of the frontoparietal network are connected through several axonal bundles, which can be investigated with diffusion tensor imaging (DTI). However, the literature on WM training investigating effects on diffusion metrics is rather scarce. A few studies indicate WM training might induce changes to frontoparietal network relevant white matter fiber tracts such as the superior longitudinal fasciculus (SLF) (Burzynska et al., 2011; Salminen et al., 2016), the inferior fronto-occipital fasciculus (IFOF) (Salminen et al., 2016), the inferior longitudinal fasciculus (ILF) (Nagy et al., 2004) and the forceps minor (Takeuchi et al., 2011; Salminen et al., 2016). However, only Salminen included an active control condition, so it is unknown whether previously reported training effects were specific to WM training or if the neural changes would have been seen with any intensive training task, regardless of cognitive domain. Moreover, white matter changes following WM training were only investigated in young participants. Thus, it remains unclear to what extent WM training has the potential to induce white matter plasticity in older adults.

The present paper aims to investigate the effects of WM training on white matter integrity in young and older participants. Hereby, an improvement in white matter integrity denotes potentially beneficial changes in neuronal microstructure (e.g., increased myelination or spatial rearrangement of fibers) that are operationalized as changes in diffusivity measures from DTI scans. For this purpose, we conducted a longitudinal DTI study to investigate the effects of an intensive four-week adaptive WM training intervention on white matter integrity quantified by the diffusivity measure mean diffusivity (MD). We compared MD for young and older adults who were randomly assigned to an adaptive working memory training group (WM) or an age-matched adaptive active control group (AC). The data presented here are part of a larger study in which we investigated WM training effects on cognitive performance (von Bastian et al., 2013a) and resting-state EEG (Langer et al., 2013) in which we found gains in the trained tasks in both young and older adults. In this study, using a subset of the larger study including participants that completed DTI scans, we expected to replicate the behavioral findings of an improvement in WM performance after the WM training for both age groups in the WM group compared to both age groups of the AC group. Based on previous studies, we expect that WM training induces increased white matter integrity in fiber tracts of the frontoparietal network and ventral visual WM stream, including the superior longitudinal fasciculus (SLF) (Vestergaard et al., 2011; Salminen, 2016), the inferior fronto-occipital fasciculus (IFOF) (Salminen, 2016), the inferior longitudinal fasciculus (ILF) (Salminen, 2016) and the corpus callosum (CC) (Nagy et al., 2004; Charlton et al., 2010; Takeuchi et al., 2010).

Consequently, we hypothesize that both age groups show significantly decreased MD values, reflecting increased white matter integrity, in the above white matter tracts after WM training.

METHODS

Participants

This study focuses on a subset of 52 participants who took part in a larger research project and completed additional neurophysiology recordings and neuroimaging scans. In the larger research project, we investigated the effects of WM training on cognitive outcomes and changes in EEG resting-state activity, which have been published elsewhere (von Bastian et al., 2013a; Langer et al., 2013); data from the brain imaging measures have not been analyzed previously. The sample consisted of 32 young (19 women, 13 men; mean age = 23, SD = 3.34, age range 19 – 36 years) and 20 older participants (8 women, 12 men; mean age = 69, SD = 3.57, age range 65 – 77 years). Each participant was randomly assigned either to a working memory group (WM) (16 young, 10 older) or an active control group (AC) (16 young, 10 older). The corresponding age groups did not differ regarding their demographic variables of age (young: $t(30) = -1.01$, $p = 0.32$; older: $t(17) = 0.76$, $p = 0.46$), gender ($\chi^2 = 0.02$, $p = 0.89$), education (young: $z = -0.80$, $p = 0.42$; older: $z = 0.91$, $p = 0.36$), their experience in using a computer (young: all $p > 0.26$; older: all $p > 0.07$), and cognitive activity in daily life (young: $t(30) = 0.77$, $p = 0.45$; older: $t(17) = -0.12$, $p = 0.91$). Basic demographics of the groups are listed in **Table 1**. Descriptive statistics regarding the experience using a computer are shown in **Supplementary Tables 1a,b**. Further information regarding the assessment of education is given in the **Supplementary Section B**. All participants were consistently right-handed according to the Annett-Handedness-Questionnaire (Annett, 1970), and highly proficient Swiss German or standard German speakers. Participants reported no history of psychiatric or neurological disease, neuropsychological problems, or medication and drug abuse. All participants gave written informed consent to participate in the study. This study was conducted according to the principles expressed in the Declaration of Helsinki and was approved by the Institutional Review Board of "Kantonale Ethikkommission" (EK: E-80/2008).

TABLE 1 | Participants demographics as a function of group.

Age group	Young		Older	
Group	**WM**	**AC**	**WM**	**AC**
Sample size (n)	16	16	10	10
Age (M ± SD)	22.38 ± 2.33	23.56 ± 4.10	69.70 ± 3.89	68.44 ± 3.28
Gender (w : m)	9 : 7	10 : 6	5 : 5	3 : 7
Education (M ± SD)	5 ± 1	5 ± 1	6 ± 2	5 ± 1

WM, working memory; AC, active control group; w, women; m, men. Education is shown in ranks ranging from 1 (no formal education) to 8 (doctorate).

Procedure

All participants were asked to complete 20 sessions of training (approx. 25 – 30 min per session) over a period of four weeks. Participants assigned to the WM group practiced WM tasks. Training for the participants assigned to the AC group comprised tasks that were not expected to have an effect on WM performance but were similarly challenging and motivating. The study was conducted double-blinded, so that neither participants nor experimenters were aware of the group participants were assigned to. Participants were also not informed about the intended effect of the training. An AC group was incorporated into this study design to control for expectancy effects (Oken et al., 2008) and intervention effects (Mahncke et al., 2006), which here would be the adherence to regular training and the completion of computer-based tasks that demand considerable concentration. Each session consisted of three tasks in randomized order with a duration of approximately 10 min per task. All groups started their first session at the same level of difficulty. Consecutive task difficulty (i.e., level) was adapted based on individual performance (see below in the task description for the details on difficulty adaption). The training was executed at home via the open-source software Tatool (von Bastian et al., 2013b) and results of each session were uploaded to a web server to monitor participants' commitment. Before and after the training period, participants completed a cognitive test battery comprising a series of cognitive tasks, EEG, and, for the subgroup analyzed in the present article, MRI scans (DTI and T1-weighted scans) in the laboratory. The present study only focuses on diffusivity measures from DTI that quantify neuroanatomical microstructures. Cognitive WM training outcomes were analyzed in terms of a partial replication as premise for subsequent analysis of neuroanatomical measures. The full results of the cognitive outcomes are reported in von Bastian et al. (2013a), and resting-state EEG analysis are reported in Langer et al. (2013).

Training Intervention

Working Memory Training Tasks

The WM training was based on the facet model of WM described in Oberauer et al. (2000, 2003). This model was established using a factor-analytic approach, following the framework of facet theory (Guttman, 1954; Canter, 1985) which suggests the differentiation of the WM along two main facets: *cognitive functions* and *content domains* (Oberauer et al., 2000). The function facet can further be divided into three categories: *storage and processing*, *relational integration*, and *supervision*. Storage and processing refer to simultaneously processing and storing information (Salthouse, 1990) and are often assessed with complex span tasks. Relational integration (also referred to as monitoring) is the coordination of information elements into new structures (e.g., Halford et al., 1998). Finally, supervision represents the selective activation of goal-relevant and inhibition of goal-irrelevant information, and is typically measured with task switching paradigms (Oberauer et al., 2000, 2008). The content domains comprise verbal, numerical, and spatial materials. For each of these categories

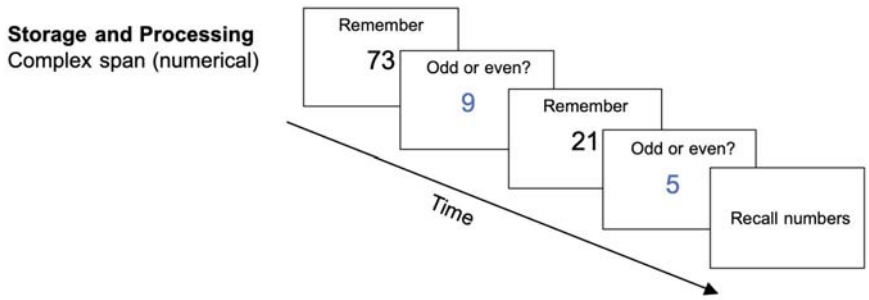


FIGURE 1 | Storage and processing task (complex span) (reused and adapted with permission from Langer et al., 2013). Participants had to memorize the blue digits which were interrupted by a distractor item on which they needed to judge the parity (Langer et al., 2013; von Bastian et al., 2013a).

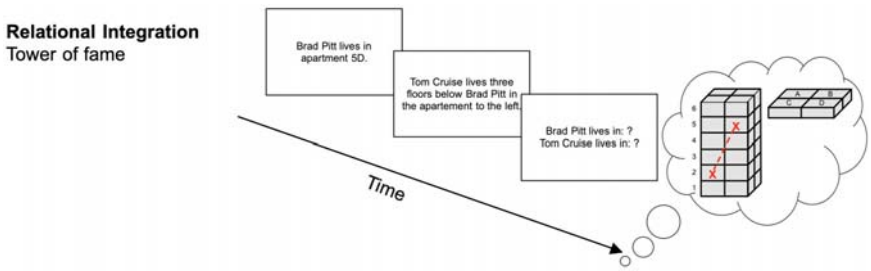


FIGURE 2 | Relational integration task (tower of fame) (reused and adapted with permission from Langer et al., 2013). Participants needed to integrate information into the tower of fame and were then asked about the apartments people lived in Langer et al. (2013); von Bastian et al. (2013a).

participants in the WM group trained one specific task in each session: a numeric complex span task for storage and processing, a task switching task for supervision, and the tower of fame task for relational integration (Langer et al., 2013; von Bastian et al., 2013a). The tasks were chosen based on the results of a previous study, which examined the effects of training on the three functional categories separately (von Bastian and Oberauer, 2013).

Storage and processing: numerical complex span

Each trial started with the central display of a memory item (a two-digit number, printed in black) for 0.5 s. After that, a distractor (a single-digit number, printed in blue) was presented for which participants had to judge the parity (odd or even) as quickly and as accurately as possible within 3 s. The digit remained on the screen until the participant's response, and a blank screen was shown until the 3 s passed. This was followed by another memory item and then another distractor-decision (see **Figure 1**). The length of the item-distractor sequence depended on the difficulty level. Afterward, participants had to recall all memory items in the correct order. There was no time limit set for the recall. Each session consisted of 12 recall trials. The achieved level of difficulty was used as a performance measure.

Relational integration: tower of fame

The Tower of Fame task (see **Figure 2**) required participants' ability to integrate single information elements and the relations between them. Participants had to imagine a tower with six floors, each consisting of four apartments (A, B, C, and D). They were then sequentially presented with statements about the location of famous people's apartments. Each trial started with a statement about one particular apartment (e.g., "Tom Cruise lives in apartment 2A.") and was then followed by statements with information relative to the first statement (e.g.,"Brad Pitt lives two floors above Tom Cruise in the apartment to the right."). At the end of each trial, participants had to recall the apartments for each of the people mentioned (e.g., "Tom Cruise lives in? – 2A" and "Brad Pitt lives in? – 4B"). At increased difficulty levels the order of recall was randomized and more statements were presented. Hence, participants had to memorize the binding between apartment numbers (e.g., "2A") and names (e.g., "Tom Cruise"). Each session comprised 15 trials. Performance was measured by the achieved level of difficulty.

Supervision: figural task switching

In this task, participants were presented with bivalent stimuli (simple geometrical shapes, see **Figure 3**). These stimuli had to be categorized based on set criteria that switched every two trials. For example, participants had to switch between determining whether a shape had a border and whether it was dotted or striped. The decision had to be made as accurately and as quickly as possible. At increased difficulty levels, display duration, and therefore the time to respond to the stimulus was set to the 99th percentile of the individual

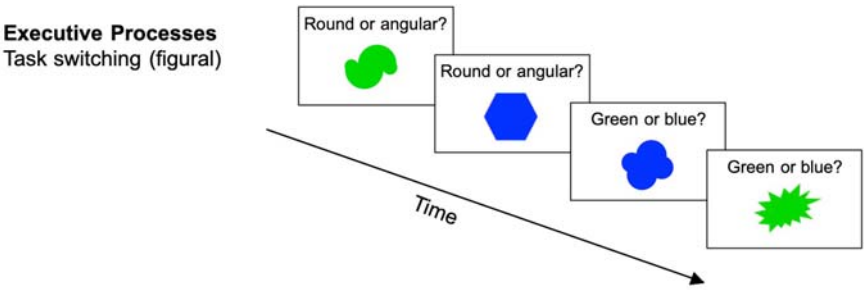


FIGURE 3 | Executive processes task (task switching) (reused and adapted with permission from Langer et al., 2013). Participants were presented with stimuli on which they had to decide if they were round vs. angular or green vs. blue dependent on the current task (Langer et al., 2013; von Bastian et al., 2013a).

reaction times (RT) based on all previous trials since the last adjustment (see von Bastian and Oberauer, 2013). Additionally, in every fifth session, the sets of stimuli were replaced (i.e., new stimuli and new categorization rules) in order to enhance variability. Each session consisted of 384 trials. Performance was measured as the proportional switch cost, which is the reaction time difference between task switch trials and repetition trials divided by the average reaction time of both switch and repetition trials.

Active Control Training Tasks

To hold the variability of the training tasks constant, the AC group completed three different tasks in each training session as well. As the WM group, the AC group performed 20 sessions of training at home with approximately the same duration. Their training consisted of a general knowledge quiz, a visual search task, and a counting task (von Bastian et al., 2013a; Langer et al., 2013).

Quiz

Participants had to answer questions taxing general knowledge by choosing one of four response options within 60 s. The training consisted of 3'507 questions provided by the Quiz-Fabrik GmbH. Each session comprised 100 trials. Task difficulty was increased according to the item difficulty provided by Quiz-Fabrik. Performance was measured by the percentage of correct answers.

Visual search

Participants were presented multiple circles with two gaps and were asked to search for a target item, a circle with only one gap. Participants then had to point out the direction of the gap with the arrow keys within 60 s. In trials without a target, participants had to press "A." At increased difficulty levels a larger number of circles was displayed. In each session, participants had to complete 70 trials. Performance was measured as the percentage of correct answers. If there was no answer given by the participant in a trial, the response was treated as incorrect.

Counting

Blocks of identical digits were presented for 60 s. These blocks followed the rule that they contained as many digits as the digit indicated (e.g., a sequence of three 3s, or a sequence of five 5s). Participants were asked to determine whether any block presented broke this rule, and then press the respective number. If all blocks were correct, participants had to press "0." Each session comprised 70 trials. The percentage of correct answers was used as the performance score. In the case of not providing a response, the trial was counted as incorrect.

Pre- and Post-assessment

Before and after the training intervention, participants completed a cognitive test battery comprising the three trained WM tasks, three structurally similar tasks with different materials, three structurally different WM tasks, and one (older adults) or two (young adults) reasoning tasks (for detailed descriptions, see von Bastian et al., 2013a). In a supplementary analysis (see **Supplementary Section B**) of the behavioral data, that is a partial replication of a previous study (see von Bastian et al., 2013a), we found that the WM training group showed performance gains in all training tasks; however, only for the complex span and the Tower of Fame task, but not for the switching task, were these improvements were significantly different from those observed in the AC group. This supplementary analysis on behavioral training effects served as a prerequisite for the consecutive DTI analysis, demonstrating WM training gains are also present in the subsample for which we had DTI data. In line with the previous larger study (see von Bastian et al., 2013a; Langer et al., 2013), in the present study, we focused exclusively on data from the complex span and the Tower of Fame task. The pretest/posttest versions of these tasks were exactly identical to the training versions, except that the difficulty levels were fixed. The complex span task comprised 15 trials with set sizes ranging from 3-7 numbers, and the Tower of Fame task comprised 18 trials with 2-4 statements and pseudo-randomized order of recall. We computed a composite WM score for pretest and posttest by averaging the z-transformed scores of both tasks as follows:

$$Composite\ WM\ Score = \frac{z\left(Storage\ and\ Processing\right) + z\left(Relational\ Integration\right)}{2}$$

MRI Data Acquisition

The magnetic resonance imaging (MRI) scans were acquired on a 3.0 T Philips Achieva whole-body scanner (Philips Medical System, Best, The Netherlands) equipped with a transmit-receive body coil and a commercial eight-element sensitivity encoding (SENSE) head coil array. Volumetric 3D T1-weighted gradient-echo sequence scans were obtained with a measured spatial resolution of 0.94 × 0.94 × 1 mm (acquisition matrix 256 x 256 pixel, 160 slices). Further imaging parameters were: field of view FOV = 240 × 240 mm, echo time TE = 3.7 ms, repetition time TR = 8.06 ms, flip angle = 8, and SENSE factor R = 2.1.

Diffusion-weighted spin echo-planar (EPI) sequence scans were obtained with a measured spatial resolution of 2.0 × 2.0 × 2.0 mm (acquisition matrix 112 × 112 pixels, 75 slices). Further imaging parameters were: field of view FOV = 224 × 224 mm, echo time TE = 55 ms, repetition time TR = 13.006 ms, flip angle FA = 90, and SENSE factor R = 2.1. Diffusion was measured in 64 non-collinear directions preceded by a non-diffusion-weighted volume (reference volume). The b-value was 1.000 s/mm^2.

Preprocessing

Diffusion-weighted images of all participants and both time points underwent identical processing steps using the FMRIB Software Library (FSL) version 6.0 (Jenkinson et al., 2012). First, the FSL Brain Extraction Tool (BET) was applied, extracting the brain from non-brain tissue from the whole head image resulting in a binary brain mask (Smith, 2002). The brain mask was created with a fractional anisotropy (FA) threshold of 0.2. In the following step, eddy current-induced distortions, which are a common artifact of diffusion images, were removed from the data using the FSL tool "eddy_cuda". This tool also corrects for artifacts of participant's in-scanner head motion (Andersson and Sotiropoulos, 2016; Andersson et al., 2016, 2017). In detail, the eddy_cuda tool was applied with the number of iterations set to 8 and decreasing smoothing full-width-half-max parameters (10,6,4,2,0,0,0,0) in each iteration. Outlier detection and replacement were enabled for slice-wise outliers which concern signal dropouts that occur within a single slice. Furthermore, the parameters of eddy_cuda were set to run additional 8 iterations for the slice-to-volume correction, which considers movement within a volume instead of between volumes by modeling 9 degrees of freedom in movement for each volume. Eddy_cuda was chosen over previous eddy versions as it additionally implements intra-volume movement correction and therefore reduces remaining artifacts and its potential impacts on measures extracted further down in the processing pipeline (Andersson et al., 2017).

The subsequent preprocessing steps included diffusion tensor fitting to obtain diffusion measures (e.g., FA and MD) at each voxel by calculating the tensors of the diffusion weights images based on a linear regression using the tool "dtifit" from FMRIB's Diffusion Toolbox. This step was followed by ac-pc alignment on a T1-weighted image from Talairach coordinates into MNI space by using the function "mrAnatAutoAlignAcpcNifti" from the vista-soft toolbox developed by the Stanford Vista Lab, 2016[1]. The T1-weighted image served as an anatomical reference for subsequent extraction of measures of white matter integrity using the Automated Fiber Tract Quantification (AFQ) toolbox version 1.2 (Yeatman et al., 2012) in MATLAB 2015b.

AFQ is a deterministic tractography algorithm that identifies 20 major fiber tracts and calculates diffusion properties for 100 equidistant nodes of each tract yielding individual fiber tract profiles. Relevant to the present study, tractography was obtained on fiber tracts, which belong to the frontoparietal network and revealed a significant relationship with WM performance in previous publications (Nagy et al., 2004; Charlton et al., 2010; Takeuchi et al., 2010; Vestergaard et al., 2011; Salminen, 2016). Specifically, tractography was conducted for the callosum forceps minor, the left and right IFOF, the left and right SLF, the left and right ILF, and the left corticospinal tract (CST). The latter was chosen as a control tract expecting to have no training-induced changes.

In brief, the AFQ tractography algorithm is implemented as follows: first, a whole-brain tractography was estimated with a deterministic streamline tracking algorithm (STT) (Mori et al., 1999) by seeding voxels with FA values greater than 0.2. The tracking was interrupted when either (1) the FA value at that specific position dropped below 0.25 or (2) the minimal angle between the current and the following segment was greater than 35° implying an unusual fiber orientation. Second, fiber tracts were segmented by a waypoint ROI procedure where all fibers passing through the same two ROIs, as identified by the participant's T1-weighted scan, were assigned to the identical fiber group (Wakana et al., 2004). Third, each fiber tract was matched with fiber tract probability maps as created by Hua et al. (2008). Fibers exceeding a maximum aberrance of 0.25 were excluded from further analyses. Fourth, fibers deviating more than 4 standard deviations from the mean fiber length or more than 5 standard deviations from the core of the fiber tracts were iteratively eliminated at this point. Fifth, after eliminating all potential outliers, the fiber tracts were clipped at the two waypoint ROIs and then resegmented into 100 equidistant nodes where diffusion properties (i.e., FA and MD) were computed (Yeatman et al., 2012). These properties were computed as weighted averages of the diffusion properties of all fibers within the node, with the weights being the probability that the fibers are part of the fiber group.

DTI allows for the extraction of various microstructural measures based on the diffusion of water molecules that is restricted by impermeable tissue. Water can diffuse best along the direction of myelinated fiber tracts, microtubules, and cell membranes. Therefore, measures of diffusivity can be leveraged to quantify neuroanatomical microstructures. These diffusivity measures include axial diffusivity (AD), radial diffusivity (RD), FA, and MD. AD represents the diffusion rate along the principal axis (Soares et al., 2013; Curran et al., 2016), whereas RD reflects the perpendicular diffusivity to the principal axis, respectively. FA characterizes the restriction of diffusion and is defined as

[1]https://github.com/vistalab/vistasoft

the degree of anisotropy of water diffusion at a given voxel. FA is commonly associated with increased myelination and white matter fiber health although it is strongly influenced by various other factors such as axon branching, packing density, axon diameter, number of axons, and fiber crossing (Zatorre et al., 2012). The latter has often been referred to as the *"crossing fiber problem"* in literature (Alexander and Seunarine, 2010; Tournier, 2010; Jones et al., 2013). Finally, MD is a measure reflecting overall diffusivity in any direction and is defined as the mean of the three eigenvectors of the diffusion tensor (Soares et al., 2013). In a study by Cox et al. (2016), they compared various microstructural diffusion measures on a large sample from the United Kingdom Biobank. They found MD to be most sensitive to identify white matter changes in aging populations. In contrast to FA, MD was non-linearly associated with increased age reflected in a steeper slope and thus stronger differentiation in older age (Cox et al., 2016). In addition, several papers have demonstrated that MD is more sensitive to study short-term plasticity. Sagi et al. (2012) and Tavor et al. (2013) also chose MD as a measure of microstructural fiber organization, to investigate short-term plasticity changes after learning. Their choice of using MD over other measures is based on a previous study by Hofstetter and Assaf (2017) in which rapid MD changes were evident after short periods of training. Therefore in this paper, statistical analysis was conducted using MD as the measure quantifying change in white matter integrity with the focus on older participants.

Here, we investigated tract-wise MD, for the eight above mentioned fiber tracts. For those we extracted MD for each of the 100 nodes between the ROI of the tract using AFQ and presented the descriptive statistics of the individual nodes. For statistical analyses, we calculated the average MD value along these 100 nodes. Hence, we obtained one MD value per tract for the eight previously mentioned tracts of interest. Furthermore, we also investigate MD changes on the whole-brain level, subsequently referred to as global MD.

Quality Metrics

Beyond the extraction of measures of interest, DTI data were also analyzed for metrics reflecting imaging quality for both time points separately. We used the "eddy_openmp" tool and the "eddy_quad" (QUality Assessment for DMRI) tool from FSL version 6.0.03 to extract estimates of in-scanner head motion and b-shell-wise CNR metrics (Bastiani et al., 2019). The final CNR score was defined as the average value across all b-shell metrics as reported by eddy_quad (Bastiani et al., 2019). In-scanner head motion was calculated as the average relative motion across all volumes.

Two participants of the older WM group showed a disproportional low CNR for the DTI scan of the first time point (i.e., 5 median standard deviations from the median over all subjects for each specific analysis), for which reason they were excluded from subsequent analyses of neuroanatomical measures. Exclusion of those two participants served to maintain high image quality regarding the small sample sizes of the subgroups and ensure data quality differences are not biasing group results.

Statistical Analyses

All statistical analyses were conducted using the Statistics and Machine Learning Toolbox version 11.5 in MATLAB 2019a. In the following formulas fixed effects are denoted by a "+" symbol and interaction effects by an "*" symbol in line with the Wilkinson notation (Wilkinson and Rogers, 1973). All effect sizes of significant effects from the generalized linear mixed-effects model analyses were calculated using Cohen's $f2$ (Selya et al., 2012).

For a replication of the behavioral training gains in each task within the subsample used in this study see **Supplementary Section B**.

Behavioral Training Effects

Behavioral training effects were analyzed with a generalized linear mixed-effects model using the *composite WM score* of the pre- and post-assessment as the dependent variable, *age* (young/older), *training* (WM group/AC group) and *time point* (pre-assessment/post-assessment) as the fixed effects and *subject* as the random effect. We also specified random slopes for *age*, *training*, and *time point*, resulting in the following formula used:

$$\begin{aligned}&\text{composite WM score} \sim \text{age} * \text{training} * \text{time point} + \\ &(1|\text{subject}) + (-1 + \text{age}|\text{subject}) + (-1 + \text{training}|\text{subject}) + \\ &(-1 + \text{timepoint}|\text{subject})\end{aligned}$$

Effects on In-Scanner Head Motion

For assessing the effects on in-scanner head motion a generalized linear mixed-effects model was used that specified the *head motion* as dependent variable and *age* (young/older), *training* (WM group/AC group) and *time point* (pre-assessment/post-assessment) as fixed effects in addition to a random effect of *subject* and random slopes for *age*, *training*, and *time point*. Therefore, the model was specified as follows:

$$\begin{aligned}&\text{head motion} \sim \text{age} * \text{training} * \text{time point} + (1|\text{subject}) + \\ &(-1 + \text{age}|\text{subject}) + (-1 + \text{training}|\text{subject}) + \\ &(-1 + \text{time point}|\text{subject})\end{aligned}$$

Effects on CNR

Effects on contrast-to-noise-ratio (CNR) were estimated using a generalized linear mixed-effects model. Hereby, the dependent variable was the CNR value and *age* (young/older), *training* (WM group/AC group), *time point* (pre-assessment/post-assessment) and in-scanner *head motion* were defined as fixed effects and *subject* as random effect. Additionally, there were also random slopes defined for *age*, *training*, *time point*, and in-scanner *head motion* yielding the following model:

$$\begin{aligned}&\text{CNR} \sim \text{age} * \text{training} * \text{time point} + \text{head motion} + \\ &(1|\text{subject}) + (-1 + \text{age}|\text{subject}) + (-1 + \text{training}|\text{subject}) + \\ &(-1 + \text{time point}|\text{subject}) + (-1 + \text{head motion}|\text{subject})\end{aligned}$$

Effects on White Matter Integrity

To investigate the effects of the WM training on measures of MD, generalized linear mixed-effects models were calculated. This was performed for global MD and for the previously mentioned tract-wise MD values. For the subsequent analysis, the following generalized linear mixed-effects model was calculated separately for *global MD* and *tract-wise MD*.

$$\begin{aligned}&\text{MD} \sim \text{age} * \text{training} * \text{time point} + \text{CNR} + \text{head motion} \\ &+ (1|\text{subject}) + (-1 + \text{age}|\text{subject}) + (-1 + \text{training}|\text{subject}) + \\ &(-1 + \text{time point}|\text{subject}) + (-1 + \text{CNR}|\text{subject}) + \\ &(-1 + \text{head motion}|\text{subject})\end{aligned}$$

Hereby, the dependent variable *global MD* and *tract-wise MD*, respectively, were estimated by the fixed effect of *age* (young/older), *training* (WM group/AC group), *time point* (pre-assessment/post-assessment), *CNR*, and in-scanner *head motion*, as well as by a random effect of *subject*, while considering random slopes for each of the above fixed effect variables.

To interpret the threefold interactions of *age*, *time point*, and *training*, the same generalized linear mixed-effects model was calculated separately for each age group, thus reducing the model to the following:

$$\begin{aligned}&\text{MD} \sim \text{training} * \text{time point} + \text{CNR} + \text{head motion} + \\ &(1|\text{subject}) + (-1 + \text{training}|\text{subject}) + \\ &(-1 + \text{time point}|\text{subject}) + (-1 + \text{CNR}|\text{subject}) + \\ &(-1 + \text{head motion}|\text{subject})\end{aligned}$$

RESULTS

Pre- and Post-assessment of Working Memory Performance

First, we applied an unpaired *t*-test independently for the young and the older group to review potential pre-assessment differences between the WM and AC group. The results indicated no difference in the composite WM score between the WM group and the AC group at the time of pre-assessment for the young group, $t(30) = -1.54$, $p = 0.13$, and the older group, $t(18) = -0.22$, $p = 0.83$. To investigate the effect of *age, training,* and *time point*, we conducted a generalized linear mixed-effects model analysis as introduced above.

The analysis from pre- to post-assessment (see **Table 2**) yielded a significant main effects of *age* on the composite WM score, $\beta = 1.13$, $p = 2.481 \times 10^{-6}$, 95% *CI* = [0.68, 1.58], $f^2 = 1.17$, indicating higher WM scores for the young participants compared to the older participants. The interaction of *training* × *time point* yielded significant effects on the composite WM score, $\beta = 0.74$, $p = 0.005$, 95% *CI* = [0.23, 1.26], $f^2 = 0.10$, indicating for both age groups greater increases in WM performance for the WM group than for the AC group. There was no interaction effect of *age*, *training,* and *time point*.

TABLE 2 | Generalized linear mixed-effects model results for the composite WM score.

Name	Estimate	*p*	95% CI	f^2
age	1.134	$2.481 \times 10^{-6***}$	[0.684, 1.584]	1.168
training	0.067	0.809	[−0.481, 0.616]	n/a
time point	0.350	0.060	[−0.015, 0.716]	n/a
age × training	0.282	0.426	[−0.418, 0.981]	n/a
age × time point	0.185	0.432	[−0.281, 0.651]	n/a
training × time point	0.743	0.005*	[0.226, 1.260]	0.104
age × training × time point	0.088	0.792	[−0.571, 0.746]	n/a

**p < 0.05, **p < 0.005, ***p < 0.0005. age (0 = older, 1 = young); training (0 = control, 1 = training); time point (0 = pre-assessment, 1 = post-assessment); df = 96.*

DTI Scan Quality Metrics

To assess the effect of *age*, *training*, and *time point* on the quality metrics of the DTI scans, a generalized linear mixed-effects model was fit with the dependent variable being (1) the averaged relative in-scanner head motion estimate and (2) the average CNR.

In-Scanner Head Motion

The model for in-scanner head motion revealed significance for the main effect of *age*, $\beta = 0.10$, $p = 0.03$, 95% *CI* = [0.01, 0.18], $f^2 = 0.07$, indicating increased motion in young participants. There was also a significant interaction effect of *training* × *time point*, $\beta = -0.17$, $p = 0.01$, 95% *CI* = [−0.31, −0.04], $f^2 = 0.06$, revealing a decrease of in-scanner head motion in the WM group from pre-assessment to post-assessment compared to the AC group. Therefore, in-scanner head motion was used as random effect in all subsequent analyses studying the effects of WM on white matter integrity. Complete model results are included in **Supplementary Table 2**.

Contrast to Noise Ratio (CNR)

For the average CNR the model yielded a significant effect only for the main effect of *age*, $\beta = 0.22$, $p = 9.661 \times 10^{-5}$, 95% *CI* = [0.11, 0.33], $f^2 = 0.33$, suggesting greater CNR in young participants. As a consequence, CNR was used as a random effect in all subsequent analyses. Detailed results of this model are provided in **Supplementary Table 3**.

Neuroanatomical Results

Global Mean Diffusivity

A generalized linear mixed-effects model was used to assess the effect of *age*, *training*, and *time point* on the measure of global MD. The model revealed a significant main effect of *age*, $\beta = -0.12$, $p = 0.001$, 95% *CI* = [−0.19, −0.05], $f^2 = 0.11$, indicating increased MD in the older group. No other effects and interaction effects reached significance. The results for this model are shown in **Supplementary Table 4a**. When splitting the model into the two age groups there was no effect of concern that reached significance in either of the age groups (see **Supplementary Tables 4b–c**).

Tract-Wise Averaged Mean Diffusivity

A generalized linear mixed-effects model was calculated to assess the effect of *age*, *training* and *time point* on the tract-wise averaged MD for each of the predefined fiber tracts: the callosum forceps minor, the left, and right IFOF, the left and right SLF, the left and right ILF, and the left CST, where the latter served as control tract in which we were not expecting any training-induced changes. Only the models for the right ILF and the right SLF yielded significant results. Therefore, these are presented in more detail below. The complete results for all fiber tracts are shown in **Supplementary Tables 5a–h**.

Right inferior longitudinal fasciculus

The model for the averaged MD of the right ILF revealed a significant interaction effect of *training* × *time point*, $\beta = -0.12$, $p = 0.047$, 95% $CI = [-0.24, 0.00]$, $f^2 = 0.04$, suggesting a stronger decrease in MD from pre- to post-assessment in the WM group compared to the AC group. We also observed a significant interaction effect of *age* × *training* × *time point*, $\beta = 0.17$, $p = 0.024$, 95% $CI = [0.02, 0.33]$, $f^2 = 0.05$ (see **Supplementary Table 5g**).

To interpret the interaction effect of *age* × *training* × *time point* on MD changes, separate generalized linear mixed-effects models were calculated disjoint for the young and the older group. For the older group the model for the right ILF revealed a significant interaction effect of *training* × *time point*, $\beta = -0.14$, $p = 0.028$, 95% $CI = [-0.27, -0.02]$, $f^2 = 0.15$, suggesting a decrease of MD in the right ILF in the WM group compared to the AC group after the training intervention (see **Table 3** and **Figure 4** and for a full report **Supplementary Table 7**).

TABLE 3 | Generalized linear mixed-effects model results for average MD in the right ILF for older participants.

Name	Estimate	p	95% CI	f^2
training	0.058	0.092	[−0.010, 0.126]	n/a
time point	0.102	0.021*	[0.017, 0.188]	0.029
training × time point	−0.141	0.028*	[−0.265, −0.016]	0.149

$^*p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.0005$. training (0 = control, 1 = training); time point (0 = pre-assessment, 1 = post-assessment); df = 30.

There was no significant interaction effect of *training* and *time point* for the young participants in the separate generalized linear mixed-effects model (see **Supplementary Figure B2**). See **Figure 5** for an illustration of a comparison of segment-wise MD changes in the right ILF for the young and the older participants.

Right superior longitudinal fasciculus

In the model for the averaged MD of the right SLF, there was a significant interaction effect of *training* × *time point*, $\beta = -0.14$, $p = 5.094 \times 10^{-5}$, 95% $CI = [-0.20, -0.07]$, $f^2 = 0.09$, which indicates a decrease of MD in the WM group from pre- to post-assessment. The interaction effect of *age* × *training* was also significant, $\beta = -0.13$, $p = 0.008$, 95% $CI = [-0.22, -0.03]$, $f^2 = 0.08$, suggesting a stronger decrease in MD for

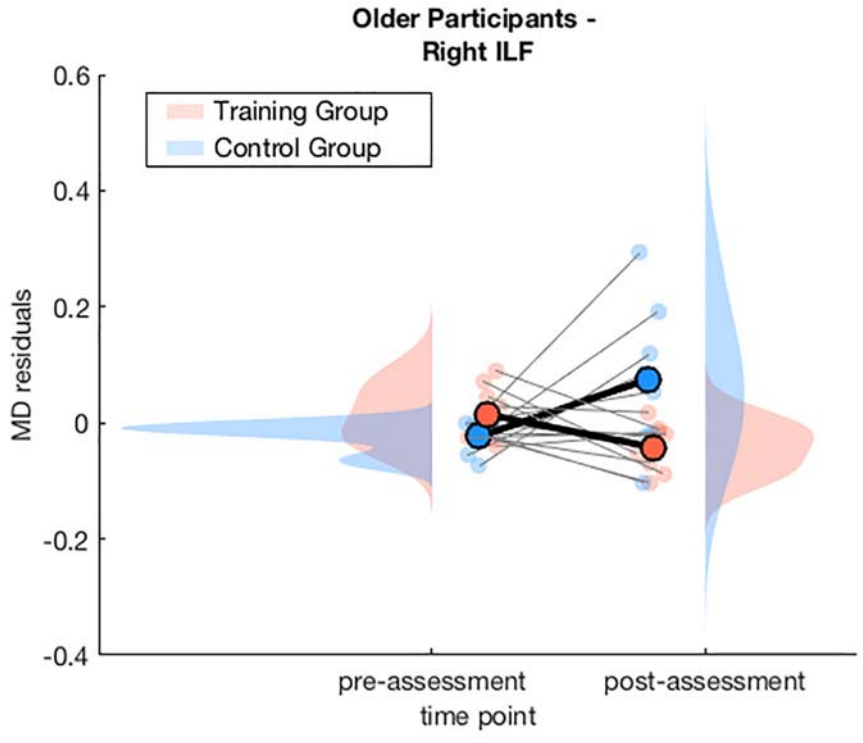


FIGURE 4 | Averaged MD residuals corrected for CNR, in-scanner head motion, and global MD for the right ILF of the WM group and AC group at pre-assessment and post-assessment of the generalized linear mixed-effects model including only older participants.

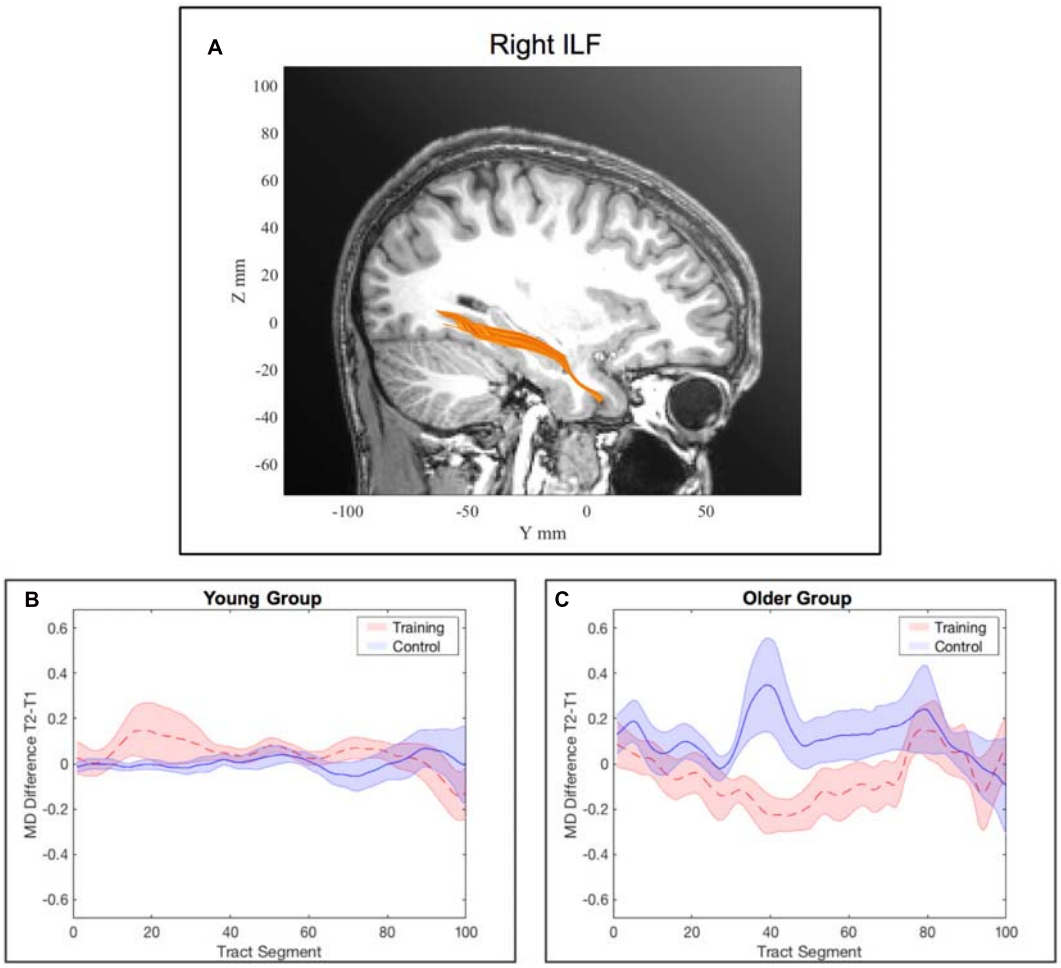


FIGURE 5 | Tract-wise averaged mean diffusivity in the right inferior longitudinal fasciculus (ILF). **(A)** location of the right ILF. **(B)** change in MD from pre- to post-assessment along the tract segments for the young WM group and young AC group separately. **(C)** change in MD from pre- to post-assessment along the tract segments for the older WM group and older AC group separately. Note: the shaded areas represent the standard error, whereas the lines show the mean across all participants within the corresponding group. Subfigure **(B)** was illustrated for completeness although there was no significant interaction effect of *training* and *time point* for the young participants for the right ILF.

the older participants of the WM group. Finally, there was a significant three-way interaction of *age* × *training* × *time point*, $\beta = 0.09$, $p = 0.026$, 95% $CI = [0.01, 0.17]$, $f^2 = 0.15$ (see **Supplementary Table 5e**).

Again, to interpret the significant three-way interaction, separate generalized linear mixed-effects models were calculated independently for the young and the older group to interpret the effect of *training* and *time point* on MD. The model for the right SLF considering only older participants yielded a significant interaction effect of *time point* × *training*, $\beta = -0.11$, $p = 0.040$, 95% $CI = [-0.20, -0.01]$, $f^2 = 0.04$, revealing that MD decreased in the right SLF in the WM group compared to the AC group after the training intervention (see **Table 4** and **Figure 6** and for a full report **Supplementary Table 8a**).

For the young participants, the model on average MD in the right SLF revealed a significant interaction effect of *training* × *time point*, $\beta = -0.05$, $p = 0.003$, 95% $CI = [-0.08, -0.02]$, $f^2 = 0.06$, showing that MD decreased in the right SLF in the WM group compared to the AC group after the training intervention (see **Table 5** and **Figure 6** and for a full report **Supplementary Table 8b**). **Figure 7** illustrates a comparison of segment-wise MD changes in the right SLF for the young and the older participants.

TABLE 4 | Generalized linear mixed-effects model results for average MD in the right SLF for older participants.

Name	Estimate	p	95% CI	f^2
training	0.134	0.042*	[0.005, 0.263]	0.095
time point	0.049	0.125	[−0.014, 0.111]	n/a
training × time point	−0.105	0.040*	[−0.204, −0.005]	0.041

$^*p < 0.05$, $^{**}p < 0.005$, $^{***}p < 0.0005$. training (0 = control, 1 = training); time point (0 = pre-assessment, 1 = post-assessment); df = 31.

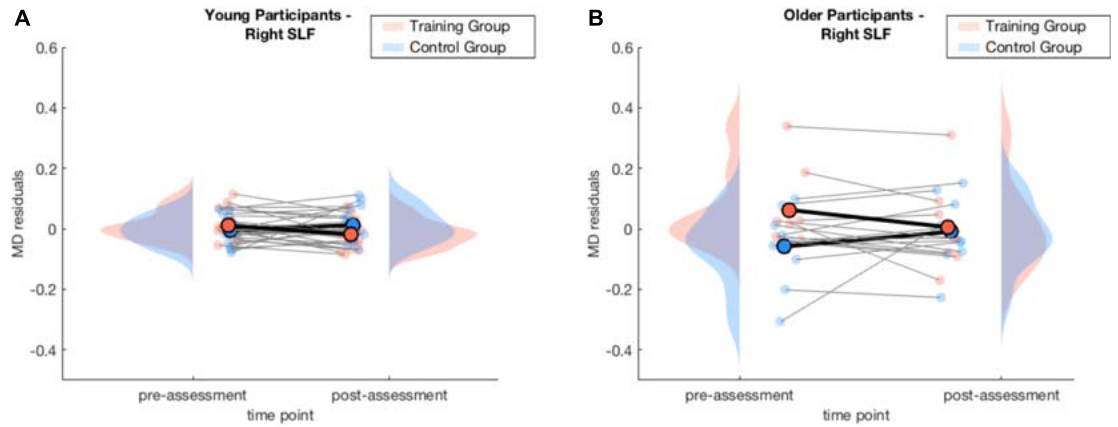


FIGURE 6 | Averaged MD residuals accounting for CNR, in-scanner head motion, and global MD for the right SLF of the WM group and AC group at pre-assessment and post-assessment of the generalized linear mixed-effects model for each age group separately. **(A)** comparison of young WM group and young AC group. **(B)** comparison of older WM group and older AC group.

DISCUSSION

The objective of the present study was to identify the neural mechanisms of white matter changes underlying the benefits of an intensive, double-blinded four-week adaptive WM training in young and older participants. Using DTI, we examined MD in fiber tracts that are associated with WM, such as the SLF, ILF, IFOF and the forceps minor of the corpus callosum (Nagy et al., 2004; Burzynska et al., 2011; Takeuchi et al., 2011; Salminen et al., 2016). During preprocessing, we applied within-volume motion correction and quality control such that throughout the statistical analyses we corrected for these potential confounding variables such as in-scanner head motion and the CNR of the scan.

We could demonstrate that for the young and the older participants the training had a positive effect on white matter integrity reflected by decreased MD in tracts associated with the frontoparietal network. The analysis revealed a decrease in MD in the right ILF in the older WM group. There was also a decrease in MD in the right SLF in both the young and the older WM group. Our control analysis demonstrated that WM training-related white matter integrity changes were specific to these tracts, as the control analysis did not show any changes in the left CST after the WM training.

In the following we will first discuss the behavioral results in more detail, followed by the interpretation of the findings of the white matter changes.

TABLE 5 | Generalized linear mixed-effects model results for average MD in the right SLF for young participants.

Name	Estimate	p	95% CI	f^2
training	0.012	0.455	[−0.019, 0.042]	n/a
time point	0.013	0.228	[−0.008, 0.035]	n/a
training × time point	−0.047	0.003**	[−0.076, −0.017]	0.063

*$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$. training (0 = control, 1 = training); time point (0 = pre-assessment, 1 = post-assessment); df = 57.

Behavioral Results

The analysis of overall WM performance included a comparison of test performance at all training levels between pre- and post-assessment (i.e., the composite WM score). The results revealed that also within this subsample of the larger study there was an improvement in overall WM performance reflected by the significant main effect of time. This partial replication is in line with a previous study that examined WM performance on the whole study sample (von Bastian et al., 2013a) and as such served as the premise for the analysis of white matter integrity changes. The results indicated an improvement for both the young and the older group. In particular, the results showed higher WM scores for the young participants in comparison to the older participants by the significant main effect of age. In the additional analysis of training gains (**Supplementary Section B**), we have also observed an increased training gain in the young group. These findings are not surprising, as age-related cognitive decline also affects WM (Mahncke et al., 2006; Wolinsky et al., 2006; Cheng et al., 2012; Anguera et al., 2013; Kelly et al., 2014) and is the main incentive of this study to which we have demonstrated an impactful intervention. Furthermore, we have also observed an improvement in WM performance in the AC groups. This observation is not surprising, as a possible explanation for an improvement in the composite WM score of the AC group might lie in the general effect of retesting, that has been shown in a recent meta-analysis (Scharfen et al., 2018). More importantly, with the design of an AC and a WM group we could differentiate between the improvement of WM performance in the WM group and the retest effects in the AC group and it further allowed us to distinguish changes that were due to intensive training, regardless of the cognitive domain, to changes that could be attributed specifically to the WM training we designed. Thereby, we observed that WM training indeed resulted in an increased improvement in WM performance compared to the improvement in the AC group, that can only be attributed to the specific WM training tasks. The results showed that the improvement of the WM composite score of the AC group

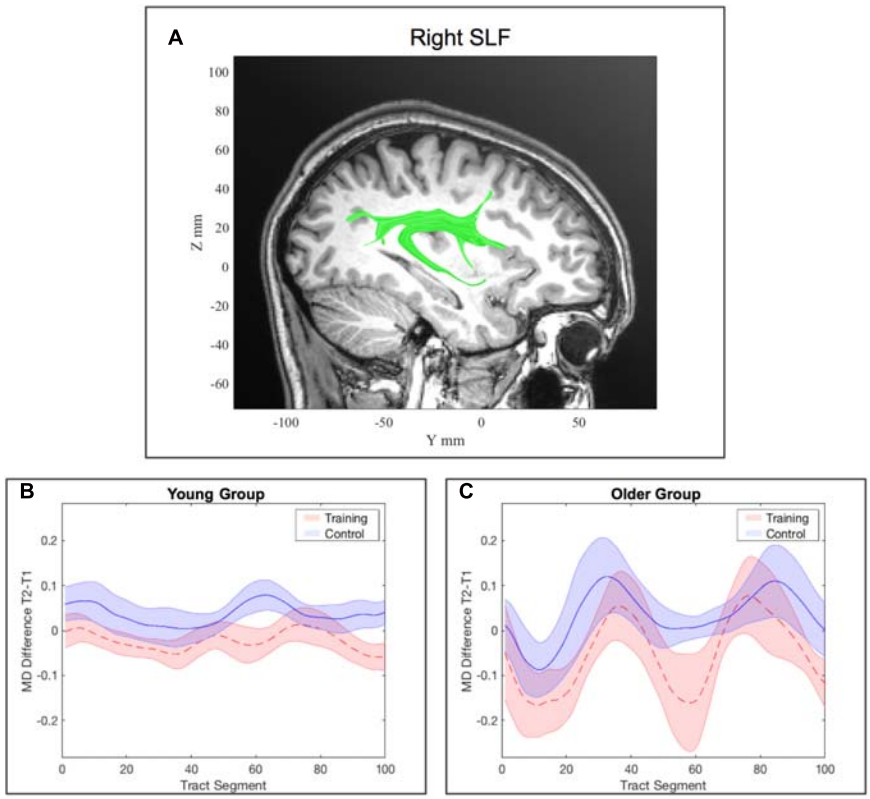


FIGURE 7 | Tract-wise averaged mean diffusivity in the right superior longitudinal fasciculus (SLF). **(A)** location of the right SLF. **(B)** change in MD from pre- to post-assessment along the tract segments for the young WM group and young AC group separately. **(C)** change in MD from pre- to post-assessment along the tract segments for the older WM group and older AC group separately. Note: the shaded areas represent the standard error, whereas the lines show the mean across all participants within the corresponding group.

was significantly lower than in the WM group, demonstrated by the significant interaction effect of time point and training. Nonetheless, the improvement on WM performance must be considered domain specific. Although in a previous study we demonstrated that the training in one domain resulted in a near transfer to other modalities of the same task (numeric complex span task and verbal complex span task), in the present study, we have not investigated these transfer effects. Instead we leveraged the result of improved performance after the four-week adaptive WM training as a premise for subsequent analysis of white matter integrity changes.

White Matter Integrity Results

Diffusion metrics are a frequently used measure to quantify neuroanatomical characteristics of white matter structures. MD is a rotationally invariant diffusivity measure reflecting the average amount of water diffusion or displacement of water molecules in a voxel (Soares et al., 2013). Biological factors like cell membranes, myelin sheath, and microtubules hinder water diffusion. Yet, diffusion is fairly unrestricted along the axons of a fiber tract. Dense packing of cell and axonal membranes decrease MD. Therefore, its value is the lowest in complex tissue (Curran et al., 2016). In previous studies on healthy aging, increase in MD has been also linked to aging (Sexton et al., 2014). In particular, the sensitivity of MD compared to other diffusivity measures in identifying healthy aging-related white matter changes was strongest in advanced age, reflected by a pronounced increase of MD in this population (Cox et al., 2016). Thus, we have focused on MD in the present study. This study is the first to investigate WM training effects in white matter in older adults. We have investigated several fiber tracts that were previously associated with WM performance. Observing a decrease in MD after intensive four-week training is in contrast to the general age-related increase in MD (Cox et al., 2016). Accordingly, the observed WM training-related decrease in MD in tracts that are related to WM might be interpreted as plastic neuroanatomical changes in response to the increased WM demand, counteracting the aging-related increase of MD. We could demonstrate that the WM training-induced white matter integrity changes in the older adults in tracts which are part of the frontoparietal network and the ventral visual WM stream. In detail, after four weeks of intensive adaptive WM training, MD was decreased in the right ILF and the right SLF in the WM group compared to the AC group. In the young participants, we observed white matter integrity changes reflected in a decrease of MD in the right SLF, which is a tract of the frontoparietal network, in the WM group compared to the AC group. In addition, we also inspected the left CST as a control tract in which we did not expect any changes and accordingly did not find any WM training-induced changes in both age groups, as this tract is not directly associated with WM.

Inferior Longitudinal Fasciculus (ILF)

We have identified WM training-related changes in the right ILF. The ILF is an associating fiber tract located in the temporal and occipital lobe and is part of the ventral visual WM stream (Latini et al., 2017). It resembles a U-shape that connects the extrastriate visual association areas with the temporal regions involved in transferring visual signals to anterior temporal regions. In more detail, the ILF mostly connects the lingual and the superior occipital gyrus to the middle and superior temporal gyrus (Panesar et al., 2018). As such, the ILF is a crucial component of the ventral visual stream responsible for object recognition identification (Latini et al., 2017). The function of the ILF also comprises attention, integration of visual information in visually guided behavior, as it is the case in scanning and discriminating and ordering visual features, object processing, lexical and semantic processing (Choi et al., 2012; Herbet et al., 2018). These are cognitive components essential for the visuo-spatial WM (Herbet et al., 2018).

Although previous studies linked the ILF with the visuo-spatial WM and visuoconstructive abilities, their samples consisted of adolescents (Krogsrud et al., 2018), young adults (Salminen et al., 2016), or a pathological group of participants with multiple sclerosis (Dineen et al., 2009). The present study is the first DTI study to investigate white matter integrity changes after WM training in older adults and further contributes to previous literature providing evidence that the ILF is a relevant tract of visuo-spatial WM, by demonstrating its plasticity due to WM training. The observed decrease in MD in the ILF in the older participants after the WM training intervention can be interpreted as a plasticity response causing improved communication between visual areas and anterior temporal regions responsible for processing of visual information. The absence of a significant reduction of MD in the right ILF in the AC group allows us to account the observed changes in the WM group to the adaptive WM training.

In fact, the effect was driven by a combination of an MD decrease in the training group and an MD increase in the control group. Since there were no significant baseline differences in MD for both groups (see **Supplementary Section B**), this observation is due to the interaction effect of *training* × *time point*. This is in line with previous studies that have demonstrated an age-related increase in MD that is in the normal course of healthy aging (Cox et al., 2016). Whereas the decrease in MD in the training group can be interpreted as being indicative for neural changes that were induced by the WM training and thus potentially indicate a means for counteracting cognitive decline in aging.

Particularly, the relational integration task and the storage and processing task highly require the integration of visual and lexical as well as semantic information. Therefore, it is not surprising to see a decrease in MD and thus improved white matter integrity of the right ILF in the older adults after the training.

Superior Longitudinal Fasciculus (SLF)

The SLF is an association fiber tract that mainly projects from the temporal and parietal regions to the frontal cortex. It further receives input from the occipital lobe that it projects to the frontal lobe (Schmahmann et al., 2008). The SLF consists of three components: the dorsal, the major, and the ventral component. In particular, the major and the ventral part are involved in WM. The major component bidirectionally links areas of the prefrontal with areas of the parietal cortex and thus constitutes the main fiber tract of the frontoparietal network, which is considered a main neurocircuit of WM. This connection is necessary for the control of spatial attention and visual perceptions and hence is particularly important for visual WM (Vestergaard et al., 2011).

Our observation of decreased MD after the WM training intervention in the older participants suggests that increased white matter integrity of the right SLF is a consequence of neuronal plasticity providing an improvement in processing and retrieval of spatial information. This is in line with the role of the SFL in visuo-spatial awareness and attention (Schmahmann et al., 2008; Chang et al., 2015). Potentially these changes can be associated with the relational integration task that demands visuo-spatial WM. Information about the location of famous people's apartments in the Tower of Fame task had to be integrated to recall other people's apartment locations. Similarly, to the interpretation in 4.2.1, MD increased in the AC group that can be explained on a physiological basis in the sense of the normal age-related course of cognitive aging (Cox et al., 2016), since there were no statistically significant baseline differences in MD between both groups at the time point of pre-assessment (see **Supplementary Section B**).

There is a substantial body of literature linking the SLF to various facets of WM. In line with our results, Salminen et al. (2016) found increased white matter integrity in the SLF to be related to WM training, yet their study focused on young participants. Another example is a previous lesion study, which identified the SLF to be correlated with spatial WM deficits (Kinoshita et al., 2016). In addition, a study by Burzynska et al. (2011) on young adults also identified that greater WM performance is connected to higher white matter integrity in the SLF. However, the present study is the first to demonstrate white matter integrity changes in the SLF after adaptive WM training in older adults.

Similarly to the ILF, also for this tract, we can disentangle the general learning effects from the effects which are attributed to the WM training as there was no significant decrease in MD in the AC group.

Limitations and Future Directions

As with all diffusivity measures derived from DTI (i.e., FA, AD, and RD), MD comes with the risk of incorrectly quantifying diffusion in voxels that contain more than one dominant orientation of a fiber (e.g., crossing or bending fibers). Whilst a voxel-wise change in MD cannot be directly linked to the specific type of change in neuronal microstructure (e.g., increased myelination or spatial rearrangement of fibers), MD as a whole is still suitable to identify any change of such kind. Although we cannot directly infer increased myelination or denser packing of membranes from a decreased MD value, we can infer neuroanatomical changes improving the magnitude of water diffusion along a defined direction, such as the direction of a tract. A decrease in MD in association with an improvement in WM performance is likely to be the effect of beneficial microstructural changes in terms of increased white matter integrity.

A major limitation of this study is the small number of participants in each of the four groups (older/young vs. training/control). Although the overall sample size of 52 participants is comparable to previous neuroimaging studies on WM training effects (Burzynska et al., 2011; Takeuchi et al., 2011; Lampit et al., 2015; Jiang et al., 2016; Metzler-Baddeley et al., 2016; Román et al., 2016; Salminen et al., 2016), when being split into the groups, each group size is rather small. However, the small group size is a consequence of the study design that included two age groups and an active control group in both age groups in a longitudinal design. In the future, larger studies are required to confirm these results before the findings can be generalized beyond the context of this study. With additional studies including larger sample sizes cumulatively generalizable outcomes can be derived to design clinical interventions for maintaining or enhancing cognitive capabilities in older people and improving their quality of life.

Furthermore, whilst these results demonstrated that healthy, older people's improvements in WM tasks (see also previous study von Bastian et al. (2013a)) manifest in neuroanatomical changes in fiber tracts, we did not investigate the long-term continuity of the behavioral and neuroanatomical WM-related changes. Thus, it remains unclear how long these changes will persist and to what extent they can be preserved with or without a continuation of training. Similarly, this study was not designed to assess the course of improvement and determine at which point gains would reach a plateau. Therefore, future research may focus on the optimal amount of training and follow-up training for sustainable and efficient WM improvement. It may also consider DTI scans during the course of learning to gain more detailed insights into the trajectory of WM improvements and its underlying neuroanatomical changes. Finally, it is yet unknown how changes in fiber tracts translate into everyday life functioning, such as in accomplishing daily activities. Future research would benefit from investigating whether WM training can help older people to live independent lives for longer.

CONCLUSION

This study is one of a few that longitudinally examined white matter plasticity using diffusion metrics to quantify the structural change in response to a WM training intervention. Although there are several studies that investigated WM training-induced white matter changes, these only included young participants. The present paper offered the first insight into neuroanatomical effects of WM training in older adults.

In this project, we investigated white matter integrity changes using DTI after four weeks of adaptive WM training in older participants. The results of the study first show that an adaptive WM training lead to an improved WM performance. In addition, this study confirms the hypothesis that components of the frontoparietal network, including the dorsal (SLF) and ventral (ILF) visual WM pathway, underwent structural change after WM training. Thus, in line with previous studies, the present findings suggest that these fiber tracts are essential for WM. Incorporating an AC group into the study design allowed for increasing the explanatory power of the observed white matter integrity changes. In particular, this allowed for differentiating changes that were due to intensive training, regardless of the cognitive domain, to changes that were attributed to WM training, thereby increasing the explanatory power of the observed white matter integrity changes.

The present results support the idea that age-related cognitive decline can be counteracted or partially reversed by adequate training that provokes white matter integrity improvements on the basis of neural plasticity. This outcome helps understanding the potential of neural plasticity in older adults. Although healthy aging is supposed to be unavoidably associated with cognitive decline, this study's evidence of efficacious neural plasticity in old age can provide a basis that cumulatively with additional studies can be used to develop clinical interventions and age-appropriate therapy that is potentially enhancing end prolonging cognitive abilities.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because we do not have permission from the participants to share the raw data. We can only share derivatives of the data. Requests to access the datasets should be directed to NL, n.langer@psychologie.uzh.ch.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of "Kantonale Ethikkommission" (EK: E-80/2008). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SD did the validation, formal analysis, data curation, writing original draft, writing—review and editing, visualisation, and supervision. SA did the validation, formal analysis, data curation, writing original draft, writing—review and editing, and visualization. CB did the conceptualization, methodology, software, investigation, writing—review and editing, resources, project administration, and funding acquisition. LJ did the conceptualization, resources, project administration, and funding acquisition. NL did the conceptualization, methodology, software, validation, formal analysis, investigation, resources, writing—review and editing, supervision, and project administration. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants of the University of Zurich [Swiss National Science Foundation: project "Short-term and long-term plasticity," as part of the National Center

of Competence in Research (NCCR "Neural plasticity and re-pair")], the Forschungskredit of the University of Zurich, and the Suzanne and Hans Biäsch Foundation for Applied Psychology.

ACKNOWLEDGMENTS

The authors would like to thank all participants that took part in this study. Moreover, they would also like to thank André Locher for programming the training tasks in Tatool, and Sabrina Guye, Veronica Heusser, Melanie Künzli, Eszter Montvai, Katharina Vogt, Helen Wirz, and Marc Züst for their assistance with collecting the data, and Zofia Baranczuck for advice on statistical analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.605213/full#supplementary-material

REFERENCES

Alexander, D. C., and Seunarine, K. K. (2010). "Mathematics of crossing fibers," in *Diffusion MRI: Theory, Methods, and Application*, ed. D. K. Jones (Oxford: Oxford University Press), 451–464. doi: 10.1093/med/9780195369779.003.0027

Andersson, J. L., Graham, M. S., Drobnjak, I., Zhang, H., Filippini, N., and Bastiani, M. (2017). Towards a comprehensive framework for movement and distortion correction of diffusion MR images: within volume movement. *Neuroimage* 152, 450–466. doi: 10.1016/j.neuroimage.2017.02.085

Andersson, J. L., Graham, M. S., Zsoldos, E., and Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage* 141, 556–572. doi: 10.1016/j.neuroimage.2016.06.058

Andersson, J. L., and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* 125, 1063–1078. doi: 10.1016/j.neuroimage.2015.10.019

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., et al. (2013). Video game training enhances cognitive control in older adults. *Nature* 501, 97–101. doi: 10.1038/nature12486

Annett, M. (1970). A classification of hand preference by association analysis. *Br. J. Psychol.* 61, 303–321. doi: 10.1111/j.2044-8295.1970.tb01248.x

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839.

Baddeley, A. D. (2002). Is working memory still working? *Eur. Psychol.* 7:85.

Bastiani, M., Cottaar, M., Fitzgibbon, S. P., Suri, S., Alfaro-Almagro, F., Sotiropoulos, S. N., et al. (2019). Automated quality control for within and between studies diffusion MRI data using a non-parametric framework for movement and distortion correction. *Neuroimage* 184, 801–812. doi: 10.1016/j.neuroimage.2018.09.073

Burzynska, A. Z., Nagel, I. E., Preuschhof, C., Li, S. C., Lindenberger, U., Bäckman, L., et al. (2011). Microstructure of frontoparietal connections predicts cortical responsivity and working memory performance. *Cereb. Cortex* 21, 2261–2271. doi: 10.1093/cercor/bhq293

Canter, D. (ed.) (1985). *Facet Theory: Approaches to Social Research*. Berlin: Springer Science & Business Media.

Chang, E. F., Raygor, K. P., and Berger, M. S. (2015). Contemporary model of language organization: an overview for neurosurgeons. *J. Neurosurg.* 122, 250–261. doi: 10.3171/2014.10.JNS132647

Charlton, R. A., Barrick, T. R., Lawes, I. N. C., Markus, H. S., and Morris, R. G. (2010). White matter pathways associated with working memory in normal aging. *Cortex* 46, 474–489. doi: 10.1016/j.cortex.2009.07.005

Cheng, Y., Wu, W., Feng, W., Wang, J., Chen, Y., Shen, Y., et al. (2012). The effects of multi-domain versus single-domain cognitive training in non-demented older people: a randomized controlled trial. *BMC Med.* 10:30. doi: 10.1186/1741-7015-10-30

Choi, J., Jeong, B., Polcari, A., Rohan, M. L., and Teicher, M. H. (2012). Reduced fractional anisotropy in the visual limbic pathway of young adults witnessing domestic violence in childhood. *Neuroimage* 59, 1071–1079. doi: 10.1016/j.neuroimage.2011.09.033

Cocchi, L., Zalesky, A., Fornito, A., and Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends Cogn. Sci.* 17, 493–501. doi: 10.1016/j.tics.2013.08.006

Cox, S. R., Ritchie, S. J., Tucker-Drob, E. M., Liewald, D. C., Hagenaars, S. P., Davies, G., et al. (2016). Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nat. Commun.* 7:13629.

Curran, K. M., Emsell, L., and Leemans, A. (2016). "Quantitative DTI measures," in *Diffusion Tensor Imaging*, eds W. Van Hecke, L. Emsell, and S. Sunaert (New York, NY: Springer), 65–87. doi: 10.1007/978-1-4939-3118-7_5

de Jonge, P., and de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Pers. Individ. Differ.* 21, 1007–1020. doi: 10.1016/s0191-8869(96)00161-4

D'Esposito, M., Postle, B. R., and Rypma, B. (2000). "Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies," in *Executive Control and the Frontal Lobe: Current Issues*, eds W. X. Schneider, A. M. Owen, and J. Duncan (Berlin: Springer), 3–11. doi: 10.1007/978-3-642-59794-7_2

Dima, D., Jogia, J., and Frangou, S. (2014). Dynamic causal modeling of load—dependent modulation of effective connectivity within the verbal working memory network. *Hum. Brain Mapp.* 35, 3025–3035. doi: 10.1002/hbm.22382

Dineen, R. A., Vilisaar, J., Hlinka, J., Bradshaw, C. M., Morgan, P. S., Constantinescu, C. S., et al. (2009). Disconnection as a mechanism for cognitive dysfunction in multiple sclerosis. *Brain* 132, 239–249. doi: 10.1093/brain/awn275

Duda, B. M., and Sweet, L. H. (2019). Functional brain changes associated with cognitive training in healthy older adults: a preliminary ALE meta-analysis. *Brain Imaging Behav.* 14, 1247–1262. doi: 10.1007/s11682-019-00080-0

Emch, M., von Bastian, C. C., and Koch, K. (2019). Neural correlates of verbal working memory: an fMRI meta-analysis. *Front. Hum. Neurosci.* 13:180. doi: 10.3389/fnhum.2019.00180

Finc, K., Bonna, K., He, X., Lydon-Staley, D. M., Kühn, S., Duch, W., et al. (2020). Dynamic reconfiguration of functional brain networks during working memory training. *Nat. Commun.* 11:2435.

Geda, Y. E., Roberts, R. O., Knopman, D. S., Christianson, T. J., Pankratz, V. S., Ivnik, R. J., et al. (2010). Physical exercise, aging, and mild cognitive impairment: a population-based study. *Arch. Neurol.* 67, 80–86.

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Guerin, S. A., and Miller, M. B. (2011). Parietal cortex tracks the amount of information retrieved even when it is not the basis of a memory decision. *Neuroimage* 55, 801–807. doi: 10.1016/j.neuroimage.2010.11.066

Guttman, L. (1954). "A new approach to factor analysis: the radex," in *Mathematical Thinking in the Social Sciences*, ed. P. F. Lazarsfeld (New York, NY: Free Press), 258–348.

Halford, G. S., Wilson, W. H., and Phillips, S. (1998). Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behav. Brain Sci.* 21, 803–831. doi: 10.1017/s0140525x98001769

Hampson, M., Driesen, N. R., Skudlarski, P., Gore, J. C., and Constable, R. T. (2006). Brain connectivity related to working memory performance. *J. Neurosci.* 26, 13338–13343. doi: 10.1523/jneurosci.3408-06.2006

Hebart, M. N., and Hesselmann, G. (2012). What visual information is processed in the human dorsal stream? *J. Neurosci.* 32, 8107–8109. doi: 10.1523/jneurosci.1462-12.2012

Herbet, G., Zemmoura, I., and Duffau, H. (2018). Functional anatomy of the inferior longitudinal fasciculus: from historical reports to current hypotheses. *Front. Neuroanat.* 12:77. doi: 10.3389/fnana.2018.00077

Hofstetter, S., and Assaf, Y. (2017). The rapid development of structural plasticity through short water maze training: a DTI study. *Neuroimage* 155, 202–208. doi: 10.1016/j.neuroimage.2017.04.056

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., et al. (2008). Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage* 39, 336–347. doi: 10.1016/j.neuroimage.2007.07.053

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage* 62, 782–790.

Jiang, L., Cao, X., Li, T., Tang, Y., Li, W., Wang, J., et al. (2016). Cortical thickness changes correlate with cognition changes after cognitive training: evidence from a chinese community study. *Front. Aging Neurosci.* 8:118. doi: 10.3389/fnagi.2016.00118

Jones, D. K., Knösche, T. R., and Turner, R. (2013). White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *Neuroimage* 73, 239–254. doi: 10.1016/j.neuroimage.2012.06.081

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., and Moore, K. S. (2008). The mind and brain of short-term memory. *Annu. Rev. Psychol.* 59, 193–224.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., et al. (1998). The role of parietal cortex in verbal working memory. *J. Neurosci.* 18, 5026–5034.

Jung, R. E., and Haier, R. J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* 30:135. doi: 10.1017/s0140525x07001185

Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., and Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: a systematic review and meta-analysis. *Ageing Res. Rev.* 15, 28–43. doi: 10.1016/j.arr.2014.02.004

Kinoshita, M., Nakajima, R., Shinohara, H., Miyashita, K., Tanaka, S., Okita, H., et al. (2016). Chronic spatial working memory deficit associated with the superior longitudinal fasciculus: a study using voxel-based lesion-symptom mapping and intraoperative direct stimulation in right prefrontal glioma surgery. *J. Neurosurg.* 125, 1024–1032. doi: 10.3171/2015.10.jns1591

Krogsrud, S. K., Fjell, A. M., Tamnes, C. K., Grydeland, H., Due-Tønnessen, P., Bjørnerud, A., et al. (2018). Development of white matter microstructure in relation to verbal and visuospatial working memory — A longitudinal study. *PLoS One* 13:e0195540. doi: 10.1371/journal.pone.0195540

Lampit, A., Hallock, H., Suo, C., Naismith, S. L., and Valenzuela, M. (2015). Cognitive training-induced short-term functional and long-term structural plastic change is related to gains in global cognition in healthy older adults: a pilot study. *Front. Aging Neurosci.* 7:14. doi: 10.3389/fnagi.2015.00014

Langer, N., von Bastian, C. C., Wirz, H., Oberauer, K., and Jäncke, L. (2013). The effects of working memory training on functional brain network efficiency. *Cortex* 49, 2424–2438. doi: 10.1016/j.cortex.2013.01.008

Latini, F., Mårtensson, J., Larsson, E. M., Fredrikson, M., Åhs, F., Hjortberg, M., et al. (2017). Segmentation of the inferior longitudinal fasciculus in the human brain: a white matter dissection and diffusion tensor tractography study. *Brain Res.* 1675, 102–115. doi: 10.1016/j.brainres.2017.09.005

Lautenschlager, N. T., Cox, K. L., Flicker, L., Foster, J. K., Van Bockxmeer, F. M., Xiao, J., et al. (2008). Effect of physical activity on cognitive function in older adults at risk for Alzheimer disease: a randomized trial. *JAMA* 300, 1027–1037. doi: 10.1001/jama.300.9.1027

Ma, L., Steinberg, J. L., Hasan, K. M., Narayana, P. A., Kramer, L. A., and Moeller, F. G. (2012). Working memory load modulation of parieto−frontal connections: evidence from dynamic causal modeling. *Hum. Brain Mapp.* 33, 1850–1867. doi: 10.1002/hbm.21329

Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., et al. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: a randomized, controlled study. *Proc. Natl. Acad. Sci.* 103, 12523–12528. doi: 10.1073/pnas.0605194103

Metzler-Baddeley, C., Caeyenberghs, K., Foley, S., and Jones, D. K. (2016). Task complexity and location specific changes of cortical thickness in executive and salience networks after working memory training. *Neuroimage* 130, 48–62. doi: 10.1016/j.neuroimage.2016.01.007

Mori, S., Crain, B. J., Chacko, V. P., and Van Zijl, P. C. (1999). Three−dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annal. Neurol.* 45, 265–269. doi: 10.1002/1531-8249(199902)45:2<265::aid-ana21>3.0.co;2-3

Nagy, Z., Westerberg, H., and Klingberg, T. (2004). Maturation of white matter is associated with the development of cognitive functions during childhood. *J. Cogn. Neurosci.* 16, 1227–1233. doi: 10.1162/0898929041920441

Nielsen, J. D., Madsen, K. H., Wang, Z., Liu, Z., Friston, K. J., and Zhou, Y. (2017). Working memory modulation of frontoparietal network connectivity in first-episode schizophrenia. *Cereb. Cortex* 27, 3832–3841. doi: 10.1093/cercor/bhx050

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., and Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Pers. Individ. Differ.* 29, 1017–1045. doi: 10.1016/s0191-8869(99)00251-2

Oberauer, K., Süß, H. M., Wilhelm, O., and Wittman, W. W. (2003). The multiple faces of working memory: storage, processing, supervision, and coordination. *Intelligence* 31, 167–193. doi: 10.1016/s0160-2896(02)00115-0

Oberauer, K., Süβ, H. M., Wilhelm, O., and Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence* 36, 641–652. doi: 10.1016/j.intell.2008.01.007

Oken, B. S., Flegal, K., Zajdel, D., Kishiyama, S., Haas, M., and Peters, D. (2008). Expectancy effect: impact of pill administration on cognitive performance in healthy seniors. *J. Clin. Exp. Neuropsychol.* 30, 7–17. doi: 10.1080/13803390701775428

Owen, A. M., McMillan, K. M., Laird, A. R., and Bullmore, E. (2005). N−back working memory paradigm: a meta−analysis of normative functional neuroimaging studies. *Hum. Brain Mapp.* 25, 46–59. doi: 10.1002/hbm.20131

Panesar, S. S., Yeh, F. C., Jacquesson, T., Hula, W., and Fernandez-Miranda, J. C. (2018). A quantitative tractography study into the connectivity, segmentation and laterality of the human inferior longitudinal fasciculus. *Front. Neuroanat.* 12:47. doi: 10.3389/fnana.2018.00047

Pickering, S. J. (2006). *Working Memory and Education*. San Diego, CA: Academic Press.

Pieramico, V., Esposito, R., Cesinaro, S., Frazzini, V., and Sensi, S. L. (2014). Effects of non-pharmacological or pharmacological interventions on cognition and brain plasticity of aging individuals. *Front. Syst. Neurosci.* 8:153. doi: 10.3389/fnsys.2014.00153

Román, F. J., Lewis, L. B., Chen, C. H., Karama, S., Burgaleta, M., Martínez, K., et al. (2016). Gray matter responsiveness to adaptive working memory training: a surface-based morphometry study. *Brain Struct. Funct.* 221, 4369–4382. doi: 10.1007/s00429-015-1168-7

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., et al. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis. *Neuroimage* 60, 830–846. doi: 10.1016/j.neuroimage.2011.11.050

Sagi, Y., Tavor, I., Hofstetter, S., Tzur-Moryosef, S., Blumenfeld-Katzir, T., and Assaf, Y. (2012). Learning in the fast lane: new insights into neuroplasticity. *Neuron* 73, 1195–1203. doi: 10.1016/j.neuron.2012.01.025

Salmi, J., Nyberg, L., and Laine, M. (2018). Working memory training mostly engages general-purpose large-scale networks for learning. *Neurosci. Biobehav. Rev.* 93, 108–122. doi: 10.1016/j.neubiorev.2018.03.019

Salminen, T. (2016). *On the Cognitive and Neuronal Effects and Mechanisms of Working Memory Training*. M.Sc. dissertation, Humboldt-Universität zu, Berlin.

Salminen, T., Mårtensson, J., Schubert, T., and Kühn, S. (2016). Increased integrity of white matter pathways after dual n-back training. *Neuroimage* 133, 244–250. doi: 10.1016/j.neuroimage.2016.03.028

Salthouse, T. A. (1990). Working memory as a processing resource in cognitive aging. *Dev. Rev.* 10, 101–124. doi: 10.1016/0273-2297(90)90006-p

Scharfen, J., Peters, J. M., and Holling, H. (2018). Retest effects in cognitive ability tests: a meta-analysis. *Intelligence* 67, 44–66. doi: 10.1016/j.intell.2018.01.003

Schmahmann, J. D., Smith, E. E., Eichler, F. S., and Filley, C. M. (2008). Cerebral white matter: neuroanatomy, clinical neurology, and neurobehavioral correlates. *Annal. N. Y. Acad. Sci.* 1142:266.

Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., and Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f2, a measure of local effect size, from PROC MIXED. *Front. Psycholo.* 3:111. doi: 10.3389/fpsyg.2012.00111

Sexton, C. E., Walhovd, K. B., Storsve, A. B., Tamnes, C. K., Westlye, L. T., Johansen-Berg, H., et al. (2014). Accelerated changes in white matter microstructure during aging: a longitudinal diffusion tensor imaging study. *J. Neurosci.* 34, 15425–15436. doi: 10.1523/jneurosci.0203-14.2014

Shah, P., and Miyake, A. (1999). "Models of working memory–An introduction," in *Models of working Memory: Mechanisms of Active Maintenance and Executive Control*, eds P. Shah and A. Miyake (Cambridge: Cambridge University Press).

Shen, M., Xu, H., Zhang, H., Shui, R., Zhang, M., and Zhou, J. (2015). The working memory Ponzo illusion: involuntary integration of visuospatial information stored in visual working memory. *Cognition* 141, 26–35. doi: 10.1016/j.cognition.2015.04.010

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Soares, J., Marques, P., Alves, V., and Sousa, N. (2013). A hitchhiker's guide to diffusion tensor imaging. *Front. Neurosci.* 7:31. doi: 10.3389/fnins.2013.00031

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., et al. (2010). Training of working memory impacts structural connectivity. *J. Neurosci.* 30, 3297–3303. doi: 10.1523/jneurosci.4611-09.2010

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., et al. (2011). Verbal working memory performance correlates with regional white matter structures in the frontoparietal regions. *Neuropsychologia* 49, 3466–3473. doi: 10.1016/j.neuropsychologia.2011.08.022

Tavor, I., Hofstetter, S., and Assaf, Y. (2013). Micro-structural assessment of short term plasticity dynamics. *Neuroimage* 81, 1–7. doi: 10.1016/j.neuroimage.2013.05.050

Tournier, J. D. (2010). "The biophysics of crossing fibres," in *Diffusion MRI: Theory, Methods, and Applications*, ed. D. K. Jones (New York, NY: Oxford University Press), 465–482.

Vestergaard, M., Madsen, K. S., Baaré, W. F., Skimminge, A., Ejersbo, L. R., Ramsøy, T. Z., et al. (2011). White matter microstructure in superior longitudinal fasciculus associated with spatial working memory performance in children. *J. Cogn. Neurosci.* 23, 2135–2146. doi: 10.1162/jocn.2010.21592

von Bastian, C. C., Langer, N., Jäncke, L., and Oberauer, K. (2013a). Effects of working memory training in young and old adults. *Memory Cogn.* 41, 611–624. doi: 10.3758/s13421-012-0280-7

von Bastian, C. C., Locher, A., and Ruflin, M. (2013b). Tatool: a java-based open-source programming framework for psychological studies. *Behav. Res. Methods* 45, 108–115. doi: 10.3758/s13428-012-0224-y

von Bastian, C. C., and Oberauer, K. (2013). Distinct transfer effects of training different facets of working memory capacity. *J. Mem. Lang.* 69, 36–58. doi: 10.1016/j.jml.2013.02.002

Wager, T. D., and Smith, E. E. (2003). Neuroimaging studies of working memory. *Cogn. Affect. Behav. Neurosci.* 3, 255–274.

Wakana, S., Jiang, H., Nagae-Poetscher, L. M., Van Zijl, P. C., and Mori, S. (2004). Fiber tract–based atlas of human white matter anatomy. *Radiology* 230, 77–87. doi: 10.1148/radiol.2301021640

Wilkinson, G. N., and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *J. R. Stat. Soc. Ser. C (Appl. Stati.)* 22, 392–399. doi: 10.2307/2346786

Wolinsky, F. D., Unverzagt, F. W., Smith, D. M., Jones, R., Stoddard, A., and Tennstedt, S. L. (2006). The ACTIVE cognitive training trial and health-related quality of life: protection that lasts for 5 years. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* 61, 1324–1329. doi: 10.1093/gerona/61.12.1324

World Health Organization (2015). *World Report on Ageing and Health*. Geneva: World Health Organization.

Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., and Feldman, H. M. (2012). Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS One* 7:e49790. doi: 10.1371/journal.pone.0049790

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 16 April 2021
doi: 10.3389/fnhum.2021.635687

Generalizing Longitudinal Age Effects on Brain Structure – A Two-Study Comparison Approach

Christiane Jockwitz[1,2], Susan Mérillat[3], Franziskus Liem[3], Jessica Oschwald[3], Katrin Amunts[1,4], Lutz Jäncke[3,5†] and Svenja Caspers[1,2†]*

[1] Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany, [2] Institute for Anatomy I, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University, Düsseldorf, Germany, [3] University Research Priority Program "Dynamics of Healthy Aging", University of Zurich, Zurich, Switzerland, [4] C. and O. Vogt Institute for Brain Research, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [5] Division of Neuropsychology, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Duygu Tosun, University of California, San Francisco, United States

Reviewed by:
Deborah Finkel, Indiana University Southeast, United States
Evan Fletcher, University of California Davis Medical School, United States

***Correspondence:**
Christiane Jockwitz
c.jockwitz@fz-juelich.de

† These authors have contributed equally to this work

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

Received: *30 November 2020*
Accepted: *03 March 2021*
Published: *16 April 2021*

Citation:
Jockwitz C, Mérillat S, Liem F, Oschwald J, Amunts K, Jäncke L and Caspers S (2021) Generalizing Longitudinal Age Effects on Brain Structure – A Two-Study Comparison Approach. Front. Hum. Neurosci. 15:635687. doi: 10.3389/fnhum.2021.635687

Cross-sectional studies indicate that normal aging is accompanied by decreases in brain structure. Longitudinal studies, however, are relatively rare and inconsistent regarding their outcomes. Particularly the heterogeneity of methods, sample characteristics and the high inter-individual variability in older adults prevent the deduction of general trends. Therefore, the current study aimed to compare longitudinal age-related changes in brain structure (measured through cortical thickness) in two large independent samples of healthy older adults (n = 161 each); the Longitudinal Healthy Aging Brain (LHAB) database project at the University of Zurich, Switzerland, and 1000BRAINS at the Research Center Juelich, Germany. Annual percentage changes in the two samples revealed stable to slight decreases in cortical thickness over time. After correction for major covariates, i.e., baseline age, sex, education, and image quality, sample differences were only marginally present. Results suggest that general trends across time might be generalizable over independent samples, assuming the same methodology is used, and similar sample characteristics are present.

Keywords: brain structure, aging, cognition, longitudinal change, old age, cortical thickess

INTRODUCTION

Normal aging can be accompanied by a decline in cognitive abilities (Hedden and Gabrieli, 2004) and changes in brain structure (Sowell et al., 2003). Both phenomena show high inter-individual variability, especially during later decades of life (Habib et al., 2007; Dickie et al., 2013). Results derived from cross-sectional studies have revealed a negative relationship between age and brain structure across adulthood, with differential effect sizes for specific brain regions (Fjell et al., 2009; Jockwitz et al., 2019), depending on the functional properties of the brain region of interest as well as the brain structure metric investigated (e.g., brain volume-based versus surface-based metrics or cortical thickness versus surface area) (O'Sullivan et al., 2001; Sowell et al., 2003; Salat et al., 2005; Walhovd et al., 2011; Ziegler et al., 2012; Dickie et al., 2013; Hogstrom et al., 2013; Fjell et al., 2014a,b; Liem et al., 2015).

While the associations between brain structure and age are rather heterogenous across studies, we recently showed consistent cross-sectional age associations for two different cohorts when

applying the same analysis protocol [e.g., age range, processing of the neuroimaging data (Jockwitz et al., 2019)]. At the same time, cross-sectional studies inherit a potential problem concerning the validity of inferences: Cross-sectional studies assess age-related differences between individuals, which is not comparable to age-related changes within individuals. One important disadvantage of cross-sectional studies concerns interindividual differences that might obscure intraindividual changes of aging (Raz and Lindenberger, 2011).

Longitudinal studies are still relatively rare and inconsistent with respect to their outcomes, preventing the deduction of general trends of age-related changes in brain structure. When comparing cross-sectional and longitudinal research designs, different patterns were shown for structural brain aging (Hedden and Gabrieli, 2004; Pfefferbaum and Sullivan, 2015). Large between-study heterogeneity of designs and methods, differences in sample characteristics and the generally larger inter-individual variability in samples of older adults make it difficult to extract general trends. However, general decreases in brain structure have been reported, although to a lesser degree than those reported in cross-sectional research designs [for a recent review, see Oschwald et al. (2019)].

To extract general age trends for brain structure, comparability between independent study samples is necessary. A few studies have already performed comparability analyses of cross-sectional age-related differences in brain structure metrics (i.e., brain volume or cortical thickness) between different samples, e.g., Fjell et al. (2009); Jockwitz et al. (2019). These studies indicate that general associations between age and brain structure are similar across independent samples, assuming that the same methodology and analysis protocol was used. However, such between study comparisons are lacking for investigations of longitudinal aging trajectories, especially in the older adult population, where inter-individual variability is particularly high. With the growing trend of large imaging consortia, e.g., UK Biobank (Miller et al., 2016), ENIGMA (Thompson et al., 2014), German National Cohort Study [NaKo; Bamberg et al. (2015)], or ADNI [Alzheimer's Disease Neuroimaging Initiative; Jack et al. (2008)] which aim at pooling datasets from a variety of study centers to increase sample size and statistical power, it will be crucial to establish the validity of age-related changes in brain structure. Therefore, the current study aimed to compare longitudinal age-related changes in brain structure in two large independent samples of healthy older adults: The Longitudinal Healthy Aging Brain (LHAB) database project at the University of Zurich (Switzerland; Zollig et al. (2011)] and 1000BRAINS at the Research Centre Juelich (Germany; Caspers et al. (2014)].

MATERIALS AND METHODS

Participants included in the current research project were recruited from two longitudinal studies investigating brain-behavior relationships in older adults located in the larger Zurich area (Switzerland) and in the Ruhr district (Germany).

The first sample comprised the ongoing LHAB database project at the University Research Priority Program (URPP) "Dynamics of Healthy Aging" of the University of Zurich (Zollig et al., 2011). LHAB investigates age-related dynamics of brain-behavior relationships in healthy older adults. A particular focus is placed on assessing and explaining interindividual variability in the observed aging trajectories. For this purpose, a broad spectrum of factors assumed to influence such trajectories (e.g., lifestyle, sleep, and nutrition) is collected. In LHAB, older adults from Zurich and surrounding areas are observed longitudinally with between-measurement intervals of one to 2 years. Inclusion criteria for study participation at baseline were age $\geq$ 64, right-handedness, fluent German language proficiency, a score of $\geq$ 26 on the Mini Mental State Examination [MMSE; Folstein et al. (1975)], no self-reported neurological disease of the central nervous system and no contraindications to MRI. The study was approved by the ethical committee of the canton of Zurich. Participation was voluntary and all participants gave written informed consent in accordance with the declaration of Helsinki. The initial sample of LHAB was comprised of 232 participants ranging from 64 to 87 years of age. Data acquisition in the LHAB project started in 2011. Currently the dataset covers an observation period of 7 years.

The second sample comprised 1000BRAINS at the Institute of Neuroscience and Medicine, Research Centre Juelich. 1000BRAINS is a longitudinal population-based study that assesses variability in brain structure and function during aging with respect to various influencing factors (Caspers et al., 2014). The 1000BRAINS sample is drawn from the 10-year follow-up cohort of the Heinz Nixdorf Recall Study, an epidemiological population-based study of risk factors for atherosclerosis, cardiovascular disease, cardiac infarction, and death (Schmermund et al., 2002) and the affiliated MultiGeneration study. In 1000BRAINS, adults aged 55 and older (at baseline) from the Heinz Nixdorf Recall study and their relatives (spouses and offspring; sampled from MultiGeneration study) were recruited, and were examined two times over a period of about 3 to 4 years. In contrast to the LHAB study, inclusion in the study was only dependent on the eligibility requirements for the MR acquisition based on the MR safety guidelines (e.g., stents and heart pacemakers led to exclusion from the study). The study protocol was approved by the University of Duisburg-Essen. Participation was voluntary and all participants gave written informed consent in accordance with the declaration of Helsinki. The initial sample of 1000BRAINS was comprised of 1,315 participants ranging from 18 to 87 years of age.

For the current study, we focused on two time points in both samples (LHAB: baseline and 4-year follow-up; 1000BRAINS: baseline and 3 to 4-years follow-up). Participants with missing values for the brain data were excluded. In order to assure comparability between the two samples, we matched them with respect to baseline age and sex using propensity score matching implemented in R (Stuart et al., 2011).

This resulted in 161 participants for each of the two final samples with the following demographic characteristics: LHAB: mean age = 69.9 $\pm$ 4.1; 85 females, mean interval = 4.2 $\pm$ 0.1; 1000BRAINS: mean age = 69.2 $\pm$ 4.6, 76 females, mean interval = 3.7 $\pm$ 0.7. For an overview of demographic variables

of the two samples at both timepoints, see **Table 1**. Education was measured according to the international classification of education (ISCED) and afterward divided into three educational classes: 1. school and/or vocational training, 2. grammar school or vocational baccalaureate, specialized secondary school/diploma, or commercial school degree, and 3. Bachelor, Master, Doctorate or equivalent.

Data Acquisition

For LHAB, anatomical T1-weighted images of both timepoints were acquired on a 3.0 T Philips Ingenia scanner (Philips Medical Systems, Best, The Netherlands). T1-weighted structural brain images were measured per visits with: TR = 8.18 ms, TE = 3.8 ms, Flip Angle = 8°, FoV = 240 mm × 240 mm, isotropic voxel size = 1 mm × 1 mm × 1 mm, 160 slices per volume. For 1000BRAINS, anatomical T1-weighted images of both timepoints were acquired on a 3.0 Tesla TIM-Trio MR scanner (Siemens Medical System, Erlangen, Germany). The T1-weighted structural brain images were scanned per visit with: TR = 2.25 s, TE = 3.03 ms, flip angle = 9°, FoV = 256 mm × 256 mm, voxel resolution = 1 mm × 1 mm × 1 mm, 176 slices per volume. In both studies, T1-imaging was part of a larger MR imaging protocol [see Caspers et al. (2014); Zollig et al. (2011)].

Preprocessing

Anatomical images from both samples were preprocessed using the same automated surface-based processing stream for longitudinal analyses of the FreeSurfer Software package [1000BRAINS: version 6.0.0; LHAB: FreeSurfer BIDS App v6.0.0-2; Gorgolewski et al. (2017)]. A detailed description of this pipeline is provided by Reuter et al. (2012); Dale et al. (1999), Fischl et al. (1999) as well as on http://surfer.nmr.mgh.harvard.edu. In short, first the cross-sectional surface reconstruction pipeline was applied to every subject, which includes (a) the segmentation of the structural brain images into gray matter, white matter, and cerebrospinal fluid, (b) motion correction, (c) intensity normalization, (d) transformation into Talairach space, (e) tessellation of the gray/white matter boundary, and (f) correction of topological defects. The gray/white matter interface was then (g) expanded to create the pial surface (boundary between gray matter and cerebrospinal fluid), which finally consists of about 150,000 vertices per hemisphere with an average surface area of 0.5 mm^2. Afterwards, each subject was preprocessed using the longitudinal surface reconstruction pipeline (Reuter et al., 2012) in which, based on the results of the cross-sectional preprocessing pipeline, a within-subject anatomical template was built across the two timepoints. Subsequently, cortical thickness was calculated based on the cross-sectional as well as longitudinal information from each subject. This procedure has previously been shown to be more sensitive in calculating surface-based brain metrics, since, due to the common template for the two timepoints, within-subject variability is reduced (Reuter et al., 2012). No manual correction of the reconstructed surfaces (white matter and pial surface) was performed in the two studies.

TABLE 1 | Demographics of the two samples and group comparisons (Independent *T*-test for continuous and Wilxon-Cox test for categorical variables) with corresponding *T/W* and *p*-values.

	1000BRAINS	LHAB	*T/W* (*P*-Values)
Age (TP1)	69.2 ± 4.6	69.9 ± 4.1	−1.39 (0.166)
Sex	0.53 ± 0.5	0.47 ± 0.5	13685 (0.317)
ISCED 3	2.0 ± 1.0	2.3 ± 0.8	11000 (0.010)
Age (TP2)	72.9 ± 4.7	74 ± 4.1	−2.28 (0.024)
Intervall (TP1 – TP2)	3.7 ± 0.7	4.2 ± 0.1	−8.02 (<0.001)

Regions of Interest

For the current study, we used the widely used Desikan-Killiany atlas (Desikan et al., 2006) as implemented in FreeSurfer to extract cortical thickness from left and right cortices. Specifically, for each of the 68 regions of interest (ROIs), mean cortical thickness was calculated as the average shortest distance between the white matter surface and the corresponding vertex within the respective ROIs on the pial surface.

Cognitive Performance

Participants from both LHAB and 1000BRAINS took part in a large neuropsychological assessment consisting of tests in the domains attention, executive functions, working memory, episodic memory and language functions. For comparison between the two samples, the following tasks were chosen: Trail Making Test A: processing speed, B: concept shifting; Morris et al. (1989), LPS50 + subtest three [reasoning; Sturm et al. (1993)] and [Regensburger Wortflüssigkeitstest (RWT), semantic condition (verbal fluency); Aschenbrenner et al. (2000)]. For descriptives of cognitive tasks, see **Table 2**.

Statistical Analysis

The purpose of the current research project was to compare intra-individual changes in brain structure (cortical thickness) across the ROIs of two independent population-based cohort studies. We calculated annual percentage changes to estimate yearly changes in cortical thickness and cognitive performance. Annual percentage changes were calculated as the following: $[(\text{Value at last measurement occasion in the study/Value at baseline})^{1/(totalyearsinstudy)} - 1] \times 100$. Positive values represent increases and negative values represent decreases. We next identified outliers for all annual percentage changes (mean annual percentage change ± 3 SD) and excluded those values that deviated more than 3 SD from the mean.

To examine whether the two samples showed similar changes in cortical thickness over time, we first used a one sample *t*-test to estimate general changes in cortical thickness for the two groups separately. To investigate whether the two samples differed concerning their variances, we conducted Levene's test for sample homogeneity. Finally, between sample differences in cortical thickness annual percentage changes were assessed using a General Linear Model (GLM) with cortical thickness as the dependent variable and sample and sex as fixed factors. Baseline age (TP1), education, and Euler number were included as covariates of non-interest. Euler number represents a marker of image quality that summarizes the topological complexity of the reconstructed cortical surface (Rosen et al., 2018).

TABLE 2 | Raw cognitive performance values for TP1 and 2, as well as the APC together with *T* and *p*-values for the APC (Sig. of APC; one sample *T*-test) and *F* and *p*-values for sample homogeneity (Levene's test).

	1000BRAINS				LHAB				
	Tp1	Tp2	APC	Sig. of APC	Tp1	Tp2	APC	Sig. of APC	Levene's test
Processing speed	40.22 ± 12.46	41.12 ± 14.12	0.34 ± 7.06	0.61 (0.54)	37.16 ± 12.90	39.37 ± 16.15	1.07 ± 6.88	1.93 (0.056)	0.25 (0.614)
Concept shifting	93.20 ± 41.55	96.87 ± 43.33	0.84 ± 7.98	1.32 (0.188)	86.69 ± 33.86	94.22 ± 39.77	2.04 ± 6.83	3.63 (<0.001)	2.40 (0.122)
Verbal fluency	23.96 ± 6.67	22.81 ± 6.73	−1.31 ± 5.76	−2.81 (0.006)	26.06 ± 6.46	25.98 ± 5.83	0.17 ± 4.41	0.47 (0.633)	9.59 (0.002)
Reasoning	20.99 ± 4.65	20.56 ± 5.42	−0.13 ± 5.14	−0.31 (0.757)	24.02 ± 4.45	26.48 ± 4.75	2.35 ± 3.70	7.99 (<0.001)	10.66 (0.001)

Subsequently, we assessed the cortical thickness annual percentage changes with the mentioned covariates (baseline age, sex, education, and Euler number) separately for the two samples to examine whether changes in cortical thickness would be driven by one sample. Finally, we additionally assessed the relation between annual percentage changes of cortical thickness and cognitive performance for the two samples separately.

RESULTS

When matching the two samples for baseline age and sex, the two samples did not differ in the respective variables (baseline age: $T = -1.39$, $p = 0.166$; and sex: $W = 13{,}685$, $p = 0.317$). However, we found significant differences in terms of education ($W = 11{,}000$, $p = 0.01$), with participants included in LHAB generally showing a higher formal education as compared to participants included in 1000BRAINS. Furthermore, the time intervals between the two measurements differed, with a longer interval between measurements in the LHAB project (1000BRAINS: 3.7 ± 0.7 years; LHAB: 4.2 ± 0.1 years; $T = -8.02$; $p < 0.001$; for group differences, see **Table 1**). To address this difference in time intervals we calculated annual percentage changes of cortical thickness. **Table 3** includes cortical thickness values for the two hemispheres at both timepoints as well as the annual percentage change in cortical thickness for the two samples separately (for all ROIs see **Supplementary Table 1**).

Cortical Thickness

With respect to cortical thickness, the LHAB sample showed slightly stronger annual percentage changes (i.e., decreases) in cortical thickness over time as compared to 1000BRAINS (see **Figures 1A,B**). On the other hand, we found 1000BRAINS to generally show more variance between participants regarding the annual percentage change in most of the ROIs (for Levene's test, **Supplementary Table 1**), although variances in mean CT did not differ significantly between the two samples (see **Table 3**). **Figure 1C** shows difference maps in terms of standard deviations of the annual percentage changes. For example, one of the most significant differences in standard deviations is observed in the right postcentral gyrus (see **Figure 1C** for a density plot; 1000BRAINS: SD = 0.7, LHAB: SD = 0.5; Levene's test: $F = 14.64$, $p < 0.001$).

Next, we again used GLMs to examine sample differences in annual percentage changes in cortical thickness with age, sex, education and Euler number as covariates (for all significant influences, see **Table 4** and **Supplementary Table 2**). Overall, after correcting for the different covariates and for multiple comparisons, only very few sample differences in terms of annual percentage change were present, i.e., inferior frontal gyrus pars triangularis (lh: $F = 13.67$, rh: $F = 16.54$) and inferior frontal gyrus pars opercularis (rh: $F = 21.43$) and transverse temporal gyrus (rh: $F = 20.47$).

In addition, after correcting for the above-mentioned variables, only a few regions showed significant intercepts (i.e., main effects of time), age effects or relations to sex, education or the Euler number (almost no effects did survive correction for multiple comparisons). **Figure 2** shows age-related annual percentage changes in cortical thickness for left and right hemispheres. As one can see in the two plots, the annual percentage change was not significantly related to baseline age for the left hemisphere ($F = 2.41$; $p = 0.121$) but was at trend level for the right hemisphere ($F = 4.95$; $p = 0.027$). The plots also show that the relationship between age and annual percentage change follows a linear, rather than a non-linear trend.

For a better understanding of the regional specificity of sample differences in the cortical thickness annual percentage changes, we projected the effect sizes (partial eta squared) of the sample differences onto the brains surface (**Figure 3**). Effect sizes ranged from 0 to 0.06, being interpreted as small to medium effects. Regarding the covariates, we only found sporadic effects on cortical thickness annual percentage change. After correcting for these subtle, mostly non-significant influences, and even the intercepts (i.e., main effects of annual percentage change) became non-significant. To verify that these influences were not driven by only one of the two samples, we further calculated the GLMs for the two samples separately (see **Supplementary Table 3**).

Finally, we assessed the relation between annual percentage changes of cortical thickness and cognitive performance for the two samples separately, which, after correcting for multiple comparisons, revealed non-significant results (see **Tables 2**, **5** and **Supplementary Table 4**).

DISCUSSION

Generalizability and replicability of age effects on brain and behavior are vital requirements to understand major aging mechanisms in our older adult population. The complexity of the aging process, in which the effect of single contributing factors, i.e., lifestyle or genetics, is assumed to be highly individual and rather small. To unravel even subtle brain-behavior relationships

TABLE 3 | Cortical thickness values for TP1 and 2, as well as the annual percentage change (APC) together with *T* and *p*-values for the APC (Sig. of APC; one sample *T*-test) and *F* and *p*-values for sample homogeneity (Levene's test).

	1000BRAINS				LHAB				
	Tp1	Tp2	APC	Sig. of APC	Tp1	Tp2	APC	Sig. of APC	Levene's test
Mean CT left	2.46 ± 0.09	2.45 ± 0.09	−0.15 ± 0.45	−4.17 (<0.001)	2.4 ± 0.08	2.37 ± 0.09	−0.29 ± 0.45	−8.21 (<0.001)	0.17 (0.677)
Mean CT right	2.46 ± 0.09	2.45 ± 0.10	−0.14 ± 0.40	−4.49 (<0.001)	2.41 ± 0.08	2.38 ± 0.09	−0.3 ± 0.42	−9.07 (<0.001)	0.19 (0.664)

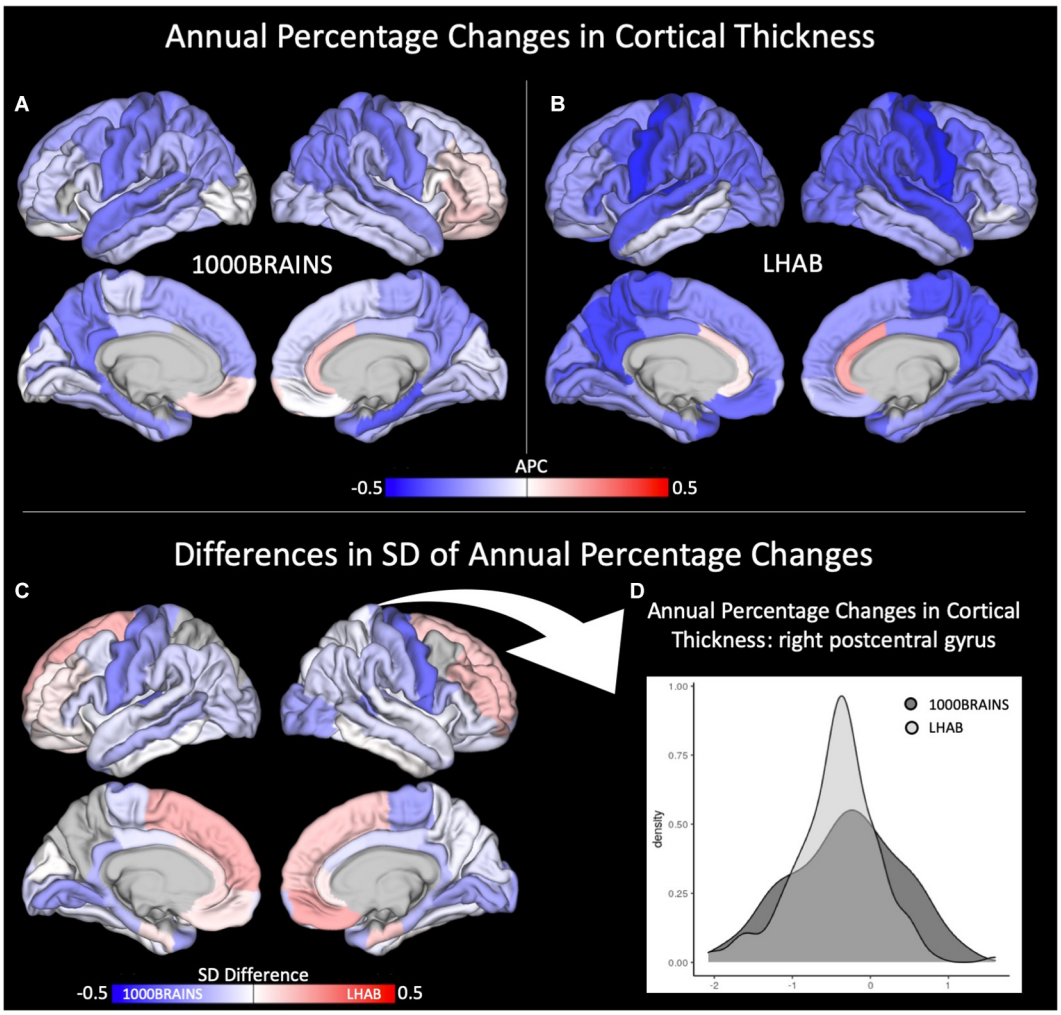


FIGURE 1 | Annual percentage changes (APC) in cortical thickness for **(A)** 1000BRAINS and **(B)** LHAB. Differences in SD between the two samples is shown in **(C)** together with a corresponding density plot **(D)** showing the variance in cortical thickness for 1000BRAINS and LHAB within the postcentral gyrus.

TABLE 4 | *F* and *p*-values derived from general linear models assessing annual percentage changes in cortical thickness in relation to sample, age, sex, education, and data quality (Euler number).

	Intercept	Age (TP1)	Sex	Education	Euler	Sample
Mean CT left	1.83 (0.177)	2.41 (0.121)	0.00 (0.966)	0.10 (0.756)	0.94 (0.334)	7.5 (0.007)
Mean CT right	4.35 (0.038)	4.95 (0.027)	0.44 (0.508)	0.95 (0.331)	0.32 (0.572)	8.85 (0.003)

during aging (Button et al., 2013; Wiseman et al., 2019) there is an upcoming trend of data pooling approaches to increase statistical power. However, data pooling procedures, particularly in imaging consortia, require proof of generalizability of observed age-related brain changes. The present study set out to meet this need and assessed age-related changes in brain structure

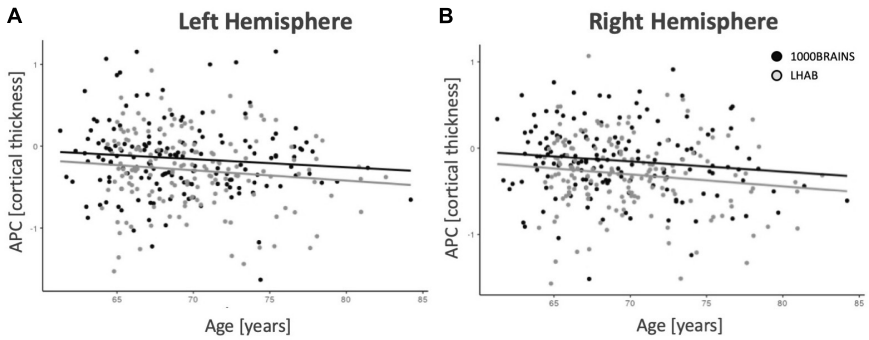


FIGURE 2 | Mean thickness annual percentage changes for the left **(A)** and right **(B)** hemispheres. With increasing age, there are slightly decreasing annual percentage changes for both samples.

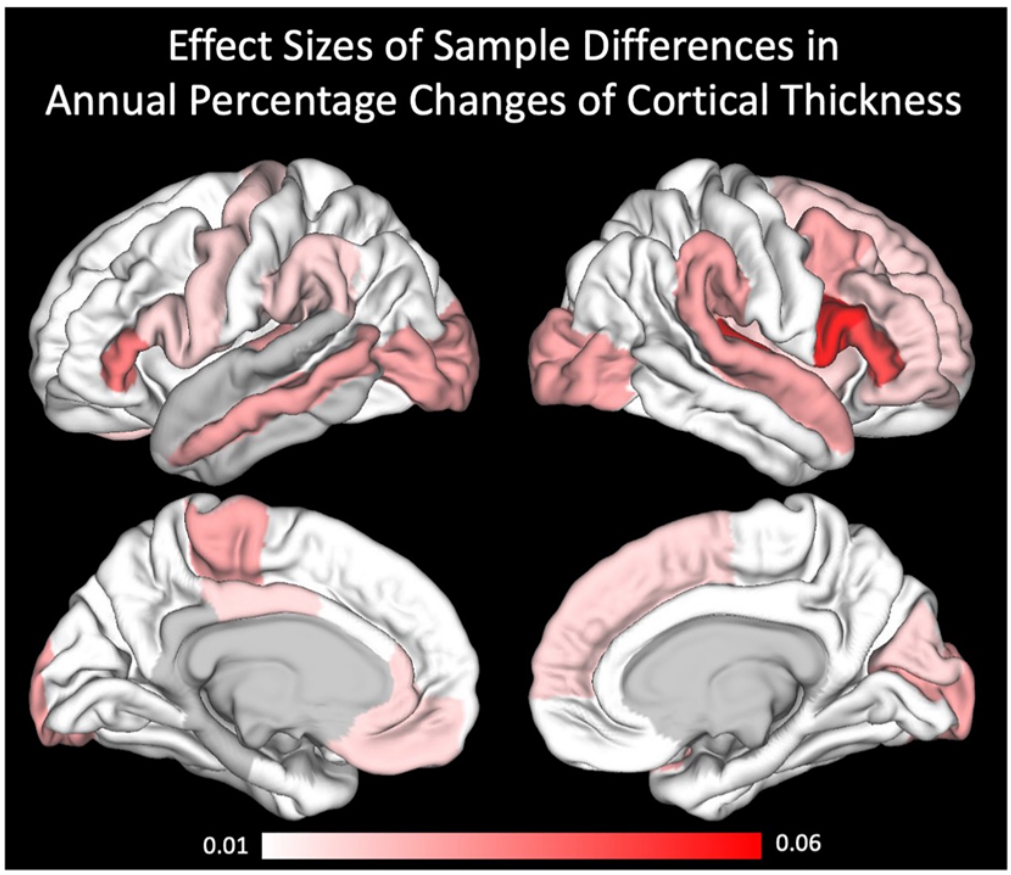


FIGURE 3 | Effect sizes of sample differences using partial eta square.

(measured by global and regional cortical thickness) in two closely matched samples of older adults over an average time period of three to four years. Despite significant differences in demographics between the two independent samples, we observed highly similar patterns of age-related changes in brain structure, when using the same methodology and analysis.

Cross-sectional age-related atrophy patterns have been reported by many previous studies (Walhovd et al., 2011; Storsve et al., 2014; Jancke et al., 2015). From those studies we would have expected to see a pattern of small but consistent cortical thickness decline in our two studies.

Overall, this decrease was found for both studies (cf. **Figure 1**) with participants included in the LHAB study showing a slightly more pronounced decline in cortical thickness. Highest annual percentage changes were found for pre- and postcentral gyri together with medial and lateral temporal and parietal

TABLE 5 | *F* and *p*-values derived from general linear models assessing the relation between annual percentage changes in cortical thickness with annual percentage changes in cognitive performance, calculated separately for the two samples, corrected for age, sex, education, and data quality (Euler number).

	Processing speed		Concept shifting		Verbal fluency		Reasoning	
	1000BRAINS	LHAB	1000BRAINS	LHAB	1000BRAINS	LHAB	1000BRAINS	LHAB
Mean CT left	0.21 (0.651)	5.45 (0.021)	1.27 (0.263)	0.31 (0.581)	0.26 (0.609)	0.00 (0.997)	2.40 (0.124)	1.03 (0.311)
Mean CT right	1.55 (0.215)	2.63 (0.107)	0.03 (0.864)	0.00 (0.971)	0.45 (0.505)	0.41 (0.522)	0.49 (0.484)	3.50 (0.063)

brain regions in both samples. In turn, the anterior cingulate cortex showed slight increases in cortical thickness over time. Importantly, the results are in line with previous longitudinal studies on cortical thickness investigating the whole adult lifespan (Storsve et al., 2014). Further, sample inhomogeneity testing revealed a higher between-subject variance for 1000BRAINS as compared to the LHAB study.

When adjusting the longitudinal effects of time for sex, education, baseline age and data quality (Euler number), only sporadic brain areas exhibited significant sample effects in annual percentage changes, i.e., left and right inferior frontal gyrus, pars triangularis, right inferior frontal gyrus pars opercularis and the right transverse temporal gyrus. Here, participants included in the LHAB study showed a more pronounced decrease over time. Based on sample characteristics, e.g., higher education in the LHAB sample, one would expect 1000BRAINS to show a more pronounced cortical thinning. However, especially for the inferior frontal gyrus (i.e., Broca's region involved in language functions), it has been shown that a higher brain reserve, in terms of higher gray matter volume, may diminish during the aging process, i.e., at older ages (Heim et al., 2019). If this holds true, then it might be the case that participants of the two samples assimilate during older ages in terms of brain structure. However, further research is necessary to unravel this complex relationship of age and brain structure.

Thus, the analysis of cortical thickness in two samples of healthy older adults revealed only marginal changes over time and only minimal sample differences. We are aware that our models include more covariate variables (age, sex, education, and data quality) than previous studies [e.g., Walhovd et al. (2011); Storsve et al. (2014); Thambisetty et al. (2010)]. We deliberately decided to include this set of variables since we know from previous research that cross-sectionally, the factors age, sex, education and data quality have an impact on brain structure (Sowell et al., 2003; Jancke et al., 2015; Jockwitz et al., 2019). Interestingly, when examining "raw annual percentage changes," these changes were partly in accordance with previous studies investigating changes in cortical thickness over time (Walhovd et al., 2011; Storsve et al., 2014). For example, Storsve found a mean annual percentage change of −0.35 in a sample ranging from 23 – 87 years and Fjell et al. (2014b) reported a mean annual percentage change of −0.59 in a sample of older adults. While we found a mean annual percentage change of −0.29 for the LHAB study, in 1000BRAINS this was slightly less pronounced, i.e., −0.15. In addition, we showed that the investigated covariates, i.e., baseline age, sex, education, and image quality, might be important in the investigation of longitudinal changes of brain structure. As an example, we found slightly negative relationships between baseline age and annual percentage changes in cortical thickness for the right hemisphere, which supports previous results (e.g., Fjell et al., 2009).

Finally, it has to be mentioned that neither of the two studies showed significant relations between annual percentage changes in cortical thickness and cognitive performance (i.e., processing speed, concept shifting, verbal fluency, and reasoning). First, these results complement previous results of our research group. In this cross-sectional study, no relation between cortical thickness and cognitive performance could be established in neither of the two study samples (Jockwitz et al., 2019). Likewise, other studies also revealed no associations between cognitive performance and particularly cortical thickness (in contrast to, e.g., brain volume [Cox et al., 2019], or white matter [Ziegler et al., 2012]). Furthermore, research regarding changes in both, brain structure and cognitive performance is quite heterogeneous. In the literature review of Oschwald et al. (2019) half of the studies revealed no association between changes in brain structure and cognitive performance, which fits to the current observation. In turn, those studies showing a significant association between changes in particularly cortical thickness and cognitive performance, differed from the current study. First, other cognitive functions were investigated, such as episodic memory or composite scores of executive functions (Fjell et al., 2014b; Möller et al., 2016; Sala-Llonch et al., 2017) and second, the above-mentioned studies included less or no covariates. Thus, when correcting for major confounding effects, cortical thickness changes were not related to cognitive performance changes over time. This is also well in line with the idea that in healthy older adults, correlations between changes in brain structure and simultaneous changes in cognitive performance are expectedly small and accompanied by high amounts of variability due to potential compensation mechanisms (Oschwald et al., 2019).

Methodological Considerations

The current study assessing longitudinal changes in brain structure has several advantages as well as limitations that we would like to address. With respect to the brain metric used in the current study, we chose cortical thickness, since it represents a prominent brain metric that seems to be sensitive to the aging process. However, it should also be mentioned that other metrics might be useful when comparing effects of aging, i.e., brain volume or gray matter density (Jäncke et al., 2019). Also, future studies may adopt Deformation-Field Morphometry

methods, such as Tensor-based morphometry (TBM), in order to compute longitudinal change in structural MRI data (Hua et al., 2008). Furthermore, with regard to the atlas used in the current study, i.e., Desikan-Killiany atlas, it needs to be stressed that other atlases might be more sensitive to functionally dependent changes in brain structure, such as the cytoarchitectonic Juelich Brain Atlas (Amunts et al., 2020) or functionally derived brain parcellations (Schaefer et al., 2018). In addition, future studies should also investigate longitudinal changes in brain structure and function with samples that are matched not only for age and gender, but also education or cognitive abilities. In the current study, we showed that covariates, such as age and education might explain small parts of the changes seen over time. Future studies should elaborate on these influencing factors to explore intra-individual aging processes.

CONCLUSION

Taken together, the current study showed that age-related changes in cortical thickness are relatively small, when adjusting for the most common influencing factors. This effect was seen in both independent studies, suggesting that general patterns of longitudinal changes in brain structure may be generalizable if the same methods are used and similar study populations with similar age and sex distributions are selected. However, fine-grained change patterns differ and the question whether results can be generalized over different samples cannot easily be answered because of the between-study differences regarding demographics (e.g., age ranges and education) or methodology (e.g., time intervals, different brain metrics, and such as brain volume versus cortical thickness). Furthermore, differences in covariates often hamper the extraction of generalizable age trends in different samples. With our study, we contribute to the field by showing that patterns of age-related changes in brain structure in two independent cohorts of older adults are highly similar when using the same methodological approach.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the used consent does not allow for the public sharing of the data. Requests to access the datasets should be directed to LJ, lutz.jaencke@uzh.ch (LHAB) and SC, s.caspers@fz-juelich.de (1000BRAINS).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of the Canton of Zurich, Switzerland (LHAB) and the University of Duisburg-Essen, Germany (1000BRAINS). The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SM and LJ contributed to the design, set-up, maintenance, and support of the LHAB project. SC and KA contributed to the design, set-up, maintenance, and support of the 1000BRAINS study. FL and CJ performed processing of the longitudinal neuroimaging data and wrote the first draft of the manuscript. CJ and JO performed the statistical analysis. LJ and SC supervised the project. All authors discussed the results, contributed to manuscript revision, and read and approved the submitted version.

FUNDING

The LHAB database project has received funding from the Velux Stiftung (project no. 369) and from the University Research Priority Program "Dynamics of Healthy Aging" of the UZH. Furthermore, this project was partially funded by the German National Cohort and the 1000BRAINS study of the Institute of Neuroscience and Medicine, Research Center Jülich, Germany. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 945539 (HBP SGA3; 897 KA, SC) as well as from the Initiative and Networking Fund of the Helmholtz Association (SC).

ACKNOWLEDGMENTS

The current analysis incorporates data from the Longitudinal Healthy Aging Brain (LHAB) database project carried out at the University of Zurich (UZH). The following researchers at the UZH were involved in the design, set up, maintenance, and support of the LHAB database: Anne Eschen, LJ, FL, Mike Martin, SM, Christina Röcke, and Jacqueline Zöllig. The authors would like to thank the Heinz Nixdorf Foundation (Germany) for the generous support of the Heinz Nixdorf Study. The authors would also like to thank the investigative group and the study staff of the Heinz Nixdorf Recall Study and 1000BRAINS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.635687/full#supplementary-material

REFERENCES

Amunts, K., Mohlberg, H., Bludau, S., and Zilles, K. (2020). Julich-brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* 369, 988–992.

Aschenbrenner, A., Tucha, O., and Lange, K. (2000). *RWT Regensburger Wortflüssigkeits-Test*. Göttingen.

Bamberg, F., Kauczor, H.-U., Weckbach, S., Schlett, C. L., Forsting, M., Ladd, S. C., et al. (2015). Whole-body MR imaging in the German National Cohort:

rationale, design, and technical background. *Radiology* 277, 206–220. doi: 10.1148/radiol.2015142272

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Caspers, S., Moebus, S., Lux, S., Pundt, N., Schutz, H., Muhleisen, T. W., et al. (2014). Studying variability in human brain aging in a population-based German cohort-rationale and design of 1000BRAINS. *Front. Aging Neurosci.* 6:149. doi: 10.3389/fnagi.2014.00149

Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M., and Deary, I. J. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence* 76:101376. doi: 10.1016/j.intell.2019.101376

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Dickie, D. A., Job, D. E., Gonzalez, D. R., Shenkin, S. D., Ahearn, T. S., Murray, A. D., et al. (2013). Variance in brain volume with advancing age: implications for defining the limits of normality. *PLoS One* 8:e84093.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999). Cortical surface-based analysis. II: inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207. doi: 10.1006/nimg.1998.0396

Fjell, A. M., McEvoy, L., Holland, D., Dale, A. M., Walhovd, K. B., and Alzheimer's Disease Neuroimaging Initiative, (2014a). What is normal in normal aging? Effects of aging, amyloid and Alzheimer's disease on the cerebral cortex and the hippocampus. *Prog. Neurobiol.* 117, 20–40. doi: 10.1016/j.pneurobio.2014.02.004

Fjell, A. M., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., et al. (2009). High consistency of regional cortical thinning in aging across multiple samples. *Cereb. Cortex* 19, 2001–2012. doi: 10.1093/cercor/bhn232

Fjell, A. M., Westlye, L. T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., et al. (2014b). Accelerating cortical thinning: unique to dementia or universal in aging? *Cereb Cortex* 24, 919–934. doi: 10.1093/cercor/bhs379

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capota, M., Chakravarty, M. M., et al. (2017). BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol.* 13:e1005209. doi: 10.1371/journal.pcbi.1005209

Habib, R., Nyberg, L., and Nilsson, L. G. (2007). Cognitive and non-cognitive factors contributing to the longitudinal identification of successful older adults in the betula study. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* 14, 257–273. doi: 10.1080/13825580600582412

Hedden, T., and Gabrieli, J. D. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci.* 5, 87–96. doi: 10.1038/nrn1323

Heim, S., Stumme, J., Bittner, N., Jockwitz, C., Amunts, K., and Caspers, S. (2019). Bilingualism and "brain reserve": a matter of age. *Neurobiology of Aging* 81, 157–165. doi: 10.1016/j.neurobiolaging.2019.05.021

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

Hua, X., Leow, A. D., Parikshak, N., Lee, S., Chiang, M. C., Toga, A. W., et al. (2008). Alzheimer's Disease Neuroimaging Initiative. Tensor-based morphometry as a neuroimaging biomarker for Alzheimer's disease: an MRI study of 676 AD, MCI, and normal subjects. *Neuroimage*. 43, 458–469. doi: 10.1016/j.neuroimage.2008.07.013

Jack, C. R. Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691. doi: 10.1002/jmri.21049

Jancke, L., Merillat, S., Liem, F., and Hanggi, J. (2015). Brain size, sex, and the aging brain. *Hum. Brain Mapp.* 36, 150–169. doi: 10.1002/hbm.22619

Jäncke, L., Liem, F., and Merillat, S. (2019). Scaling of brain compartments to brain size. *Neuroreport.* 30, 573–579. doi: 10.1097/WNR.0000000000001249

Jockwitz, C., Mérillat, S., Liem, F., Oschwald, J., Amunts, K., Caspers, S., et al. (2019). Generalizing age effects on brain structure and cognition: a two-study comparison approach. *Hum. Brain Mapp.* 40, 2305–2319.

Liem, F., Merillat, S., Bezzola, L., Hirsiger, S., Philipp, M., Madhyastha, T., et al. (2015). Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. *Neuroimage* 108, 95–109. doi: 10.1016/j.neuroimage.2014.12.035

Miller, K. L., Alfaro-Almagro, F., Bangerter, N. K., Thomas, D. L., Yacoub, E., Xu, J., et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat. Neurosci.* 19, 1523–1536. doi: 10.1038/nn.4393

Möller, C., Hafkemeijer, A., Pijnenburg, Y. A. L., Rombouts, S. A. R. B., van der Grond, J., Dopper, E., et al. (2016). Different patterns of cortical gray matter loss over time in behavioral variant frontotemporal dementia and Alzheimer's disease. *Neurobiol. Aging* 38, 21–31. doi: 10.1016/j.neurobiolaging.2015.10.020

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., et al. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD-NP). Part 1. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 39, 1159–1165. doi: 10.1212/WNL.39.9.1159

O'Sullivan, M., Jones, D. K., Summers, P. E., Morris, R. G., Williams, S. C., and Markus, H. S. (2001). Evidence for cortical "disconnection" as a mechanism of age-related cognitive decline. *Neurology* 57, 632–638.

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Rocke, C., et al. (2019). Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev. Neurosci.* 31, 1–57. doi: 10.1515/revneuro-2018-0096

Pfefferbaum, A., and Sullivan, E. V. (2015). Cross-sectional versus longitudinal estimates of age-related changes in the adult brain: overlaps and discrepancies. *Neurobiol. Aging* 36, 2563–2567. doi: 10.1016/j.neurobiolaging.2015.05.005

Raz, N., and Lindenberger, U. (2011). Only time will tell: cross-sectional studies offer no solution to the age-brain-cognition triangle: comment on Salthouse (2011). *Psychol. Bull.* 137, 790–795. doi: 10.1037/a0024503

Reuter, M., Schmansky, N. J., Rosas, H. D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* 61, 1402–1418. doi: 10.1016/j.neuroimage.2012.02.084

Rosen, A. F. G., Roalf, D. R., Ruparel, K., Blake, J., Seelaus, K., Villa, L. P., et al. (2018). Quantitative assessment of structural image quality. *Neuroimage* 169, 407–418. doi: 10.1016/j.neuroimage.2017.12.059

Sala-Llonch, R., Idland, A. V., Borza, T., Watne, L. O., Wyller, T. B., Brækhus, A., et al. (2017). Inflammation, amyloid, and atrophy in the aging brain: relationships with longitudinal changes in cognition. *J. Alzheimer's Dis.* 58, 829–840. doi: 10.3233/JAD-161146

Salat, D. H., Tuch, D. S., Hevelone, N. D., Fischl, B., Corkin, S., Rosas, H. D., et al. (2005). Age-related changes in prefrontal white matter measured by diffusion tensor imaging. *Ann. N. Y. Acad. Sci.* 1064, 37–49.

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., et al. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* 28, 3095–3114. doi: 10.1093/cercor/bhx179

Schmermund, A., Möhlenkamp, S., Stang, A., Grönemeyer, D., Seibel, R., Hirche, H., et al. (2002). Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL study. *Am. Heart J.* 144, 212–218. doi: 10.1067/mhj.2002.123579

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., and Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat. Neurosci.* 6, 309–315.

Storsve, A. B., Fjell, A. M., Tamnes, C. K., Westlye, L. T., Overbye, K., Aasland, H. W., et al. (2014). Differential longitudinal changes in cortical thickness, surface area and volume across the adult life span: regions of accelerating and decelerating change. *J. Neurosci.* 34, 8488–8498. doi: 10.1523/JNEUROSCI.0391-14.2014

Sturm, W., Willmes, K., and Horn, W. (1993). *Leistungsprüfungssystem für 50-90jährige*. Handanweisung, 2nd. Göttingen: Hogrefe.

Stuart, E. A., King, G., Imai, K., and Ho, D. (2011). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Stat. Softw.* 42, 1–28.

Thambisetty, M., Wan, J., Carass, A., An, Y., Prince, J. L., and Resnick, S. M. (2010). Longitudinal changes in cortical thickness associated with

normal aging. *Neuroimage* 52, 1215–1223. doi: 10.1016/j.neuroimage.2010.04.258

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The ENIGMA consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Walhovd, K. B., Westlye, L. T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., et al. (2011). Consistent neuroanatomical age-related volume differences across multiple samples. *Neurobiol. Aging* 32, 916–932. doi: 10.1016/j.neurobiolaging.2009.05.013

Wiseman, S. J., Meijboom, R., Valdes Hernandez, M. D. C., Pernet, C., Sakka, E., Job, D., et al. (2019). Longitudinal multi-centre brain imaging studies: guidelines and practical tips for accurate and reproducible imaging endpoints and data sharing. *Trials* 20:21. doi: 10.1186/s13063-018-3113-6

Ziegler, G., Dahnke, R., Jäncke, L., Yotter, R. A., May, A., and Gaser, C. (2012). Brain structural trajectories over the adult lifespan. *Hum. Brain Mapp.* 33, 2377–2389. doi: 10.1002/hbm.21374

Zollig, J., Merillat, S., Eschen, A., Rocke, C., Martin, M., and Jancke, L. (2011). Plasticity and imaging research in healthy aging: core ideas and profile of the international normal aging and plasticity imaging center (INAPIC). *Gerontology* 57, 190–192. doi: 10.1159/000324307

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 22 June 2021
doi: 10.3389/fnhum.2021.621263

Fractional Anisotropy in Selected, Motor-Related White Matter Tracts and Its Cross-Sectional and Longitudinal Associations With Motor Function in Healthy Older Adults

Jessica Oschwald[1], Susan Mérillat[1], Lutz Jäncke[1,2] and Rachael D. Seidler[3]*

[1] University Research Priority Program "Dynamics of Healthy Aging", University of Zurich, Zurich, Switzerland, [2] Department of Neuropsychology, Psychological Institute, University of Zurich, Zurich, Switzerland, [3] Department of Applied Physiology and Kinesiology, University of Florida, Gainesville, FL, United States

OPEN ACCESS

Edited by:
Ariel Rokem,
University of Washington,
United States

Reviewed by:
Stephanie Lefebvre,
University of Bern, Switzerland
Hakon Grydeland,
University of Oslo, Norway

****Correspondence:***
Jessica Oschwald
jessica.oschwald@gmail.com

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

Received: *25 October 2020*
Accepted: *03 May 2021*
Published: *22 June 2021*

Citation:
Oschwald J, Mérillat S, Jäncke L and Seidler RD (2021) Fractional Anisotropy in Selected, Motor-Related White Matter Tracts and Its Cross-Sectional and Longitudinal Associations With Motor Function in Healthy Older Adults. Front. Hum. Neurosci. 15:621263. doi: 10.3389/fnhum.2021.621263

Background: While it is well-known that deficits in motor performance and brain structural connectivity occur in the course of healthy aging, it is still unclear if and how these changes are related to each other. While some cross-sectional studies suggest that white matter (WM) microstructure is positively associated with motor function in healthy older adults, more evidence is needed. Moreover, longitudinal data is required to estimate whether similar associations can be found between trajectories of change in WM microstructure and motor function. The current study addresses this gap by investigating age-associations and longitudinal changes in WM microstructure and motor function, and the cross-sectional (level-level) and longitudinal (level-change, change-change) association between these two domains.

Method: We used multiple-occasion data (covering 4 years) from a large sample (N = 231) of healthy older adults from the Longitudinal Healthy Aging Brain (LHAB) database. To measure WM microstructure, we used diffusion-weighted imaging data to compute mean FA in three selected WM tracts [forceps minor (FMIN); superior longitudinal fasciculus (SLF); corticospinal tract (CST)]. Motor function was measured via two motor speed tests (grooved pegboard, finger tapping) and one motor strength test (grip force test), separately for the left and the right hand. The statistical analysis was conducted with longitudinal growth curve models in the structural equation modeling framework.

Results: The results revealed longitudinal decline and negative cross-sectional age-associations for mean WM FA in the FMIN and SLF, and for motor function in all tests, with a higher vulnerability for left than right hand motor performance. Regarding cross-domain associations, we found a significant positive level-level correlation among mean WM FA in the FMIN with motor speed. Mean FA in SLF and CST was not correlated with motor performance measures, and none of the level-change or change-change

associations were significant. Overall, our results (a) provide important insights into aging-related changes of fine motor abilities and FA in selected white matter tracts associated with motor control, (b) support previous cross-sectional work showing that neural control of movement in older adults also involves brain structures outside the core motor system and (c) align with the idea that, in healthy aging, compensatory mechanisms may be in place and longer time delays may be needed to reveal level-change or change-change associations.

Keywords: white matter microstructure, motor function, longitudinal, correlated change, healthy aging, fractional anisotropy, structural equation modeling (SEM), latent growth curve model (LGCM)

INTRODUCTION

Life expectancy has risen steadily due to innovations in medicine and improved living standards. In 2015, life expectancy at birth exceeded 80 years in 22 European countries (World Health Organization, [WHO], 2016). Globally, it is estimated to increase by a further 6 years until 2050 (United Nations, 2017). Understanding how the central nervous system changes with age contribute to declines in function is critically important for enhancing productivity and quality of life for this aging population. It is well known that aging is associated with degeneration of the central nervous system and decreases in motor performance (Seidler et al., 2010). To date, however, work in this area has been largely cross-sectional and more focused on regional measures of brain structure and function rather than network connectivity (Oschwald et al., 2019a). The brain's network structure underlies neural communication and functional activity; thus, studying how it changes over time may provide key insights into age-related functional declines.

Diffusion-weighted MRI (DW-MRI) allows investigation of structural integrity of the brain's white matter (WM) connectivity pathways. This technique is sensitive to diffusion of water molecules, which is spatially bounded by large WM tracts in the brain. Cross-sectional DW-MRI studies generally report lower fractional anisotropy (FA) in older individuals (reviewed in Oschwald et al., 2019a). A few longitudinal DW-MRI studies have been conducted; similar to what has been reported with other imaging modalities, prefrontal WM exhibits accelerated declines relative to other areas of the brain (cf. Barrick et al., 2010; Sullivan et al., 2010; Teipel et al., 2010) while sensorimotor WM exhibits less change (de Groot et al., 2016). In contrast, a recent, large ($n > 900$) cross-sectional study challenges the notion that sensorimotor regions exhibit reduced aging effects relative to more anterior prefrontal cortex; Taubert et al. (2020) found disproportionately reduced brain volume, iron, and myelin in the pre- and postcentral gyri in older individuals. This study did not evaluate WM tracts, however, leaving open the questions of how sensorimotor WM tracts change over time and whether such changes are correlated with motor function.

Whether or not the sensorimotor fibers are spared with age, there is certainly evidence of age effects on motor function. Gait and balance (Studenski et al., 2011), grip force (Bohannon, 2008), and other activities in everyday life decline with age and impact quality of life. Interestingly, performance of these behaviors is associated with prefrontal activity (Heuninckx et al., 2008; Seidler et al., 2010; Carson, 2018) as well as with prefrontal (Verlinden et al., 2016; Moscufo et al., 2018; Massa et al., 2019) and corpus callosum (Fling et al., 2011; Fling and Seidler, 2012). WM integrity in older adults, potentially reflecting compensation (Heuninckx et al., 2008). However, more evidence on the association between WM connectivity and motor function in healthy older adults is needed. Importantly, longitudinal data are required to more precisely delineate the trajectories of decline and to better understand if the associations between WM microstructure and motor behavior, and particularly associations of their changes, are reflective of compensation, maintenance, or other patterns in healthy older adults (Zahodne and Reuter-Lorenz, 2019).

In the current study, we leverage data from the Longitudinal Healthy Aging Brain (LHAB) database to evaluate cross-domain associations between brain WM microstructure and measures of manual motor function (motor strength: grip force; motor speed: grooved pegboard test and tapping speed) in healthy older adults. The LHAB database project is currently conducted at the University Research Priority Program (URPP) "Dynamics of Healthy Aging" of the University of Zurich (Zöllig et al., 2011). Our analyses include data that were acquired at four time points spanning over 4 years.

We used latent growth curve models (LGC) estimated in the structural equation modeling framework (SEM) to examine change in WM microstructure and motor function, as well as cross-sectional, and longitudinal associations among the two domains.

LGC is a statistical technique for the analysis of longitudinal data (McArdle and Epstein, 1987; McArdle, 2009; for a tutorial see Ghisletta and McArdle, 2012). LGC models estimate longitudinal growth processes as latent (i.e., unobserved) variables, including a latent intercept which reflects the initial level of a variable of interest (e.g., WM microstructural properties or motor performance at baseline) and a latent slope, which reflects the rate of change in this variable over time. An advantage of LGC models over traditional regression models is that besides such average (i.e., fixed) effects, they can capture interindividual variances (i.e., random effects) in intraindividual change. Of specific interest in the present study, two univariate LGCs (i.e., LGC that estimate the growth process in one variable) can be combined into a bivariate LGC to model parallel change processes, including cross-domain associations between baseline

levels of two variables (level-level), baseline level in one variable and changes in the other (level-change) and between two change processes (change-change) (Oschwald et al., 2019a). Importantly, advanced statistical techniques such as LGC are required to appropriately estimate within-person change (King et al., 2018), and to disentangle the complex longitudinal associations between changes in multiple variables – both questions of pivotal importance in the field of aging neuroscience.

Cross-sectional age-associations and longitudinal decline was estimated in the forceps minor (FMIN), the superior longitudinal fasciculus (SLF), and the corticospinal tract (CST). The CST is the main motor control projection tract. It plays a critical role in fine motor control of hand and finger movements (cf. Isa, 2012) such as those required for the tasks studied here. The FMIN connects the two prefrontal cortices via the anterior corpus callosum, whereas the SLF connects prefrontal cortex largely with parietal regions. While all three tracts have been implicated in motor function (Farbota et al., 2012; Henley et al., 2014; Wang et al., 2016; Reid et al., 2017; Giacosa et al., 2019; Maltais et al., 2020), the SLF and FMIN largely support executive functions such as attention and working memory (Mamiya et al., 2018; Nakajima et al., 2020) and do not have direct projections to spinal motor neurons. Thus, here, we refer to the CST as belonging to the motor system, and FMIN and SLF as being outside of the motor system. To measure WM microstructure, we chose to follow a region-of-interest (ROI) based approach in the present study, averaging WM FA across the selected fiber tracts, since the extraction of ROIs (as opposed to voxel-wise estimates) enabled us to estimate parallel change in the SEM framework. We used FA as an index of WM microstructure as it is a comparatively well-researched metric and provides a general estimate of the change in the WM fiber organization (see Jones et al., 2013, for considerations on the interpretation of FA).

Age-associations and change in motor performance was assessed with finger tapping, pegboard performance and grip force given that performance in these tests has been previously shown to decline with age. Based on the findings of de Groot et al. (2016), we hypothesized that the FMIN and SLF would exhibit greater change over time than the CST, and would be more correlated with changes in manual motor function. Furthermore, we hypothesized that baseline levels and longitudinal WM microstructural change in the tracts of interest would be most correlated with left hand and weakly or not correlated with right hand motor performance levels/change, since right-hand performance is well-trained and might be more adept at compensation.

MATERIALS AND METHODS

Participants

Longitudinal motor and MRI data were taken from the Longitudinal Healthy Aging Brain (LHAB) database (Zöllig et al., 2011). We used data from the first four measurement occasions (baseline, 1-year follow-up, 2-year follow-up, 4-year follow-up). The baseline dataset included 232 participants (*M* age = 70.8; range: 64–87; females: 114). At each measurement occasion, participants completed an extensive battery of neuropsychological and psychometric cognitive and motor assessments and underwent brain imaging. The brain imaging session was conducted in close temporal proximity to the behavioral assessments [difference between behavioral and MRI assessments in days (M ± SD): baseline: 2.2 ± 5.2, 1-year follow-up: 2.6 ± 5.2, 2-year follow-up: 4.3 ± 13.0, 4-year follow-up: 4.6 ± 9.3]. Inclusion criteria for study participation at baseline were age ≥ 64, right-handedness, fluent German language proficiency, a score of ≥26 on the Mini Mental State Examination (MMSE; Folstein et al., 1975), no self-reported neurological disease of the central nervous system and no contraindications to MRI. The study was approved by the Ethics Committee of the Canton of Zurich. Participation was voluntary and all participants gave written informed consent in accordance with the declaration of Helsinki. Self-reported physical and mental health of the sample at baseline, as measured by the SF-12 (Ware et al., 1996), were 50.9 ± 7.4 (M ± SD) and 54.8 ± 6.3, respectively, which indicates above-average health compared to a normative population (Ware et al., 1995). As expected, sample means for these general health indicators slightly declined over time, but still indicated above-average health at 4-year follow-up (physical health score: 50.5 ± 6.9, mental health score: 53.1 ± 8.0, MMSE = 28.3 ± 1.3). At 4-year follow-up, the dataset still comprised 74.57% of the baseline sample (n = 173). As reported in other publications with this sample (Oschwald et al., 2019b; Malagurski et al., 2020), selectivity analyses showed that the participants remaining in the study at the 4-year follow-up did not substantially differ from the baseline sample in terms of age, education, physical and mental health, or head motion in the scanner.

For the present analysis, participants were excluded if either motor behavior or DW-MRI data were missing for all measurement occasions. With this criterion we were able to include 231 participants from the LHAB baseline sample (*M* age at baseline = 70.8; females: 113). Of those 231 participants, 172 were still participating at the 4-year follow-up. Participant characteristics at each measurement occasion are presented in **Table 1**.

Brain Measures

MR Imaging

MRI measurements were conducted on a Philips Ingenia 3T scanner equipped with a commercial 32-element sensitivity encoding (SENSE) head coil array. The DW-MRI protocol employed an echo-planar (EPI) sequence [TR = 23.918 s, TE = 55 ms, FoV = 224 × 224 mm, acquisition matrix = 112 × 112, slice thickness = 2 mm, 75 contiguous slices, 2 mm^3 isotropic voxel, flip angle = 90°, Echo Train Length (ETL) = 59, NSA = 1, SENSE factor R = 2.0]. One non-weighted image (b-value = 0 s/mm^2) and 32 diffusion-weighted images (b-value of 1,000 s/mm^2) were acquired. The diffusion-weighted directions were equally distributed in space. The same scanner and sequence were used at all measurement occasions.

TABLE 1 | Participant characteristics of the full sample at baseline and at each follow-up wave.

Variable	Baseline (n = 231)			1-year follow-up (n = 210)			2-year follow-up (n = 196)			4-year follow-up (n = 172)		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Baseline age (years)	231	70.82	5.08	210	70.92	5.15	196	70.64	4.80	172	70.12	4.43
Gender (m/f)	231	118/113	–	210	109/101	–	196	105/91	–	172	93/79	–
Education (1–3)	224	2.23	0.86	209	2.24	0.86	194	2.23	0.87	170	2.28	0.84
Mental health	211	54.78	6.26	194	54.60	6.40	183	54.54	6.26	158	54.68	5.74
Physical health	211	50.85	7.37	194	50.97	7.37	183	51.11	6.86	158	51.52	6.32
Head motion[a]	228	0.24	0.15	206	0.25	0.16	189	0.27	0.17	164	0.26	0.19

m, male; f, female.
Education was measured on a scale from 1 to 3 (1 = high school with or without vocational education, 2 = higher education entrance qualification, business school or university of applied sciences, or 3 = university degree). Mental and physical health scores were computed based on the SF12 questionnaire, which participants filled out at home (Ware et al., 1996).
[a]*Head motion was assessed at each measurement occasion. All other variables were assessed at baseline.*

MRI Data Preprocessing

To facilitate analysis, data were organized according to the brain imaging data structure (BIDS) (Gorgolewski et al., 2016). Diffusion data were processed with a nipype pipeline (v0.14.0) (Gorgolewski et al., 2011) using tools from MRtrix (3-rc2) (Tournier et al., 2012), FSL (v5.0.9) (Jenkinson et al., 2012), and ANTs (2.1.0) (Avants et al., 2011). The analysis code is publicly available: https://github.com/fliem/extract_FA, and a BIDS-Apps-compatible (Gorgolewski et al., 2017) software container to reproduce the analysis can be found here: http://hub.docker.com/r/fliem/extract_fa/.

The diffusion data were denoised (Veraart et al., 2016a,b) and corrected for eddy current distortions and head motion (Andersson and Sotiropoulos, 2016; Andersson et al., 2016). Subsequently, the data were bias-corrected (Tustison et al., 2010) and a white matter mask was created (Dhollander et al., 2016). Tensor maps were calculated (Veraart et al., 2013) and FA maps were derived (Basser et al., 1994; Westin et al., 1997). ANTs was used to register FA maps to the JHU-ICBM-FA template (included in FSL). Mean FA was extracted for tracts of the JHU WM tractography atlas (thresholded at 25% probability) for voxels with FA $>$ 0.2 (Hua et al., 2008). The tracts considered here are: Forceps minor (FMIN), left and right hemispheric superior longitudinal fasciculi (SLF) and corticospinal tracts (CST). The size of the FMIN was 19407 voxels. For the SLF and CST, we averaged left and right hemispheres, weighted by the total number of voxels of the respective tract (SLF: 17657, CST: 10739). For the statistical analyses, we multiplied the FA values by 100 (now ranging from 0 to 100) to ensure a more intuitive interpretability of these scores for the reader and avoid any model estimation problems that might occur when bringing two domains together that are on very different scales.

Head Motion Control

As a means of ensuring sufficient quality of the data, we removed FA values for 56 individual observations (i.e., 7% of the total N of 787 observations) at which participants showed excessive head motion. As a measure of head motion, we used the summary statistic of average RMS motion as compared to the previous slice in a volume, which was calculated during preprocessing (Andersson and Sotiropoulos, 2016). Excessive values were defined as any value more than three median absolute deviations (MADs) above the median of the sample distribution across measurement occasions (Leys et al., 2013). We used the median as a reference, since it is more robust to the influence of extreme values than the mean.

Motor Performance Measures

Motor performance was assessed by three motor tests described in detail below. Specifically, we used two tests to measure motor speed and one test to measure motor strength. First, all dependent values of interest were individually standardized by using the mean and standard deviation of the first measurement occasion, and then transformed into a T-statistic to achieve better interpretability.

Motor Speed

Grooved pegboard test

The grooved pegboard test (Merker and Podell, 2010) comprises a board with a cavity in the upper region, in which small metal sticks in key-like shapes (pegs) are stored. In the bottom area, the board has 25 holes (5 columns $\times$ 5 holes), each including a slit on the side to fit the key-shape of the pegs. However, these holes are oriented in a random fashion, such that the matching of a peg to the hole might require a turn of the peg. The test requires participants to sequentially place pegs into the corresponding holes in the board as quickly as possible. Participants are asked to complete the test with their left hand (proceeding from right to left) and with the right hand (proceeding from left to right). The dependent measure is the time the participant needs to place all pegs into the holes, separately measured for the left and the right hand.

Finger tapping

Finger tapping speed was assessed with the MLS (Motorische Leistungsserie; Schoppe, 1974). This test is based on a work panel that allows participants to perform a series of different uni- or bimanual tasks designed to assess fine motor skills. Specifically, the finger tapping task required participants to tap as quickly as possible on one of two small squares located on the bottom left and right of the panel with the tip of a pen. Performance

was separately measured for the right and the left hand and participants were asked to target the square that was located on the same side as their respective hand. The dependent measure of interest was the number of taps within 32 s, separately measured for the left and the right hand.

Motor Strength

Grip force

Grip force was measured with a hydraulic hand dynamometer (Merker and Podell, 2010) (Model SH5001, Sae-han Corporation, Korea) to assess the isometric maximum grip force of the right and the left hand. Participants were instructed to sit upright, with their feet positioned flat on the ground, their shoulders and forearm in a neutral position, their elbow in 90° flexion and the wrist in an extension of 0° to 30°. Participants were asked to press the dynamometer for 4 s with maximum force, beginning with the dominant hand. After a 30 s break, they had to switch hands and repeat the task with the non-dominant hand. Overall, three repetitions were conducted for each hand. If the maximum grip force in one hand was higher in the third as compared to the first two rounds, data collection was continued until the final measurement was smaller than the previous one. The dependent measure of interest was the average grip force across the three highest measurements, separately computed for the left and the right hand.

Covariates

To control for potential confounding influences, we included age at baseline (Age_{base}), level of education (on a scale from 1 to 3; 1 = high school with or without vocational education, 2 = higher education entrance qualification, business school or university of applied sciences, or 3 = university degree) and gender (0 = female, 1 = male) as covariates on the intercept and slope terms in all statistical analyses. Furthermore, for the FA models, we also included head motion in the scanner as a time-varying covariate on the manifest indicators at each measurement occasion. To facilitate model interpretation, age was centered at 70 years (median of the sample), and education at level 2. Head motion was left uncentered, since a value of zero was meaningful (i.e., reflecting no head motion).

Statistical Analysis

All statistical analyses were run in R version 3.3.3 (R Core Team, 2019). Outlier correction in each motor performance measure was done using a cut-off of three MADs above or below the median of the sample distribution across measurement occasions, resulting in the removal of 58 individual values [i.e., between 0.1% and 3.2-% of the total *N* of observations for each test; Grooved Pegboard left: *n* (%) = 25 (3.2%) and right *n* (%) = 21 (2.7%); Finger tapping left: *n* (%) = 3 (0.4%) and right *n* (%) = 7 (0.9%); Grip Force left: *n* (%) = 1 (0.1%) and right *n* (%) = 1 (0.1%)]. We refrained from outlier control in FA measures, since FA can largely vary between individuals (e.g., Veenith et al., 2013), and no clear consensus exists on normative cut-offs. However, we excluded individual observations with excessive head motion values (see section on "Head motion control" above).

Latent Growth Curve Modeling (LGC)

In the present study, we first used univariate LGC to model change in FA and motor performance measures individually, and bivariate LGC to model cross-domain interactions between these measures. We estimated the LGC models in the SEM framework using the *lavaan* package version 0.5-23.1097 (Rosseel, 2012) in R.

Univariate models

We estimated separate univariate LGC models for FA in each of the WM tracts and for motor function in the tests assessed. For each univariate LGC we estimated the (1) initial level, i.e., the intercept of the measure of interest, (2) its rate of (linear) change, i.e., the slope, and (3) the association between the initial level and the rate of change of the measure of interest. To avoid confounding of initial level/rate of change estimation, we added baseline age and gender as covariates to all models. For the univariate FA models, we additionally added head motion as a time-varying covariate. To illustrate, **Figure 1** shows a path diagram of a univariate LGC model for motor strength.

From bottom to top, the Figure shows the observed measurements of grip force at each time point ($Grip_0 \ldots Grip_{4yr}$). These measurements load on the latent estimates of motor strength ($Ms_0 \ldots Ms_{4yr}$). In other words, the motor strength variables represent latent estimations of grip force at each measurement occasion, separated from measurement error. Further, a latent intercept (I) and slope factor (S) was estimated on top of the latent estimates of motor strength, to capture initial levels and overall rate of change across time. The means of these factors reflect the average baseline value (μ_I) and change (μ_S) in a variable across the entire sample (i.e., fixed effects). The variances of these latent factors reflect the variability between persons (i.e., random effects) in their individual baseline values ($\sigma^2{}_I$) and change trajectories ($\sigma^2{}_S$). We estimated the loadings of the change slope to reflect a linear change trajectory (i.e., slope loadings of 0, 1, 2, 4). As is the standard in longitudinal SEM modeling, we treated missing values as missing at random (MAR; Little, 1995) and retained them in the model by using the full information maximum likelihood estimation (FIML; Finkbeiner, 1979; Schafer and Graham, 2002) to deal with incomplete data.

Bivariate models

We estimated a series of bivariate LGC models, combining the univariate LGC models for each of the WM tracts and for each of the motor function tests, resulting in 18 models. To illustrate, **Figure 2** shows a path diagram of a bivariate LGC model for motor strength (i.e., grip force test) and WM FA in the SLF.

In each bivariate LGC model, we estimated the following cross-domain interactions between FA and motor function (blue pathways in **Figure 2**): (1) the level-level (i.e., intercept-intercept) correlation, to investigate the association between baseline FA values in each of the WM tracts and baseline motor function in each of the motor function tests. As an example, a positive level-level association would suggest that at study initiation, individuals with higher WM FA show better motor performance than individuals with lower WM FA values. (2) The level-change (i.e., intercept-slope) correlation to examine the association between baseline FA

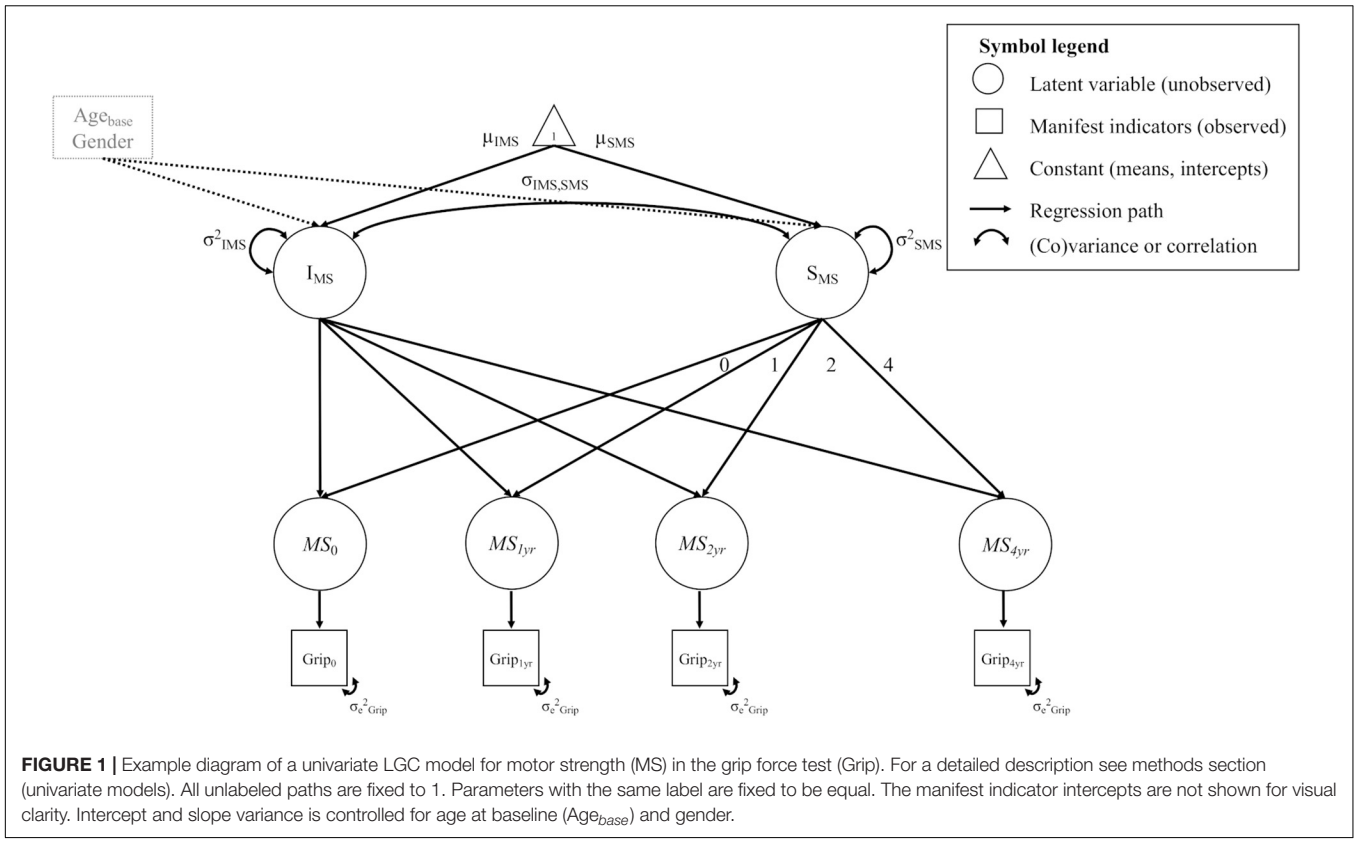


FIGURE 1 | Example diagram of a univariate LGC model for motor strength (MS) in the grip force test (Grip). For a detailed description see methods section (univariate models). All unlabeled paths are fixed to 1. Parameters with the same label are fixed to be equal. The manifest indicator intercepts are not shown for visual clarity. Intercept and slope variance is controlled for age at baseline (Age$_{base}$) and gender.

values and motor performance changes. As an example, a positive level-change association would indicate that people with higher WM FA at baseline show reduced motor performance declines over the study period (3) the change-change (i.e., slope-slope) correlation, to investigate the association between change in WM FA and change in motor function. As an example, a positive change-change association would mean that people with steeper declines in WM FA also show steeper declines in motor performance over the study period. Note that these examples are formulated based on the assumption that individuals experience declines in WM FA and motor performance over time.

We only evaluated these longitudinal cross-domain effects, if there was sufficient variance in the intercepts and slopes of the respective univariate models of the combined variables. We defined sufficient variance as intercept or slope variance that is significantly ($p < .05$) different from zero, suggesting that individuals show substantial heterogeneity with regards to their baseline levels or their longitudinal change trajectories in WM FA or motor performance. While intercept variance was substantial in all univariate models, this was not the case for slope variance in some of the models. As a significance test relies on an arbitrary cut-off, we additionally calculated effective curve reliability (ECR) for each univariate model. ECR is a reliability index for LGC model slope variance that can be interpreted as a standardized effect size statistic of the slope variance (Kelley and Preacher, 2012; Brandmaier et al., 2018). It is computed as the slope variance scaled as a proportion of the sum of slope variance and slope measurement error. ECR ranges from 0 to 1, with larger values reflecting increased true population slope variance and/or increased study design precision (and thus reduced effective error) (Brandmaier et al., 2018).

Evaluation of model fit

Overall model fit was evaluated by the χ^2 test, specifically, by the ratio of the χ^2 test statistic to the respective degrees of freedom (Jöreskog and Sörbom, 1993). Furthermore, the Comparative Fit Index (CFI; Bentler, 1990), and the root mean square error of approximation (RMSEA; Steiger and Lind, 1980) were used to evaluate goodness-of-fit. Good model fit was defined as a ratio of $\chi^2/df \leq 2$, CFI > 0.97, RMSEA ≤ 0.05, and adequate fit was defined as $\chi^2/df \leq 3$, CFI > 0.95, RMSEA between 0.05 and 0.08 (see Jöreskog and Sörbom, 1993; Schermelleh-Engel et al., 2003). Models were compared using the difference χ^2 test (for nested models) and the sample size adjusted Bayesian Information Criterion (BIC; Raftery, 1995). The BIC is not interpretable in isolation, however, in model comparisons, smaller values indicate a closer fit of the model to the data (Kass and Raftery, 1995; Raftery, 1995). Given that we tested a large number of hypotheses, we applied a correction for multiple comparisons using the False Discovery Rate (FDR) correction, to reduce the likelihood of false positive findings, i.e., Type 1 errors (Benjamini and Hochberg, 1995). The FDR correction was applied to each LGC model separately, and across all effects of interest within the structural part of the model. For the univariate LGC models, we included the means and variances of intercept and slope,

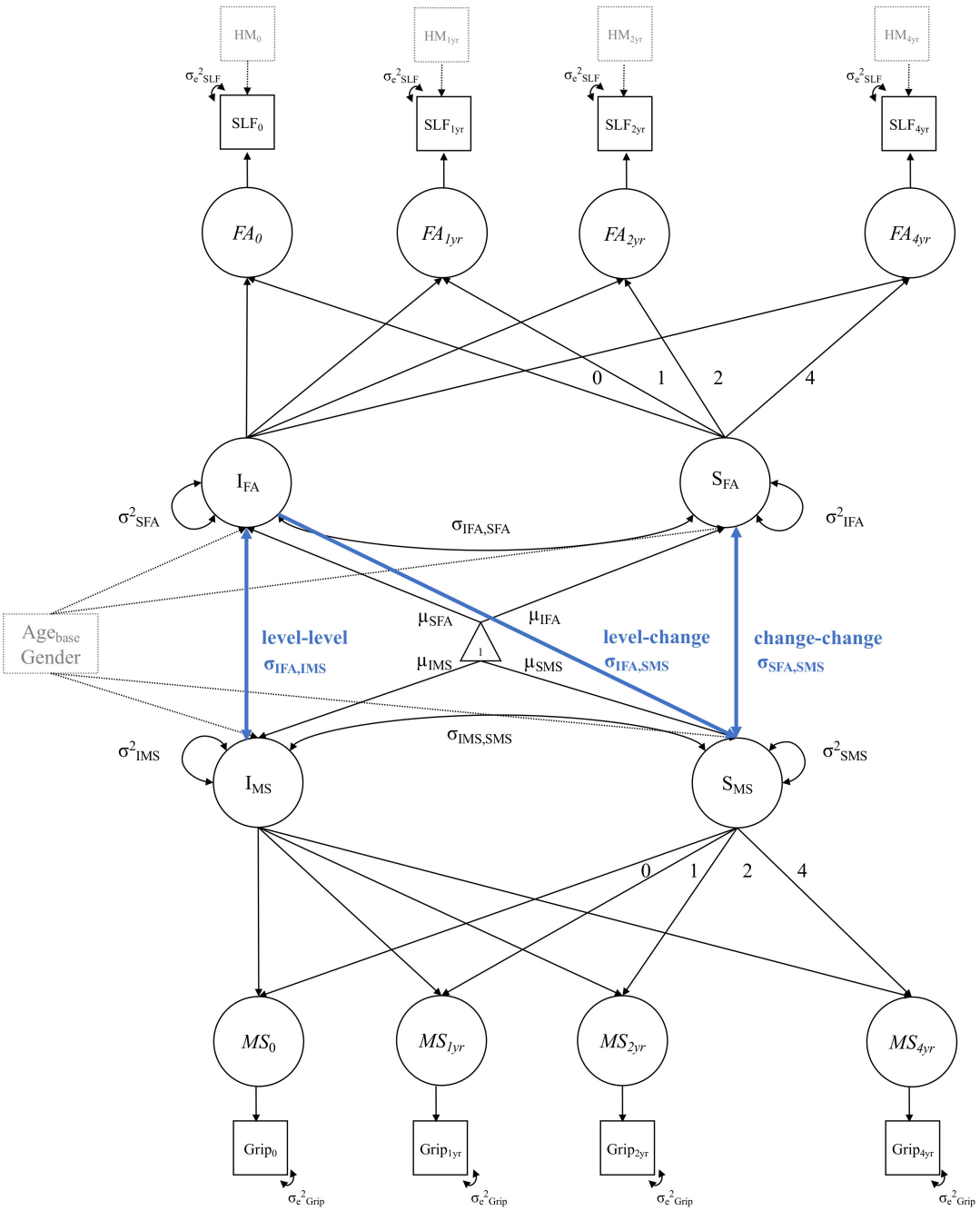


FIGURE 2 | Example diagram of a bivariate LGC model for motor strength (MS) in the grip force test (Grip) and WM FA in the SLF. Blue paths illustrate the cross-domain associations between (1) initial WM FA and motor performance at study baseline (level–level), (2) initial WM FA and subsequent change in motor performance (level-change), change in WM FA and change in motor performance (change–change). Single-headed arrows represent regression effects and double-headed arrows represent (co)variances and correlations. Circles represent latent, unobserved variables and squares represent manifest, observed variables. Triangles stand for constants, such as means and intercepts. All unlabeled paths are fixed to 1. Parameters with the same label are fixed to be equal. The manifest indicator intercepts are not shown for visual clarity. Intercept and slope variance for motor strength and WM FA is controlled for age at baseline (Age_{base}) and gender. The observed variables of WM FA in the SLF are controlled for head motion (HM) at each measurement occasion.

covariate effects and within-domain correlations of intercept and slope as effects of interest. For the bivariate LGC models, we only considered cross-domain correlations between intercept and slopes as effects of interest.

RESULTS

Raw FA declined annually for the FMIN and SLF by −0.20 ± 1.18% and −0.56 ± 0.88%, respectively. In the

CST, annual increases were observed in average FA values ($0.25 \pm 0.92\%$). This effect, however, did not manifest itself in the univariate LGC models described below. Potential explanations for this divergence between raw change and modeled change values will be addressed in the discussion. Average annual declines were observed in all motor function tests [annual percentage change (APC) ranging between –0.24% and 1.21%]. Grip force and tapping speed of the left hand declined more strongly compared to the right (dominant) hand, while we observed the reversed pattern (higher annual decline in the right hand) for performance in the pegboard test. **Table 2** contains a detailed overview of the raw annual change and APC in FA and motor functioning.

Univariate LGC Models: FA

We fit univariate LGC models for mean FA in each of three WM tracts (FMIN, SLF, CST). Model fit statistics and parameter estimates are shown in **Table 3**.

Fit statistics ranged from adequate to good for all models [$\chi^2_{(24)}$ = 31.733–67.697, χ^2/df = 1.32–2.82, RMSEA = 0.037–0.089, CFI = 0.960–0.988]. With regards to the parameter estimates, substantial individual differences (i.e., significant intercept variance) were observed for baseline mean FA in all WM tracts. On average, mean scaled FA declined in the FMIN and SLF (–0.15 and –0.20 per year). In contrast, as we hypothesized, no substantial average FA changes were observed in the CST. Moreover, substantial interindividual differences in longitudinal change (i.e., significant slope variance) were observed for the FMIN and SLF. However, the univariate LGC model for the CST initially converged with a negative slope variance. Constraining the CST slope variance to a positive value resulted in the estimation of a slope variance of zero (see **Table 3**), suggesting the absence of interindividual differences in longitudinal average FA change trajectories for the CST. Accordingly, ECR could not be calculated for the CST. As sufficient variance in change trajectories is necessary to estimate cross-domain parallel change correlations, only the FMIN and SLF were retained for the bivariate modeling of change-change associations. ECR for the FMIN and SLF were 0.44 and 0.30 (see **Table 3**), respectively, suggesting a small to medium effect size of the slope variance for these tracts (Cohen, 1988, 1992). Covariance between intercept and slope was not significant for any of the WM tracts.

With regard to covariate effects on intercept and slope (see **Table 4**), baseline age was significantly associated with baseline mean FA in the FMIN and SLF, in the direction that older individuals had lower FA values in these tracts at baseline than their younger peers. Baseline age was also significantly related to changes in mean FA over time in the FMIN and CST, suggesting that participants showed accelerated annual FA decline with increasing age (mean scaled FA of –0.01 and –0.02). Specifically, as no significant main effect of mean FA change was observed for the CST, this result suggests that FA decline was predominantly observed for the older participants in the sample. Gender had a significant effect on average baseline FA in the FMIN, and FA changes in the CST: Male participants had lower mean FA in the FMIN at baseline (–0.64, SE = 0.29, p = 0.026), and showed less annual mean FA change in the CST (0.22, SE = 0.06, $p < 0.001$) than female participants.

Finally, we also investigated the impact of head motion at each measurement occasion (see **Table 5**). Head motion was significantly associated with FA in all WM tracts on at least one measurement occasion, such that more motion in the scanner was associated with significantly lower average FA. This effect was most consistently observed for the CST. Note that one unit increase in head motion amounts to almost five times the average head motion present in the sample (from 0.24 to 0.27 across measurement occasions; cf. **Table 1**).

Univariate LGC Models: Motor Function

We fit univariate LGC models to estimate longitudinal performance change in each of the motor function tests (motor speed: pegboard, tapping; motor strength: grip force), and separately for each hand. Model fit statistics and parameter estimates of these models are presented in **Table 6**.

Fit statistics ranged from adequate to good for all models [$\chi^2_{(12)}$ = 15.972–31.070, χ^2/df = 1.33–2.59, RMSEA = 0.038–0.083, CFI = 0.980–0.993]. Substantial interindividual differences were observed for motor performance at baseline in all tests. On average, motor performance declined in all tests (between –0.39 and –0.56 per year) but one (Pegboard left hand performance). Mirroring the raw performance changes reported in **Table 2**, larger declines were observed for left hand than right hand motor function. Regarding longitudinal change trajectories, substantial interindividual variance was only observed for motor strength in the grip force test (left and right hand). In contrast, the slope variance was not significantly different from zero for any of the motor speed tests[1]. Mirroring this result from the significance test, ECR and therefore slope variance effect size was below the cut-off of a small effect size for the motor speed tests (between 0.03–0.15) and of medium size for the Grip Force test (left hand: 0.47, right hand: 0.60) (Cohen, 1988, 1992). Due to the lack of substantial change variability in the motor speed tests, only the grip force test (left and right hand) will be retained for the estimation of change-change associations in the bivariate modeling section. The covariance between intercept and slope was significant only in the grip force test (left hand: –1.15, SE = 0.53, p = 0.029, right hand: –1.57, SE = 0.60, p = 0.009), in the direction that individuals with higher motor function at baseline tended to show accelerated declines in motor function over 4 years.

Baseline age was significantly negatively associated with motor function across all tests, such that older individuals had a lower performance at baseline than younger peers (**Table 7**). In addition, baseline age was significantly negatively associated with

[1]In case of the tapping test (left hand), the univariate LGC model did initially converge with a negative slope variance. Constraining the slope variance to a positive value still did not result in an ideal solution, as it produced a warning message (covariance matrix of latent variables non-positive definite). Since these issues were most likely produced by the lack of sufficient slope variance, we added further constraints to the model, fixing the intercept-slope covariance to zero. This resulted in convergence without any warning messages, and a good model fit (see **Table 6**).

TABLE 2 | Raw annual change and annual percentage change of WM microstructure and motor function.

Variable/change score	Raw annual change			APC (%)	
	n	*M*	*SD*	*M*	*SD*
WM microstructure (raw FA scores × 10^2)					
FMIN	197	−0.07	0.42	−0.20	1.18
SLF	197	−0.20	0.31	−0.56	0.88
CST	197	0.13	0.50	0.25	0.92
Motor function (standardized *T*-scores)					
Pegboard (left)	202	−0.22	2.79	−0.24	7.16
Pegboard (right)	206	−0.46	2.96	−0.78	7.38
Tapping (left)	205	−0.45	2.37	−1.09	5.07
Tapping (right)	204	−0.32	2.53	−0.67	6.02
Grip force (left)	210	−0.63	1.63	−1.21	3.68
Grip force (right)	210	−0.49	1.74	−0.91	4.22

*Raw annual change, raw annual change score, calculated as [(Value at last measurement occasion in the study-Value at baseline)/total years in study]; APC, annual percentage change, calculated as [(Value at last measurement occasion in the study/Value at baseline)$^{1/(\text{total years in study})}$-1]*100.*

TABLE 3 | Parameter estimates and model fit statistics of best fitting univariate LGC models for WM FA.

		Mean (μ)			Variance (σ^2)				Model fit				
LGC model/WM tract		**Estimate**	***SE***	***p* value**	**Estimate**	***SE***	***p* value**	**ECR**	**χ^2(*df*)**	**χ^2/*df***	**RMSEA [95% *CI*]**	**CFI**	**BIC**
FMIN	Intercept	37.36	0.24	**<0.001**	4.30	0.44	**<0.001**		42.695(24)	1.78	0.058 [0.028–0.086]	0.983	3257.441
	Slope	−0.15	0.06	**0.014**	0.04	0.01	**0.001**	0.44					
SLF	Intercept	36.19	0.21	**<0.001**	3.27	0.33	**<0.001**		67.697(24)	2.82	0.089 [0.064–0.114]	0.960	3023.458
	Slope	−0.2	0.05	**<0.001**	0.02	0.01	**0.024**	0.30					
CST	Intercept	54.52	0.27	**<0.001**	3.90	0.44	**<0.001**		31.733(24)	1.32	0.037 [0.000–0.069]	0.988	3634.188
	Slope	−0.04	0.08	0.628	0.00[a]	0.00	**<0.001**	– [b]					

ECR, Effective Curve Reliability; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; BIC, Bayesian Information Criterium. Parameter estimates are unstandardized, and adjusted for effects of age at baseline, and gender (on intercept and slope), and for head motion at each measurement occasion. FA values are raw scores, multiplied by 100. Results that were significant after FDR correction for multiple comparisons ($p < 0.05$) are highlighted in bold font.

[a] *The slope variance was constrained to be positive, which resulted in a slope variance of zero.*

[b] *ECR could not be calculated because the slope variance was zero.*

TABLE 4 | Effects of Age at baseline and Gender on intercept and slope of best fitting univariate LGC models for WM FA.

		Age_{base}			Gender		
Tract		**Estimate**	***SE***	***p*-value**	**Estimate**	***SE***	***p*-value**
FMIN	Intercept	−0.18	0.03	**<0.001**	−0.64	0.29	**0.026**
	Slope	−0.01	0.01	**0.029**	0.05	0.05	0.316
SLF	Intercept	−0.13	0.03	**<0.001**	0.36	0.25	0.148
	Slope	−0.01	0.00	0.122	0.01	0.04	0.692
CST	Intercept	0.00	0.03	0.892	0.51	0.29	0.075
	Slope	−0.02	0.01	**0.017**	0.22	0.06	**<0.001**

Age_{base}, age at baseline. Parameter estimates are unstandardized. Results that were significant after FDR correction for multiple comparisons ($p < 0.05$) are highlighted in bold font.

performance changes in two motor speed tests (pegboard right hand: –0.13 and left hand tapping: –0.07), indicating that older participants' motor function declined more rapidly as compared to their younger peers.

Gender was significantly related with motor function at baseline, such that male participants showed better motor performance in the tapping (left hand: 5.43, $SE = 1.32$, $p < 0.001$; right hand: 6.16, $SE = 1.20$, $p < 0.001$) and grip force tests (left hand: 15.34, $SE = 0.78$, $p < 0.001$; right hand: 15.66, $SE = 0.77$, $p < 0.001$). In addition, male participants showed steeper declines in left hand grip force performance over 4 years than female participants (–0.34, $SE = 0.17$, $p = 0.045$), however, this effect was no longer significant after correction for multiple comparisons.

TABLE 5 | Effects of head motion on FA at each measurement occasion of best fitting univariate LGC models for WM FA.

	Head Motion$_{base}$			Head Motion$_{1year}$			Head Motion$_{2years}$			Head Motion$_{4years}$		
Tract	**Estimate**	***SE***	***p*-value**	**Estimate**	***SE***	***p*-value**	**Estimate**	***SE***	***p*-value**	**Estimate**	***SE***	***p*-value**
FMIN	–1.76	0.57	**0.002**	–0.49	0.46	0.289	–0.70	0.46	0.128	–1.10	0.73	0.134
SLF	–0.90	0.49	0.069	–2.02	0.40	**<0.001**	–1.73	0.39	**<0.001**	–0.52	0.60	0.384
CST	–2.51	0.83	**0.002**	–2.11	0.68	**0.002**	–2.29	0.66	**0.001**	–0.89	0.95	0.346

Base, baseline. Parameter estimates are unstandardized. Results that were significant after FDR correction ($p < 0.05$) are highlighted in bold font.

TABLE 6 | Parameter estimates and model fit statistics of best fitting univariate LGC models for motor function.

		Mean (μ)			Variance (σ^2)				Model fit				
LGC model/WM tract		**Estimate**	***SE***	***p*-value**	**Estimate**	***SE***	***p*-value**	**ECR**	χ^2***(df)***	χ^2***/df***	**RMSEA [95% *CI*]**	**CFI**	**BIC**
Motor Speed													
Pegboard (*l*)	Intercept	51.06	0.84	**<0.001**	60.50	7.48	**<0.001**		15.972(12)	1.33	0.038 [0.000–0.082]	0.993	6907.057
	Slope	–0.31	0.21	0.138	0.27	0.44	0.545	0.09					
Pegboard (*r*)	Intercept	51.83	0.84	**<0.001**	59.07	7.44	**<0.001**		19.685(12)	1.64	0.053 [0.000–0.093]	0.987	6972.606
	Slope	–0.56	0.21	**0.007**	0.25	0.46	0.584	0.09					
Tapping (*l*)[a]	Intercept	47.60	0.95	**<0.001**	85.14	8.81	**<0.001**		27.957(13)	2.15	0.071 [0.034–0.107]	0.980	6875.361
	Slope	–0.53	0.18	**0.003**	0.06	0.29	0.842	0.03					
Tapping (*r*)	Intercept	47.15	0.86	**<0.001**	68.98	7.68	**<0.001**		23.735(12)	1.98	0.065 [0.024–0.103]	0.984	6813.738
	Slope	–0.44	0.18	**0.014**	0.35	0.33	0.289	0.15					
Motor Strength													
Grip Force (l)	Intercept	42.31	0.56	**<0.001**	30.21	3.27	**<0.001**		31.070(12)	2.59	0.083 [0.048–0.119]	0.986	6164.275
	Slope	–0.45	0.13	**<0.001**	0.50	0.16	**0.002**	0.47					
Grip Force (r)	Intercept	42.20	0.56	**<0.001**	28.82	3.19	**<0.001**		21.785(12)	1.82	0.059 [0.013–0.099]	0.993	6260.735
	Slope	–0.39	0.14	**0.005**	0.80	0.20	**<0.001**	0.60					

ECR, Effective Curve Reliability; RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; BIC, Bayesian Information Criterium; l, left; r, right. Parameter estimates are unstandardized, and adjusted for effects of age at baseline, and gender (on intercept and slope). Motor function values are T-distributed for all tests (pegboard, tapping and grip force), with higher values reflecting better performance. Results that were significant after FDR correction for multiple comparisons ($p < 0.05$) are highlighted in bold font.

[a] *The slope variance of this model was constrained to be positive and the intercept-slope covariance was fixed to zero.*

TABLE 7 | Effects of Age at baseline and Gender on intercept and slope of best fitting univariate LGC models for motor function.

		Age$_{base}$			Gender		
Tract		**Estimate**	***SE***	***p*-value**	**Estimate**	***SE***	***p*-value**
Motor Speed							
Pegboard (*l*)	Intercept	–0.86	0.12	**<0.001**	–0.57	1.17	0.629
	Slope	–0.03	0.03	0.374	–0.27	0.28	0.345
Pegboard (*r*)	Intercept	–0.86	0.12	**<0.001**	–2.26	1.17	0.053
	Slope	–0.13	0.03	**<0.001**	–0.25	0.28	0.370
Tapping (*l*)	Intercept	–0.43	0.13	**0.001**	5.43	1.32	**<0.001**
	Slope	–0.07	0.03	**0.006**	0.21	0.24	0.387
Tapping (*r*)	Intercept	–0.60	0.12	**<0.001**	6.16	1.20	**<0.001**
	Slope	–0.01	0.03	0.740	0.17	0.25	0.490
Motor Strength							
Grip Force (*l*)	Intercept	–0.47	0.08	**<0.001**	15.34	0.78	**<0.001**
	Slope	–0.01	0.02	0.632	–0.34	0.17	0.045[a]
Grip Force (*r*)	Intercept	–0.45	0.08	**<0.001**	15.66	0.77	**<0.001**
	Slope	–0.01	0.02	0.549	–0.22	0.19	0.246

Age$_{base}$, age at baseline; l, left; r, right. Parameter estimates are unstandardized. Results that were significant after FDR correction for multiple comparisons ($p < 0.05$) are highlighted in bold font.

[a] *No longer significant ($p < 0.05$) after correction for multiple comparisons.*

TABLE 8 | Model fit statistics of bivariate LGC models (for parameter estimates see **Table 9**).

WM tract	Motor function	χ^2 (*df*)	χ^2/df	RMSEA [95% *CI*]	CFI	BIC
FMIN	Pegboard (*l*)	94.089 (67)	1.40	0.042 [0.019–0.061]	0.984	8421.788
	Pegboard (*r*)	79.877 (67)	1.19	0.029 [0.000–0.050]	0.992	8488.734
	Tapping (*l*)	110.001 (68)	1.62	0.052 [0.033–0.069]	0.977	8387.129
	Tapping (*r*)	92.599 (67)	1.38	0.041 [0.017–0.060]	0.986	8329.705
	Grip Force (*l*)	102.529 (65)	1.58	0.050 [0.030–0.068]	0.985	7684.128
	Grip Force (*r*)	96.723 (65)	1.49	0.046 [0.025–0.064]	0.987	7779.699
SLF	Pegboard (*l*)	113.834 (67)	1.70	0.055 [0.037–0.072]	0.972	8188.823
	Pegboard (*r*)	107.235 (67)	1.60	0.051 [0.032–0.068]	0.976	8255.944
	Tapping (*l*)	122.849 (68)	1.81	0.059 [0.042–0.076]	0.970	8156.749
	Tapping (*r*)	109.217 (67)	1.63	0.052 [0.034–0.070]	0.977	8098.732
	Grip Force (*l*)	123.705 (65)	1.90	0.063 [0.046–0.079]	0.976	7453.513
	Grip Force (*r*)	131.512 (65)	2.02	0.067 [0.050–0.083]	0.973	7549.944
CST	Pegboard (*l*)	79.414 (67)	1.19	0.028 [0.000–0.050]	0.990	8802.741
	Pegboard (*r*)	71.231 (67)	1.06	0.017 [0.000–0.043]	0.996	8868.370
	Tapping (*l*)	86.920 (68)	1.28	0.035 [0.000–0.055]	0.986	8770.575
	Tapping (*r*)	79.729 (67)	1.19	0.029 [0.000–0.050]	0.991	8709.157
	Grip Force (*l*)	96.105 (66)	1.46	0.044 [0.023–0.063]	0.985	8061.218
	Grip Force (*r*)	99.550 (66)	1.51	0.047 [0.026–0.065]	0.983	8154.132

RMSEA, Root Mean Square Error of Approximation; CFI, Comparative Fit Index; BIC, Bayesian Information Criterium; l, left; r, right.

Bivariate LGC Models: FA and Motor Function

We fit bivariate LGC models to estimate cross-domain relationships between mean WM FA in each of the three WM tracts and motor function in each of the six motor tests, resulting in overall 18 separate models. Fit statistics of these models are presented in **Table 8**, and standardized parameter estimates are shown in **Table 9**.

Model fit statistics ranged from adequate to good for all models [$\chi^2_{(65-68)}$ = 71.231 - 131.512, χ^2/df = 1.06–2.02, RMSEA = 0.017–0.067, CFI = 0.970–0.996]. First, in each of the 18 bivariate LGC models, we estimated level–level (i.e., intercept–intercept) correlations between baseline mean FA values and baseline motor function. The results revealed a significant positive correlation for the FMIN and motor speed performance: higher baseline mean FA in the FMIN was associated with faster motor performance in the pegboard (fully standardized estimate: 0.16, *SE* = 0.08, p = 0.035) and tapping tests (fully standardized estimate: 0.19, *SE* = 0.07, p = 0.006) for the left, but not for the right hand. These estimates correspond to a small to typical effect size if compared to the norms in interindividual difference research (Gignac and Szodorai, 2016). We also observed a positive correlation for the FMIN and right-hand grip force performance, however, this result was no longer significant after correction for multiple comparisons. None of the other level–level correlations were significant.

In addition, we estimated level-change (i.e., intercept-slope) correlations to investigate the association between baseline mean FA values and motor performance changes. As the motor speed tests did not show sufficient slope variance, we only estimated level-change correlations for the motor strength tests (grip force left and right hand). None of these results were significant.

Finally, to investigate the association between changes in mean WM FA and changes in motor function, we estimated change–change (i.e., slope–slope) correlations between motor strength in the grip force test and mean FA in those WM tracts that showed sufficient slope variance (FMIN, SLF). Again, none of these correlations were significant.

DISCUSSION

In the current study we followed > 200 healthy older adult participants over 4 years of longitudinal DW-MRI and manual motor performance assessments. To optimally harvest insights from longitudinal data, we conducted our analyses using LGC models in the SEM framework, which enable the separation of interindividual variability from intraindividual change trajectories. We specifically targeted the FMIN, SLF, and CST for their purported roles in motor and cognitive function. Cross sectional associations with age at baseline indicate that mean FA values in these tracts as well as motor performance indicators are lower in older individuals, similar to what has been shown previously (Ruff and Parker, 1993; Jiménez-Jiménez et al., 2011; Bennett and Madden, 2014; Cox et al., 2016). We observed longitudinal declines over time in all measures except for average FA in the CST. Baseline age was also related to changes in mean FA over time in the FMIN and CST, suggesting that changes accelerated with advancing age in the FMIN. In case of the CST, this finding suggests that longitudinal decline is restricted to the oldest participants, as no decline was

TABLE 9 | Results of bivariate LGC models (level-level, level-change, change-change).

Correlation	WM tract	Motor speed				Motor strength	
		Pegboard (*l*)	Pegboard (*r*)	Tapping (*l*)	Tapping (*r*)	Grip Force (*l*)	Grip Force (*r*)
Level-level	FMIN	**0.16 (0.08) ***	0.13 (0.07)	**0.19 (0.07)****	0.13 (0.07)	0.13 (0.07)	0.16 (0.07)*[a]
	SLF	0.14 (0.08)	0.10 (0.08)	0.14 (0.07)	0.02 (0.08)	0.01 (0.07)	0.04 (0.08)
	CST	–0.03 (0.08)	–0.01 (0.07)	–0.05 (0.07)	–0.04 (0.07)	–0.07 (0.07)	–0.14 (0.07)
Level-change	FMIN	–	–	–	–	0.00 (0.13)	–0.12 (0.11)
	SLF	–	–	–	–	0.07 (0.13)	–0.05 (0.12)
	CST	–	–	–	–	–0.00 (0.12)	–0.03 (0.11)
Change-change	FMIN	–	–	–	–	–0.13 (0.19)	–0.06 (0.16)
	SLF	–	–	–	–	0.01 (0.23)	0.04 (0.20)
	CST	–	–	–	–	–	–

Parameter estimates are standardized, with standard errors in parentheses. Results that were significant ($p < 0.05$, **$p < 0.01$) are marked with stars. Results that were still significant after FDR correction for multiple comparisons are highlighted in bold font.*
FMIN, Forceps Minor; SLF, Superior Longitudinal Fasciculus; CST, Corticospinal Tract; l, left; r, right.
[a]*No longer significant ($p < 0.05$) after correction for multiple comparisons.*

observed in average FA for the overall sample in this tract. We found that FMIN tract average FA values correlate positively with motor speed and motor strength measures, suggesting that maintenance of WM structure in the anterior corpus callosum is associated with better manual motor function. We did not observe significant longitudinal level-change or change-change associations, suggesting that – at least over the timescale studied here – age-related baseline levels and changes in WM microstructure within these tracts do not associate with changes in motor speed and grip force.

Longitudinal Change in and Cross-Sectional Age Effects on Motor Speed and Motor Strength

Our LGC models for motor function reveal both longitudinal declines over time and cross-sectional effects of age, consistent with previous literature (Kallman et al., 1990; Ruff and Parker, 1993; Vianna et al., 2007). In addition, our results point to an acceleration of motor performance declines in older individuals. In light of the recent proposition that changes in grip strength are not simply an index of muscle mass loss but should be viewed as a marker of brain health given the complex neural circuitry that it engages (Carson, 2018), the loss of left-hand grip strength with an annual decline rate of 1.2% (**Table 2**) is particularly interesting. While some studies have reported greater grip strength loss with age in the right hand (Teixeira, 2008), others have shown more selective loss of neural control of the left (non-dominant) hand (Sale and Semmler, 2005). Here, we observed greater declines for the left hand, supporting our hypothesis of right-hand performance preservation due to stronger lifelong practice of the dominant hand. With respect to gender, we expectedly found significant cross-sectional effects for baseline grip force and tapping performance, with males having higher grip strength and faster tapping speeds. This is in accordance with previous work (Hubel et al., 2013; Yorke et al., 2015; Wong, 2016).

Longitudinal Change in and Cross-Sectional Age Effects on Fractional Anisotropy

We found that both FMIN and SLF WM indices significantly declined over time, whereas CST did not. However, age was significantly negatively associated with CST change, implying that CST changes over time only occurred for the oldest participants in our sample. It should be noted that, in the case of the CST, the results from the univariate LGC model diverged from the annual change computed based on the raw FA values. While the modeled change values suggested stability in mean FA over 4 years, annual mean FA increases were observed in the raw values (0.25% ± 0.92%). This divergence most likely is a result of the inclusion of head motion and baseline age as covariates in the LGC models, but not in the calculation of the raw change values. While head motion had a significant impact on all WM tracts, the CST was most affected. Especially, head motion had an impact on the CST at the first three measurement occasions (i.e., by underestimating mean FA values), but not at the last occasion, which was also spaced further apart and thus had more weight for the change calculation. If not adjusted for, this imbalance in the FA estimation might result in the finding of FA increases over time instead of stability. The strict control for head motion, both by excluding observations with extreme motion, and including head motion as a covariate into the estimation of FA, is a major strength of this study. It has been previously reported that participant's head movements in the scanner can result in biased FA estimates (Yendiki et al., 2014; Baum et al., 2018). Despite this common knowledge, head motion is often neglected in DW-MRI studies.

We have recently shown that a global WM decline factor is not a good fit to our data (Oschwald et al., 2019b). Lövdén et al. (2013) have also reported that aging differentially impacts individual WM tracts. Further, studies investigating gray and WM volume and WM diffusion changes with age largely provide evidence for the "last in, first out" hypothesis. This conceptual model proposes that tracts which are later to mature developmentally will be more

sensitive to aging effects (Raz, 2000, 2001; Bender et al., 2016). Some studies have shown that primary motor and somatosensory cortex volumes are relatively spared by aging (Raz et al., 1997). In contrast, other studies have suggested disproportionate age effects on pre- and postcentral gyrus volume (Good et al., 2001; Taubert et al., 2020), cortical thickness (Salat et al., 2004), iron and myelin content (Taubert et al., 2020). It should be noted, however, that these studies were all cross-sectional and did not quantify longitudinal changes. In the current longitudinal investigation, we see more robust declines in WM microstructure for the FMIN and SLF than for the CST, which is consistent with the idea of a higher susceptibility of association and commissural fibers for detrimental effects of aging (Madden et al., 2012; Bender et al., 2016).

Cross-Sectional and Longitudinal Brain-Behavior Associations

With regard to cross-sectional brain-behavior associations at baseline, left hand pegboard and tapping speed measures were significantly correlated with FMIN FA, with higher average FA values being associated with faster performance. There was also a trend for a level-level association between mean FA in the FMIN and right-hand grip force, but this association did not survive correction for multiple comparisons. We have previously shown that corpus callosum WM integrity is correlated with unimanual and bimanual task performance in older adults (Fling et al., 2011; Fling and Seidler, 2012). Interestingly, the pegboard and tapping performance measures were not significantly correlated with mean FA in the CST or SLF at study baseline. CST microstructure has been linked to better motor performance in young adults; for example, CST FA increases with motor practice (Reid et al., 2017) and is higher in the hand and arm motor tracts of musicians (Giacosa et al., 2019). It is well established, however, that motor control relies more upon frontal brain activity in older adults than young adults (Seidler et al., 2010; Hawkins et al., 2018). This additional frontal activity during motor task performance in older adults is often interpreted as compensatory (Heuninckx et al., 2008); that is, it is positively associated with better task performance. Davis et al. (2012) have reported that anterior corpus callosum microstructure predicted the strength of functional connectivity between the left and right prefrontal cortex in older adults performing a letter matching task as well as task performance that relied on interhemispheric communication. It is not clear whether the same type of compensation process is taking place in our study, but it is compelling that only FMIN FA was correlated with task performance.

We hypothesized that SLF FA would also correlate with motor performance in our sample, given its potential link to frontal compensation. The SLF connects the frontal, occipital, parietal and temporal lobes (Wang et al., 2016) and SLF FA has been recently shown to predict increasing frailty over a 5 years follow-up (Maltais et al., 2020). It has also been shown to be correlated with tapping speed in individuals with frontotemporal dementia (Henley et al., 2014) and in those who have suffered traumatic brain injury (Farbota et al., 2012). However, the degree of interindividual variance in SLF microstructure in these populations is likely greater than that observed with healthy aging, potentially resulting in the lack of association seen here.

We did not observe any change-change associations in our data; that is, change in WM microstructure was not associated with change in motor performance. Moreover, there were no level-change associations, meaning that average FA at baseline did not predict future motor declines. This is perhaps to be expected given the lack of change over time in CST mean FA and the small slope variance in our measures. Of note, due to the lack of variance in the change trajectories of the motor speed tasks, we could only evaluate longitudinal level-change and change–change associations of WM FA with motor strength in the grip force task. Very recently, grip force has been associated with cognitive decline in healthy aging and demented patients (Cui et al., 2021) and is therefore being discussed also as an early marker of cognitive degradation. While we are still in the beginning of understanding how changes in brain structure impact on changes in motor function, previous longitudinal studies on brain structure and cognitive performance have reported inconsistent results with respect to change-change associations, especially when examining healthy older adults (Salthouse, 2011; Oschwald et al., 2019a).

This is perhaps a reflection of several factors: the good health status and thus high capacity for adaptive compensation for neural decline in this population (Reuter-Lorenz and Park, 2014), the complexity of the cognitive aging process itself, involving a multitude of driving factors (Grady, 2012), and the challenge of modeling the intricate temporal dynamics between two developmental aging processes (King et al., 2018). While the sample that is studied here comprises highly functioning, healthy individuals with expectedly high compensatory ability, they still showed average declines in grip force over the time period studied. These declines, however, are small if compared to previous reports (Auyeung et al., 2014; Patel et al., 2018). Moreover, age-related changes in grip force may be especially multidetermined given its above-mentioned association with cognitive decline (Cui et al., 2021), and its relation to several neural substrates (Carson, 2018). Together, this may serve as an explanation why changes in grip force and FA are not significantly related to each other in our analyses. In addition, while the longitudinal nature of this study is one of its major strength and sets it apart from most of the literature investigating the relations between brain and motor aging, modeling the temporal dynamics between these two developmentally distinct domains still presents a major challenge. In the present study, we used growth modeling to capture simultaneous change-change associations between WM microstructure and motor function. However, more fine-grained temporal investigations of leading-lagging relationships may have been more sensitive to uncover a relationship between these domains (see McArdle et al., 2004; Grimm et al., 2012; Estrada et al., 2019 for such applications in the field of cognitive neuroscience).

In a recent study with the same sample, we reported that changes in SLF FA predict changes in processing speed 2 years later (Oschwald et al., 2019b). It is possible that

effects of WM microstructure degradation do not exert immediate effects on motor function, but manifest only after a certain time lag. Unfortunately, the fact that most of the motor function measures in the present study showed only very limited between-person slope variance prohibited reliable modeling of more complex lagged dynamic change processes. However, even in our previous study, only a small subset of the WM tracts investigated (i.e., SLF and anterior thalamic radiation) showed lagged change associations over the studied 4-year interval, which promotes the hypothesis that in the context of healthy aging, with compensation mechanisms being in place, longer time delays might be needed to reveal consistent change-change associations between brain structural and behavior changes.

The high level of health in our participants can also be considered a strength, since many other studies investigate mild cognitive impairment, Alzheimer's disease patients, or other pathological samples. Understanding the trajectories of neural decline and the related motor impairment in healthy old age is of high practical relevance in our aging society.

To quantify slope variance reliability, we calculated ECR, an index reflecting the slope variance scaled on the effective error and interpretable as a standardized effect size. Effective error variance reflects the size of unsystematic variance in the measure of interest over several measurement occasions. It is influenced by the number of measurement occasions, temporal spacing between these occasions, the overall duration of the study, and the reliability of the measurement instrument itself (Brandmaier et al., 2018). Thus, small slope variance reliability as it was observed for several motor function measures in this study, can reflect a lack of true variance in change and/or low measurement precision that is influenced by a number of study design features. To gain more insights into the longitudinal associations between brain structural and behavior changes, we will have to await future studies that can exploit datasets with more measurement occasions spanning longer time periods. It will be of interest to compute similar measures of slope variance reliability as in this study, to be able to compare these effects across different study designs.

We decided to follow a strictly hypothesis-driven approach in the present study, evaluating only selected WM fiber tracts that have been reportedly involved in motor and cognitive functions relevant for the motor performance tests studied here. It is possible that an analysis approach that evaluates the detailed properties of the tracts (i.e., by running voxel-wise analyses) would have revealed effects that are masked via the averaging of FA values over the entire tract. However, this would have greatly increased the amount of multiple comparisons that would require correcting for, thus substantially lowering power to detect any effects of interest. Furthermore, ROI-based analyses allow the application of sophisticated statistical growth models to our longitudinal data to analyze parallel change processes (which is not feasible in the context of voxel-wise approaches, such as TBSS, which allow only the longitudinal modeling of one change process).

CONCLUSION AND FUTURE DIRECTIONS

The current study features the inclusion of multiple longitudinal assessments of both WM microstructure and motor function in a large sample of healthy older adults. Longitudinal assessments across multiple measurement occasions are crucial to better understand trajectories of neural and motor function change in old age and to unravel the effects of central nervous system decline on motor behavior decline over time (Oschwald et al., 2019a). We report evidence for both motor performance and mean WM FA declines over 4 years and negative associations with age. Interestingly, we observed declines in mean FA of the CST only in the older participants but not in the whole sample, which provides support for the "last in, first out" hypothesis of aging which postulates less decline for evolutionarily and developmentally older brain regions and pathways. Mean FA in the FMIN, but not the SLF or CST, correlated with motor speed at baseline. We did not find any longitudinal associations between neural and motor functioning, however. Overall, our results (a) provide important insights into aging-related changes of fine motor abilities and FA in selected WM tracts associated with motor control, (b) support previous cross-sectional work showing that neural control of movement in older adults also involves brain structures outside the core motor system and (c) align with the idea that, in healthy aging, compensatory mechanisms may be in place and longer time delays may be needed to reveal level change or change-change associations. More longitudinal assessments with multiple follow-ups are required to precisely delineate the complex dynamic change associations between neural and motor functioning in aging research.

DATA AVAILABILITY STATEMENT

The dataset presented in this article is not publicly available because the used consent does not allow for the public sharing of the data.

ETHICS STATEMENT

This study involving human participants was reviewed and approved by the Ethics Committee of the Canton of Zurich. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SM and LJ contributed to the design, set-up, maintenance, and support of the Longitudinal Healthy Aging Brain (LHAB) database. JO performed the statistical analysis. JO, RS, and SM wrote the first draft of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

This research was funded by the Velux Stiftung (project no. 369) and the University Research Priority Program "Dynamics of Healthy Aging" of the UZH.

ACKNOWLEDGMENTS

The current analysis incorporates data from the Longitudinal Healthy Aging Brain (LHAB) database project carried out at the University of Zurich (UZH). The following researchers at the UZH were involved in the design, set-up, maintenance and support of the LHAB database: Anne Eschen, LJ, Franz Liem, Mike Martin, SM, Christina Röcke, and Jacqueline Zöllig. During the work on this manuscript, JO was a pre-doctoral fellow of the International Max Planck Research School on the Life Course (LIFE, www.imprs-life.mpg.de; participating institutions: Max Planck Institute for Human Development, Freie Universität Berlin, Humboldt-Universität zu Berlin, University of Michigan, University of Virginia, University of Zurich).

REFERENCES

Andersson, J. L. R., Graham, M. S., Zsoldos, E., and Sotiropoulos, S. N. (2016). Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage* 141, 556–572. doi: 10.1016/j.neuroimage.2016.06.058

Andersson, J. L. R., and Sotiropoulos, S. N. (2016). An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* 125, 1063–1078. doi: 10.1016/j.neuroimage.2015.10.019

Auyeung, T. W., Lee, S. W. J., Leung, J., Kwok, T., and Woo, J. (2014). Age-associated decline of muscle mass, grip strength and gait speed: a 4-year longitudinal study of 3018 community-dwelling older Chinese. *Geriatr. Gerontol. Int.* 14(Suppl. 1), 76–84. doi: 10.1111/ggi.12213

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

Barrick, T. R., Charlton, R. A., Clark, C. A., and Markus, H. S. (2010). White matter structural decline in normal ageing: a prospective longitudinal study using tract-based spatial statistics. *Neuroimage* 51, 565–577. doi: 10.1016/j.neuroimage.2010.02.033

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267. doi: 10.1016/S0006-3495(94)80775-1

Baum, G. L., Roalf, D. R., Cook, P. A., Ciric, R., Rosen, A. F. G., Xia, C., et al. (2018). The impact of in-scanner head motion on structural connectivity derived from diffusion MRI. *Neuroimage* 173, 275–286. doi: 10.1016/j.neuroimage.2018.02.041

Bender, A. R., Völkle, M. C., and Raz, N. (2016). Differential aging of cerebral white matter in middle-aged and older adults: a seven-year follow-up. *Neuroimage* 125, 74–83. doi: 10.1016/j.neuroimage.2015.10.030

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Bennett, I. J., and Madden, D. J. (2014). Disconnected aging: cerebral white matter integrity and age-related differences in cognition. *Neuroscience* 276, 187–205. doi: 10.1016/j.neuroscience.2013.11.026

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol. Bull.* 107, 238–246. doi: 10.1037/0033-2909.107.2.238

Bohannon, R. W. (2008). Hand-grip dynamometry predicts future outcomes in aging adults. *J. Geriatr. Phys. Ther.* 31, 3–10. doi: 10.1519/00139143-200831010-00002

Brandmaier, A. M., von Oertzen, T., Ghisletta, P., Lindenberger, U., and Hertzog, C. (2018). Precision, reliability, and effect size of slope variance in latent growth curve models: implications for statistical power analysis. *Front. Psychol.* 9:294. doi: 10.3389/fpsyg.2018.00294

Carson, R. G. (2018). Get a grip: individual variations in grip strength are a marker of brain health. *Neurobiol. Aging* 71, 189–222. doi: 10.1016/j.neurobiolaging.2018.07.023

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillside, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi: 10.1037/0033-2909.112.1.155

Cox, S. R., Ritchie, S. J., Tucker-Drob, E. M., Liewald, D. C., Hagenaars, S. P., Davies, G., et al. (2016). Ageing and brain white matter structure in 3,513 UK biobank participants. *Nat. Commun.* 7:13629. doi: 10.1038/ncomms13629

Cui, M., Zhang, S., Liu, Y., Gang, X., and Wang, G. (2021). Grip strength and the risk of cognitive decline and dementia: a systematic review and meta-analysis of longitudinal cohort studies. *Front. Aging Neurosci.* 13:625551. doi: 10.3389/fnagi.2021.625551

Davis, S. W., Kragel, J. E., Madden, D. J., and Cabeza, R. (2012). The architecture of cross-hemispheric communication in the aging brain: linking behavior to functional and structural connectivity. *Cereb. Cortex* 22, 232–242. doi: 10.1093/cercor/bhr123

de Groot, M., Cremers, L. G. M., Ikram, M. A., Hofman, A., Krestin, G. P., van der Lugt, A., et al. (2016). White matter degeneration with aging: longitudinal diffusion MR imaging analysis. *Radiology* 279, 532–541. doi: 10.1148/radiol.2015150103

Dhollander, T., Raffelt, D., and Connelly, A. (2016). "Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion MR data without a co-registered T1 image," in *Proceedings of the ISMRM Workshop on Breaking the Barriers of Diffusion MRI*, (Lisbon).

Estrada, E., Ferrer, E., Román, F. J., Karama, S., and Colom, R. (2019). Time-lagged associations between cognitive and cortical development from childhood to early adulthood. *Dev. Psychol.* 55, 1338–1352. doi: 10.1037/dev0000716

Farbota, K. D., Bendlin, B. B., Alexander, A. L., Rowley, H. A., Dempsey, R. J., and Johnson, S. C. (2012). Longitudinal diffusion tensor imaging and neuropsychological correlates in traumatic brain injury patients. *Front. Hum. Neurosci.* 6:160. doi: 10.3389/fnhum.2012.00160

Finkbeiner, C. (1979). Estimation for the multiple factor model when data are missing. *Psychometrika* 44, 409–420. doi: 10.1007/BF02296204

Fling, B. W., and Seidler, R. D. (2012). Fundamental differences in callosal structure, neurophysiologic function, and bimanual control in young and older adults. *Cereb. Cortex* 22, 2643–2652. doi: 10.1093/cercor/bhr349

Fling, B. W., Walsh, C. M., Bangert, A. S., Reuter-Lorenz, P. A., Welsh, R. C., and Seidler, R. D. (2011). Differential callosal contributions to bimanual control in young and older adults. *J. Cogn. Neurosci.* 23, 2171–2185. doi: 10.1162/jocn.2010.21600

Ghisletta, P., and McArdle, J. J. (2012). Latent curve models and latent change score models estimated in R. *Struct. Equ. Model.* 19, 651–682. doi: 10.1080/10705511.2012.713275

Giacosa, C., Karpati, F. J., Foster, N. E. V., Hyde, K. L., and Penhune, V. B. (2019). The descending motor tracts are different in dancers and musicians. *Brain Struct. Funct.* 224, 3229–3246. doi: 10.1007/s00429-019-01963-0

Gignac, G. E., and Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Pers. Individ. Diff.* 102, 74–78. doi: 10.1016/j.paid.2016.06.069

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

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., et al. (2017). BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol.* 13:e1005209. doi: 10.1371/journal.pcbi.1005209

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44

Grady, C. L. (2012). The cognitive neuroscience of ageing. *Nat. Rev. Neurosci.* 13, 491–505. doi: 10.1038/nrn3256

Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., and Resnick, S. M. (2012). Recent changes leading to subsequent changes: extensions of multivariate latent difference score models. *Struct. Equ. Model. Multidiscip. J.* 19, 268–292. doi: 10.1080/10705511.2012.659627

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Hawkins, K. A., Fox, E. J., Daly, J. J., Rose, D. K., Christou, E. A., McGuirk, T. E., et al. (2018). Prefrontal over-activation during walking in people with mobility deficits: interpretation and functional implications. *Hum. Mov. Sci.* 59, 46–55. doi: 10.1016/j.humov.2018.03.010

Henley, S. M. D., Downey, L. E., Nicholas, J. M., Kinnunen, K. M., Golden, H. L., Buckley, A., et al. (2014). Degradation of cognitive timing mechanisms in behavioural variant frontotemporal dementia. *Neuropsychologia* 65, 88–101. doi: 10.1016/j.neuropsychologia.2014.10.009

Heuninckx, S., Wenderoth, N., and Swinnen, S. P. (2008). Systems neuroplasticity in the aging brain: recruiting additional neural resources for successful motor performance in elderly persons. *J. Neurosci.* 28, 91–99. doi: 10.1523/JNEUROSCI.3300-07.2008

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., et al. (2008). Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage* 39, 336–347. doi: 10.1016/j.neuroimage.2007.07.053

Hubel, K. A., Reed, B., Yund, E. W., Herron, T. J., and Woods, D. L. (2013). Computerized measures of finger tapping: effects of hand dominance, age, and sex. *Percept. Mot. Skills* 116, 929–952. doi: 10.2466/25.29.PMS.116.3.929-952

Isa, T. (2012). The corticospinal tract and its role in motor control. *Brain Nerve* 64, 1331–1339.

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

Jiménez-Jiménez, F. J., Calleja, M., Alonso-Navarro, H., Rubio, L., Navacerrada, F., Pilo-De-La-Fuente, B., et al. (2011). Influence of age and gender in motor performance in healthy subjects. *J. Neurol. Sci.* 302, 72–80. doi: 10.1016/j.jns.2010.11.021

Jones, D. K., Knösche, T. R., and Turner, R. (2013). White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *Neuroimage* 73, 239–254. doi: 10.1016/j.neuroimage.2012.06.081

Jöreskog, K. G., and Sörbom, D. (1993). *Structural Equation Modeling with the SIMPLIS Command Language*. Chicago: Scientific Software.

Kallman, D. A., Plato, C. C., and Tobin, J. D. (1990). The role of muscle loss in the age-related decline of grip strength: cross-sectional and longitudinal perspectives. *J. Gerontol. Med. Sci.* 45, 82–88. doi: 10.1093/geronj/45.3.M82

Kass, R. E., and Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* 90, 773–795. doi: 10.2307/2291091

Kelley, K., and Preacher, K. J. (2012). On effect size. *Psychol. Methods* 17, 137–152. doi: 10.1037/a0028086

King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., and Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: common challenges, and solutions from developmental psychology. *Dev. Cogn. Neurosci.* 33, 54–72. doi: 10.1016/j.dcn.2017.11.009

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* 49, 764–766. doi: 10.1016/j.jesp.2013.03.013

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *J. Am. Stat. Assoc.* 90, 1112–1121. doi: 10.2307/2291350

Lövdén, M., Laukka, E. J., Rieckmann, A., Kalpouzos, G., Li, T. Q., Jonsson, T., et al. (2013). The dimensionality of between-person differences in white matter microstructure in old age. *Hum. Brain Mapp.* 34, 1386–1398. doi: 10.1002/hbm.21518

Madden, D. J., Bennett, I. J., Burzynska, A., Potter, G. G., Chen, N., and Song, A. (2012). Diffusion tensor imaging of cerebral white matter integrity in cognitive aging. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1822, 386–400. doi: 10.1016/j.bbadis.2011.08.003

Malagurski, B., Liem, F., Oschwald, J., Mérillat, S., and Jäncke, L. (2020). Functional dedifferentiation of associative resting state networks in older adults – a longitudinal study. *Neuroimage* 214:116680. doi: 10.1016/j.neuroimage.2020.116680

Maltais, M., de Souto Barreto, P., Perus, L., Mangin, J.-F., Grigis, A., Chupin, M., et al. (2020). Prospective associations between diffusion tensor imaging parameters and frailty in older adults. *J. Am. Geriatr. Soc.* 68, 1050–1055. doi: 10.1111/jgs.16343

Mamiya, P. C., Richards, T. L., and Kuhl, P. K. (2018). Right forceps minor and anterior thalamic radiation predict executive function skills in young bilingual adults. *Front. Psychol.* 9:118. doi: 10.3389/fpsyg.2018.00118

Massa, R. E., Rosso, A., Metti, A. L., Sparto, P. L., Aizenstein, H., Ferrucci, L., et al. (2019). Neuroimaging correlates of lateral postural control in older ambulatory adults. *Aging Clin. Exp. Res.* 31, 611–619. doi: 10.1007/s40520-018-1028-4

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annu. Rev. Psychol.* 60, 577–605. doi: 10.1146/annurev.psych.60.110707.163612

McArdle, J. J., and Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Dev.* 58, 110–133. doi: 10.2307/1130295

McArdle, J. J., Hamgami, F., Jones, K., Jolesz, F., Kikinis, R., Spiro, A. III, et al. (2004). Structural modeling of dynamic changes in memory and brain structure using longitudinal data from the normative aging study. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 59, 294–304. doi: 10.1093/geronb/59.6.P294

Merker, B., and Podell, K. (2010). "Grooved pegboard test," in *Encyclopedia of Clinical Neuropsychology*, eds J. Kreutzer, J. DeLuca, and B. Caplan (New York, NY: Springer), 1176–1178.

Moscufo, N., Wakefield, D. B., Meier, D. S., Cavallari, M., Guttmann, C. R. G., White, W. B., et al. (2018). Longitudinal microstructural changes of cerebral white matter and their association with mobility performance in older persons. *PLoS One* 13:e0194051. doi: 10.1371/journal.pone.0194051

Nakajima, R., Kinoshita, M., Shinohara, H., and Nakada, M. (2020). The superior longitudinal fascicle: reconsidering the fronto-parietal neural network based on anatomy and function. *Brain Imaging Behav.* 14, 2817–2830. doi: 10.1007/s11682-019-00187-4

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., et al. (2019a). Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev. Neurosci.* 31, 1–57. doi: 10.1515/revneuro-2018-0096

Oschwald, J., Mérillat, S., Liem, F., Röcke, C., Martin, M., and Jäncke, L. (2019b). Lagged coupled changes between white matter microstructure and processing speed in healthy aging: a longitudinal investigation. *Front. Aging Neurosci.* 11:298. doi: 10.3389/fnagi.2019.00298

Patel, A., Edwards, M. H., Jameson, K. A., Ward, K. A., Fuggle, N., Cooper, C., et al. (2018). Longitudinal change in peripheral quantitative computed tomography assessment in older adults: the hertfordshire cohort study. *Calcif. Tissue Int.* 103, 476–482. doi: 10.1007/s00223-018-0442-0

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Raftery, A. E. (1995). Bayesian model selection in social research. *Soc. Methodol.* 25, 111–163. doi: 10.2307/271063

Raz, N. (2000). "Aging of the brain and its impact on cognitive performance: integration of structural and functional findings," in *The Handbook of Aging and Cognition*, eds F. I. M. Craik and T. A. Salthouse (Hillside, NJ: Lawrence Erlbaum Associates Publishers), 1–90.

Raz, N. (2001). *Ageing and the Brain: The Encyclopedia of Life Sciences*. London: Nature Publishing Group.

Raz, N., Gunning, F. M., Head, D., Dupuis, J. H., McQuain, J., Briggs, S. D., et al. (1997). Selective aging of the human cerebral cortex observed in vivo: differential vulnerability of the prefrontal gray matter. *Cereb. Cortex* 7, 268–282. doi: 10.1093/cercor/7.3.268

Reid, L. B., Sale, M. V., Cunnington, R., Mattingley, J. B., and Rose, S. E. (2017). Brain changes following four weeks of unimanual motor training: evidence

from fMRI-guided diffusion MRI tractography. *Hum. Brain Mapp.* 38, 4302–4312. doi: 10.1002/hbm.23514

Reuter-Lorenz, P. A., and Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychol. Rev.* 24, 355–370. doi: 10.1007/s11065-014-9270-9

Rosseel, Y. (2012). lavaan: an R Package for Structural Equation Modeling. *J. Stat. Softw.* 48, 1–36.

Ruff, R. M., and Parker, S. B. (1993). Gender- and age-specific changes in motor speed and eye-hand coordination in adults: normative values for the finger tapping and grooved pegboard tests. *Percept. Mot. Skills* 76, 1219–1230.

Salat, D. H., Buckner, R. L., Snyder, A. Z., Greve, D. N., Desikan, R. S. R., Busa, E., et al. (2004). Thinning of the cerebral cortex in aging. *Cereb. Cortex* 14, 721–730. doi: 10.1093/cercor/bhh032

Sale, M. V., and Semmler, J. G. (2005). Age-related differences in corticospinal control during functional isometric contractions in left and right hands. *J. Appl. Physiol.* 99, 1483–1493. doi: 10.1152/japplphysiol.00371.2005

Salthouse, T. A. (2011). Neuroanatomical substrates of age-related cognitive decline. *Psychol. Bull.* 137, 753–784. doi: 10.1037/a0023262

Schafer, J. L., and Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychol. Methods* 7, 147–177. doi: 10.1037//1082-989X.7.2.147

Schermelleh-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. Online* 8, 23–74. doi: 10.1002/0470010940

Schoppe, K. J. (1974). Das MLS-Gerät. ein neuer testapparat zur messung feinmotorischer leistungen. *Diagnostica* 20, 43–46.

Seidler, R. D., Bernard, J. A., Burutolu, T. B., Fling, B. W., Gordon, M. T., Gwin, J. T., et al. (2010). Motor control and aging: links to age-related brain structural, functional, and biochemical effects. *Neurosci. Biobehav. Rev.* 34, 721–733. doi: 10.1016/j.neubiorev.2009.10.005

Steiger, J. H., and Lind, J. C. (1980). "Statistically based tests for the number of common factors," in *Proceedings of the Annual meeting of the Psychometric Society*, (Iowa City, IA), doi: 10.1207/s15327906mbr2502_4

Studenski, S., Perera, S., Patel, K., Rosano, C., Faulkner, K., Inzitari, M., et al. (2011). Gait speed and survival in older adults. *JAMA - J. Am. Med. Assoc.* 305, 50–58. doi: 10.1001/jama.2010.1923

Sullivan, E. V., Rohlfing, T., and Pfefferbaum, A. (2010). Longitudinal study of callosal microstructure in the normal adult aging brain using quantitative DTI fiber tracking. *Dev. Neuropsychol.* 35, 233–256. doi: 10.1080/87565641003689556

Taubert, M., Roggenhofer, E., Melie-Garcia, L., Muller, S., Lehmann, N., Preisig, M., et al. (2020). Converging patterns of aging-associated brain volume loss and tissue microstructure differences. *Neurobiol. Aging* 88, 108–118. doi: 10.1016/j.neurobiolaging.2020.01.006

Teipel, S. J., Meindl, T., Wagner, M., Stieltjes, B., Reuter, S., Hauenstein, K.-H., et al. (2010). Longitudinal changes in fiber tract integrity in healthy aging and mild cognitive impairment: a DTI follow-up study. *J. Alzheimer's Dis.* 22, 507–522. doi: 10.3233/JAD-2010-100234

Teixeira, L. A. (2008). Categories of manual asymmetry and their variation with advancing age. *Cortex* 44, 707–716. doi: 10.1016/j.cortex.2006.10.002

Tournier, J. D., Calamante, F., and Connelly, A. (2012). MRtrix: diffusion tractography in crossing fiber regions. *Int. J. Imaging Syst. Technol.* 22, 53–66. doi: 10.1002/ima.22005

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 29, 1310–1320. doi: 10.1109/TMI.2010.2046908

Veenith, T. V., Carter, E., Grossac, J., Newcombe, V. F. J., Outtrim, J. G., Lupson, V., et al. (2013). Inter subject variability and reproducibility of diffusion tensor imaging within and between different imaging sessions. *PLoS One* 8:e65941. doi: 10.1371/journal.pone.0065941

Veraart, J., Fieremans, E., and Novikov, D. S. (2016a). Diffusion MRI noise mapping using random matrix theory. *Magn. Reson. Med.* 76, 1582–1593. doi: 10.1002/mrm.26059

Veraart, J., Novikov, D. S., Christiaens, D., Ades-aron, B., Sijbers, J., and Fieremans, E. (2016b). Denoising of diffusion MRI using random matrix theory. *Neuroimage* 142, 394–406. doi: 10.1016/j.neuroimage.2016.08.016

Veraart, J., Sijbers, J., Sunaert, S., Leemans, A., and Jeurissen, B. (2013). Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. *Neuroimage* 81, 335–346. doi: 10.1016/j.neuroimage.2013.05.028

Verlinden, V. J. A., de Groot, M., Cremers, L. G. M., van der Geest, J. N., Hofman, A., Niessen, W. J., et al. (2016). Tract-specific white matter microstructure and gait in humans. *Neurobiol. Aging* 43, 164–173. doi: 10.1016/j.neurobiolaging.2016.04.005

Vianna, L. C., Oliveira, R. B., and Araújo, C. S. (2007). Age-related decline in handgrip strength differs according to gender. *J. Strength Cond. Res.* 21, 1310–1314.

Wang, X., Pathak, S., Stefaneanu, L., Yeh, F.-C., Li, S., and Fernandez-Miranda, J. C. (2016). Subcomponents and connectivity of the superior longitudinal fasciculus in the human brain. *Brain Struct. Funct.* 221, 2075–2092.

Ware, J. E., Kosinski, M., and Keller, S. D. (1996). A 12-item short-form health survey?: construction of scales and preliminary tests of reliability and validity. *Med. Care* 34, 220–233. doi: 10.2307/3766749

Ware, J. E., Kosinski, M., and Keller, S. D. (1995). *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*, 2nd Edn. Boston: Health Institute, New England Medical Center.

Westin, C. F., Peled, S., Gudbjartsson, H., Kikinis, R., and Jolesz, F. A. (1997). "Geometrical diffusion measures for MRI from tensor basis analysis," in *Proceedings of the 5th Annual Meeting of ISMRM*, (Vancouver), 1742.

Wong, S. L. (2016). Reduced muscular strength among Canadians aged 60 to 79: Canadian health measures survey, 2007 to 2013. *Heal. Reports* 27, 3–10.

Yendiki, A., Koldewyn, K., Kakunoori, S., Kanwisher, N., and Fischl, B. (2014). Spurious group differences due to head motion in a diffusion MRI study. *Neuroimage* 88, 79–90. doi: 10.1016/j.neuroimage.2013.11.027

Yorke, A. M., Curtis, A. B., Shoemaker, M., and Vangsnes, E. (2015). Grip strength values stratified by age, gender, and chronic disease status in adults aged 50 years and older. *J. Geriatr. Phys. Ther.* 38, 115–121. doi: 10.1519/JPT.0000000000000037

Zahodne, L. B., and Reuter-Lorenz, P. A. (2019). "Compensation and brain aging: a review and analysis of evidence," in *The Aging Brain: Functional Adaptation Across Adulthood*, ed. G. R. Samanez-Larkin (Washington, DC: American Psychological Association), 185–216. doi: 10.1037/0000143-008

Zöllig, J., Mérillat, S., Eschen, A., Röcke, C., Martin, M., and Jäncke, L. (2011). Plasticity and imaging research in healthy aging: core ideas and profile of the International Normal Aging and Plasticity Imaging Center (INAPIC). *Gerontology* 57, 190–192. doi: 10.1159/000324307

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership