Statistical Analysis Plan (SAP)

"Effects of a Multi-Session Cognitive Training Combined With Brain Stimulation on Age-Associated Cognitive Decline" Acronym: **TrainStim-Cog**
Version 1
Date: 20th of January 2020

Intervention therapy: nine-session cognitive training over three weeks with tCDS over the left dorsolateral prefrontal cortex (DLPFC)

Control therapy: nine-session cognitive training over three weeks with sham stimulation

Study population: healthy older individuals

Clinical Phase: mono-centric randomized placebo-controlled trial

Clinicaltrials.gov Identifier: NCT03838211

Principal Investigator: Agnes Flöel[1,2]
Greifswald, 20.01.2020 _____
Place, Date, Signature

Biostatistician: Ulrike Grittner[3,4]
Greifswald, 20.01.2020 _____
Place, Date, Signature

Investigators and Co-Authors of SAP:
Daria Antonenko[1]
Greifswald, 20.01.2020 _____
Place, Date, Signature

Friederike Thams[1]
Greifswald, 20.01.2020 _____
Place, Date, Signature

Affiliations:
[1] Department of Neurology, Universitätsmedizin Greifswald, Greifswald, Germany,
[2] German Centre for Neurodegenerative Diseases (DZNE) Standort Greifswald, Greifswald, Germany.
[3] Berlin Institute of Health (BIH), Berlin, Germany,
[4] Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Biometry and Clinical Epidemiology, Berlin, Germany,

Content

1 Study Background

1.1 Study Objective

Developing interventions against age-associated cognitive decline is especially important, given the increase of aging populations around the world. Implementing combined cognitive training and transcranial direct current stimulation (tDCS) interventions may support training effects and enable transfer to other cognitive domains (Jones et al., 2015; Perceval et al., 2016; Berryhill, 2017; Antonenko et al., 2018a). Therefore, effects of a combined cognitive training and tDCS intervention in healthy older adults are tested.

The aim of the TrainStim-Cog trial (see Antonenko et al. (2019) for study protocol) is to investigate the immediate and delayed behavioral and neural effects of a three-week combined tDCS and cognitive training intervention compared to cognitive training and sham tDCS in healthy older adults. The analyses described in this statistical analysis plan (SAP) will demonstrate the efficacy of a three-week cognitive training intervention with concurrent tDCS in healthy older adults. This SAP was prepared in accordance with the *Guidelines for the Content of Statistical Analysis Plans in Clinical Trials* (Gamble et al., 2017).

1.2 Primary hypothesis

The primary hypothesis of the project is that the combination of cognitive training and tDCS is superior with regard to cognitive performance outcomes compared to cognitive training alone in healthy older adults operationalized by the score of the letter updating task after 3 weeks of intervention (post assessment).

1.3 Secondary hypotheses

Secondary hypotheses state that the combination of cognitive training and tDCS is superior compared to cognitive training alone with regard to cognitive training tasks, transfer tasks, MRI measures at all follow up measures as defined by the secondary outcomes in healthy older adults.

1.4 Study Design

The TrainStim-Cog trial is a randomized, single-blind, placebo-controlled monocenter study. The experimental group receives a nine-session cognitive training intervention over three weeks, accompanied by tDCS over the left dorsolateral prefrontal cortex (DLPFC). The intervention of the control group consists of the same nine-session cognitive training combined with sham stimulation.

Random allocation to experimental and control groups, respectively, will be performed with a 1:1 ratio. Stratified block randomization will be used. Strata will be chosen according to age (median split) and initial performance in the letter updating task (median split). After successful completion of telephone screening and baseline assessment (defined as meeting all eligibility criteria) and giving written informed consent, participants will be divided into four groups according to the age and performance

strata. For the randomization we will use the blockrand package in R[1].

1.5 Sample Size Calculation

Based on recent studies in the field using multi-session application of anodal tDCS during cognitive training compared to training with sham tDCS (Park et al., 2014; Jones et al., 2015; Antonenko et al., 2018), we estimated an effect size of 0.85. To demonstrate an effect in the primary outcome, 46 participants (23 per group) need to be included in the analysis with an independent t-test using a two-sided significance level of 0.05 and a power of 80%. This conservative approach using a t-test was chosen, even though we intend to analyse the primary outcome conducting analysis of covariance (ANCOVA) models (Borm et al., 2007). Assuming a drop-out rate of about 20% due to a high number of planned visits and considerably high demands put upon participants (e.g., performing challenging memory tasks and attending three sessions of 45 min MRI scans), 28 participants should be included in each tDCS group.

2 Analysis sets

2.1 Definitions

The **full analysis set** will consist of all participants who received at least one day of intervention. In case participants withdraw informed consent after baseline assessment, they will be considered as screening failures and therefore will not be included in the full analysis set. The **per protocol analysis set** comprises all subjects who received the full three weeks intervention or control intervention and completed all visits in the treatment groups they were allocated to. Safety measures will be assessed during tDCS intervention and all participants who received at least one intervention will be included according to their actual treatment in **the safety analysis set**. Since no participant received other treatment as intended or switched treatment groups during the study, and no information on safety measure is available for participants who missed intervention or follow up visits or dropped out, the safety analysis set is the same as the per protocol analysis set in this study.

2.2 Application

The primary efficacy analysis will be done using the full analysis set including estimated values from multiple imputations for missing values (Intention to treat). An analysis of the primary outcome in the per protocol analysis set will be used as sensitivity analysis. For the safety analysis, analysis will be done in the safety analysis set, which is the same as the per protocol analysis set.

3 Trial centres

Participants will be recruited in one centre: Greifswald

3.1 Recruitment

Participants will be recruited through advertisements in the local newspapers and distribution of flyers

[1] http://www.R-project.org, http://www.rstudio.com, https://CRAN.R-project.org/package=blockrand

in local senior citizen clubs. Telephone screenings will be conducted with all potential participants and study information will be provided. Eligible candidates will be invited for baseline assessment. Following baseline assessment (V0) participants will be included if neuropsychological testing is unobtrusive.

Information on recruitment flow can be found in the CONSORT flow diagram (Figure 1).

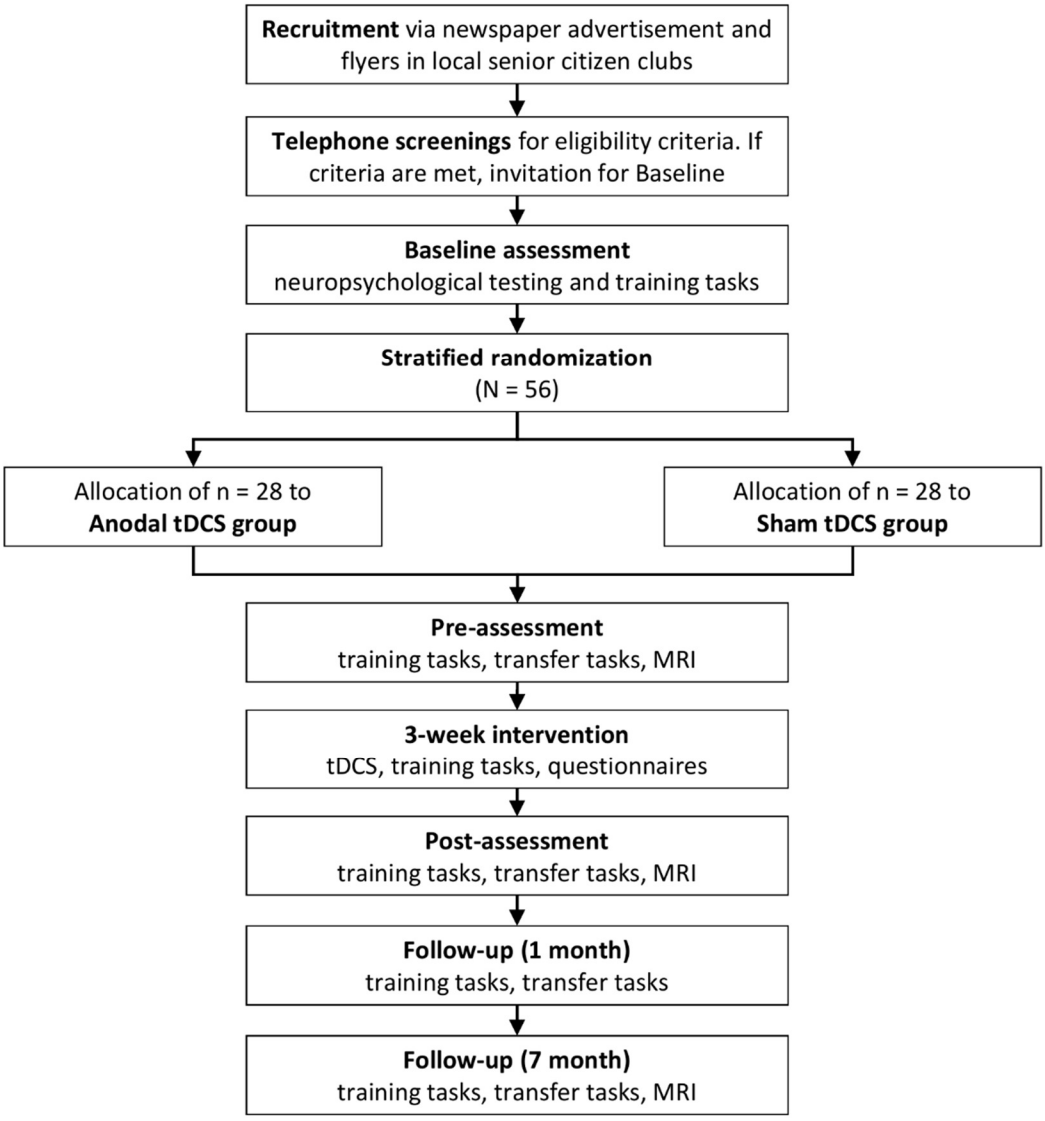

Figure 1. TrainStim-Cog study flowchart. tDCS, transcranial direct current stimulation; MRI, magnetic resonance imaging. Obtained from Antonenko et al. (2019).

4 Analysis variables

Table 1. TrainStim-Cog outcome measures.
Adapted from TrainStim-Cog study protocol (Antonenko et al., 2019)

			Post-allocation					
			Base line	Pre	T1-T9 3 weeks	Post 3 days	FU 1 month	FU 7 months
			~ 2h	~ 3h	~ 1h	~ 3h	~ 3h	~ 3h
Time point	Measurement	Mode	V0	V1	V2-V10	V11	V12	V13
Enrollment								
Eligibility screening		Paper	x					
Informed consent		Paper	x					
Neuropsychological Screening	Demographic data	Paper	x					
	Geriatric Depression scale (Brink et al., 2013)	Paper	x					
	Oldfield handedness inventory (Oldfield, 1971)	Paper	x					
	CERAD Plus Testbattery (memoryclinic.ch)	Paper	x					
	Digit span (Lezak et al., 2004)	Paper	x					
	Identical pictures (Lindenberger and Baltes, 1997) Spot-a-word tasks (Lehrl, 1977)	Computer	x					
Intervention					♦——♦			
Training tasks	Letter updating (Dahlin et al., 2008)	Tablet-PC	x	x	x	x	x	x
	Markov decision making (Eppinger et al., 2015; Wittkuhn et al., 2018)	Computer	x	x	x	x	x	x
Brain stimulation	tDCS (tDCS vs. sham)	Device			x			
Questionnaires	Self-reported well-being	Paper		x	x	x	x	x
	PANAS (Watson et al., 1988)	Paper			x			
	Adverse Events Questionnaire* (Antal et al., 2017)	Paper			x			
Additional assessments								
Transfer tasks	n-back	Computer		x		x	x	x
	AVLT (Helmstaedter et al., 2001; Lezak et al., 2004)	Paper		x		x	x	x
	Wiener Matrices Test 2 (Formann et al., 2011)	Paper		x		x	x	x
	Virtual reality task (Hartley et al., 2003)	Computer		x		x	x	x
Physical measures	MRI			x		x		x
	Blood draw			once at any of these sessions				

T1-T9, training 1-9. FU, follow-up-assessment. V0-V13, visits 0-13. CERAD: The Consortium to Establish a Registry for Alzheimer's Disease, neuropsychological battery. PANAS, Positive and negative affect schedule. VLMT, Verbaler Lern- und Merkfähigkeitstest (German version of the auditory verbal learning test). tDCS, transcranial direct current stimulation. MRI, magnetic resonance imaging. All measures were acquired on site, except for screening, which was done via telephone. * assessed only at the end of each training week (V4, 7,10).

4.2 Primary Outcome

Performance in the letter updating task at post-assessment, operationalized by number of correctly recalled lists (maximum 18 lists) will be the primary outcome measure.

4.3 Secondary Outcomes

At post- and follow-up assessments (V11, V12, V13) the following secondary outcome measures will be analyzed:

Training tasks

- Number of correctly recalled lists (as secondary outcome at follow-up sessions)
- Proportion of optimal actions in the Markov decision-making task

Transfer tasks

- Performance on numeric n-back task (% correct, d-prime)
- Performance on German version of the auditory verbal learning test (Helmstaedter et al., 2001; Lezak et al., 2004) (total amount of words learned, number of recalled words at delayed recall)
- Performance on Wiener matrices test (WMT-2) (Formann et al., 2011) (% correct)
- Performance on a virtual reality maze task (Hartley et al., 2003) (number of items found on a previous encoded route)

  All transfer measures will be corrected for performance at pre-assessment.

MRI measures

- structural neural correlates; assessed by grey matter volumes, cortical thickness, white matter microstructure (diffusion tensor imaging, DTI)
- functional neural correlates; assessed by resting-state fMRI analyses to obtain functional connectivities

Additionally, for both training tasks, effects during the intervention (V2-V10) will be analyzed:

- online effects; assessed by within session performance changes
- offline effects; assessed by performance changes from the last trial of the previous visit to the first trial of the next visit
- direct interventional effects; assessed as performance change from first to last training session (learning curves)

4.4 Safety Outcomes

Safety parameters are assessed via self-report questionnaire every third day of training (V4, V7, V10). The questionnaire was adapted from Antal et al. (2017) and includes intensity ratings with regard to itching, pain, burning, warmth/heat, metallic/iron taste, fatigue/decreased alertness and other sensations.

5 Handling of missing values

In case of missing values and under the assumption of missing at random (MAR) or missing completely at random (MCAR) as missing data mechanism, data will be estimated using multiple imputation methods with 30 imputed data sets. To estimate values in a realistic range and with values similar as in complete cases, we will use predictive mean matching.

6 Statistical analyses

For all analyses (including analysis of primary outcome) appropriate descriptive statistics (mean, standard deviation, median, interquartile range, absolute and relative frequencies) depending on the scale and distribution of the outcome variable will be presented.

Statistical analyses will be divided to analyze

1. immediate treatment effects by including all measures until include V11 (post assessment)
2. long-term treatment effects by focusing on V12 (1 month follow up) and V13 (7 months follow up)

6.1 Primary analysis

Using linear mixed models the measures of the letter updating task over the study period until include V11 (post assessment), will be used as dependent variable, stimulation group (tDCS, sham) as factor, and letter updating performance at pre-assessment as well as age as covariates. The primary outcome (letter updating task score at post assessment) will be evaluated between treatment groups based on this regression model via marginal means. We will use random intercept models that account for the clustering of measures within individuals.

6.2 Secondary analyses

Immediate treatment effects

Performance on the second training task (Markov decision-making task) will be analyzed in the same manner as the primary outcome, using linear mixed models for performance on the Markov decision-making task over the study period until include V11 (post-assessment) as dependent variable, stimulation group (tDCS, sham) as factor, and letter updating and Markov decision-making performance at pre-assessment as well as age as covariates. We will use random intercept models that account for the clustering of measures within individuals.

Transfer tasks and other secondary outcomes that are measured pre and post assessment will be compared in both groups at post-assessment (V11, dependent variable), using separate ANCOVA models for each outcome. In these models treatment allocation will be tested as covariate of interest. Age and sex as well as interaction terms will be included to adjust for possible confounders or to test subgroup differences. The pre-assessment value of the letter updating task and the particular pre-assessment value of the measure of interest will be used as covariates. Additionally a covariate for the time point of measure will be included.

Long-term treatment effects

Long term treatment effects will be analyzed using mixed models over the whole study period. These models will include the pre-assessment scores of the letter updating task and the respective measure of interest, age and sex as covariates and a random effect for the participant (random intercept). Type of link function (logistic, linear, ordinal) will depend on the scaling of the dependent variable. In case of skewed continuous data, variables will be transformed before analysis.

All secondary analyses will be done using the full analysis set with multiple imputed data in case of missing values. Per protocol analyses will be done as sensitivity analyses. All secondary analyses will be done in an exploratory framework.

Online and offline training effects

Analyses of online and offline training effects (Reis et al., 2009) for detailed examination of learning during training will be performed for the main measures of the two training tasks. Online learning is

defined as performance difference from beginning to end of a training task within each session. Offline effects will regard between session retention (overnight / over the weekend) and will be computed as performance difference from end of the previous session to the beginning of the next session. For the analysis of online training effects we will use the outcome directly after a training task as dependent variable over the whole training period as dependent variable in a linear mixed model (random intercept model). As independent variables we will use the pre-training measure of the specific training day, the pre-assessment value, age, sex, time point of measurement, group allocation. Offline effects will be analyzed similarly with measures over night / weekend after training as dependent variable, including the measure direct after training (from the day or some days before) as independent measure as well as the other covariates.

Analysis of MRI data

Structural and functional MRI data analyses will be performed using well-established pipelines from MATLAB-based toolboxes such as SPM (Statistical Parametric Mapping software, http://www.fil.ion.ucl.ac.uk/spm/), CONN toolbox (www.nitrc.org/projects/conn, Whitfield-Gabrieli and Nieto-Castanon, 2012), FSL (Analysis Group, FMRIB, Oxford, UK; fsl.fmrib.ox.ac.uk/fsl/fslwiki/, Jenkinson et al., 2012), the computational anatomy toolbox (CAT12, http://www.neuro.uni-jena.de/cat/) or Freesurfer (http://surfer.nmr.mgh.harvard.edu/). Functional connectivity within and between large-scale networks will be assessed using functional resting-state fMRI scans (Darki and Klingberg, 2015; Antonenko et al., 2018b). Segmentation on structural scans will be performed to assess volume of cortical and subcortical gray matter (Dahlin et al., 2008; Filmer et al., 2019) and white matter microstructure in white matter tracts will be extracted from diffusion-weighted images using common tractography methods (Charlton et al., 2010; Metzler-Baddeley et al., 2011; Le Bihan and Johansen-Berg, 2012; Metzler-Baddeley et al., 2017).

6.3 Safety/Tolerability

Safety outcomes will be reported separately as incidences (n, incidence rate with 95%CI) in total and by intervention group based on the safety analysis set. Participants will be grouped according to their actually received treatment. Incidence rates and 95%CI will be based on poisson regression models that account for the different observation periods for each participant. Group comparisons will be done using incidence rate ratios and 95%CI. Results of safety analysis will be interpreted and discussed thoroughly also for minor group differences, since statistically significance is not of importance here.

6.7 Planned subgroup analyses

For primary and secondary outcomes main subgroups analyses will be done by sex. Therefore as a first step we will include an interaction term of sex*intervention allocation in the regression models to test whether there are differential treatment effects with regard to sex. Similarly this will be done as first step for all subgroup analyses. All subgroup analyses will be done within an exploratory framework.

To further explore learning effects, we will perform sensitivity analyses using only measures on time points on which participants felt well enough, based on initial self-rating questionnaires at each visit. To obtain more detailed information on the two training tasks, exploratory models will be calculated on measures, such as performance dependent on list length in the letter updating task (Morris and Jones, 1990) or parameters from a drift diffusion model and change-points for the Markov decision-making task (Wagenmakers et al., 2007; Durstewitz et al., 2010; Eppinger et al., 2015; Wittkuhn et al., 2018).

To assess possible predictors of training task performance and responsiveness to the intervention, measures of cognitive reserve (e.g. education, baseline cognitive ability or neuropsychological status) will be entered into analyses.

6.8 Example table for the description of baseline characteristics

Table 2. Baseline characteristics of the study sample.

	All n =	TDCS group n =	Sham group n =
Age (years)			
Gender (n, % female)			
Education (years)			
GDS			
Semantic fluency			
BNT (max. 15)			
MMSE (max. 30)			
Word list learning			
Total (max. 30)			
Trial 1 (max. 10)			
Trial 2 (max. 10)			
Trial 3 (max. 10)			
Word list retrieval (max. 30)			
Word list intrusions			
Figure copying (max. 11)			
Figure retrieval (max. 11)			
Phonematic fluency			
Trail-making test			
Part A (sec)			
Part B (sec)			
Digit-span			
Forward			
Backward			
Identical pictures			
Accuracy			
RT			
Spot-a-word			
Accuracy			
RT			

Data are shown as the mean (SD) or n(%). GDS, Geriatric Depression Scale. BNT, Boston Naming Test. MMSE, Mini Mental Status Examination. RT, reaction time.

7. References

Antal A et al. (2017) Low intensity transcranial electric stimulation: Safety, ethical, legal regulatory and application guidelines. Clin Neurophysiol 128:1774-1809.

Antonenko D, Kulzow N, Sousa A, Prehn K, Grittner U, Floel A (2018a) Neuronal and behavioral effects of multi-day brain stimulation and memory training. Neurobiol Aging 61:245-254.

Antonenko D, Nierhaus T, Meinzer M, Prehn K, Thielscher A, Ittermann B, Flöel A (2018b) Age-dependent effects of brain stimulation on network centrality. NeuroImage 176:71-82.

Antonenko D, Thams F, Uhrich J, Dix A, Thurm F, Li SC, Grittner U, Floel A (2019) Effects of a Multi-Session Cognitive Training Combined With Brain Stimulation (TrainStim-Cog) on Age-Associated Cognitive Decline - Study Protocol for a Randomized Controlled Phase IIb (Monocenter) Trial. Front Aging Neurosci 11:200.

Berryhill ME (2017) Longitudinal tDCS: Consistency across Working Memory Training Studies. AIMS Neuroscience 4:71-86.

Brink TL, Yesavage J, Lum O (2013) Geriatric depression scale. Evidence-Based Diagnosis: A Handbook of Clinical Prediction Rules:297.

Charlton RA, Barrick TR, Lawes INC, Markus HS, Morris RG (2010) White matter pathways associated with working memory in normal aging. Cortex 46:474-489.

Dahlin E, Neely AS, Larsson A, Backman L, Nyberg L (2008) Transfer of learning after updating training mediated by the striatum. Science 320:1510-1512.

Darki F, Klingberg T (2015) The role of fronto-parietal and fronto-striatal networks in the development of working memory: a longitudinal study. Cereb Cortex 25:1587-1595.

Durstewitz D, Vittoz NM, Floresco SB, Seamans JK (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. Neuron 66:438-448.

Eppinger B, Heekeren HR, Li SC (2015) Age-related prefrontal impairments implicate deficient prediction of future reward in older adults. Neurobiology of aging 36:2380-2390.

Filmer HL, Ehrhardt SE, Shaw TB, Mattingley JB, Dux PE (2019) The efficacy of transcranial direct current stimulation to prefrontal areas is related to underlying cortical morphology. NeuroImage.

Formann A, Waldherr K, Piswanger K (2011) Wiener Matrizen-Test 2. Manual. Beltz Test GmbH.

Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Doré C, Williamson PR, Altman DG, Montgomery A, Lim P (2017) Guidelines for the content of statistical analysis plans in clinical trials. Jama 318:2337-2343.

Hartley T, Maguire EA, Spiers HJ, Burgess N (2003) The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. Neuron 37:877-888.

Helmstaedter C, Lendt M, Lux S (2001) Verbaler Lern-und Merkfähigkeitstest: VLMT; Manual: Beltz-Test.

Jones KT, Stephens JA, Alam M, Bikson M, Berryhill ME (2015) Longitudinal Neurostimulation in Older Adults Improves Working Memory. PLOS ONE 10:e0121904.

Le Bihan D, Johansen-Berg H (2012) Diffusion MRI at 25: exploring brain tissue structure and function. Neuroimage 61:324-341.

Lehrl S (1977) MWT-B: Mehrfach-Wahl-Wortschatz-Test B. Erlangen: Straube.

Lezak MD, Howieson DB, Loring DW, Fischer JS (2004) Neuropsychological assessment: Oxford University Press, USA.

Lindenberger U, Baltes PB (1997) Intellectual functioning in old and very old age: cross-sectional results from the Berlin Aging Study. Psychology and aging 12:410.

Metzler-Baddeley C, Jones DK, Belaroussi B, Aggleton JP, O'Sullivan MJ (2011) Frontotemporal connections in episodic memory and aging: a diffusion MRI tractography study. Journal of Neuroscience 31:13236-13245.

Metzler-Baddeley C, Foley S, de Santis S, Charron C, Hampshire A, Caeyenberghs K, Jones DK (2017) Dynamics of White Matter Plasticity Underlying Working Memory Training: Multimodal Evidence from Diffusion MRI and Relaxometry. J Cogn Neurosci 29:1509-1520.

Morris N, Jones DM (1990) Memory updating in working memory: The role of the central executive. British Journal of Psychology 81:111-121.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97-113.

Perceval G, Floel A, Meinzer M (2016) Can transcranial direct current stimulation counteract age-associated functional impairment? Neurosci Biobehav Rev 65:157-172.

Reis J, Schambra HM, Cohen LG, Buch ER, Fritsch B, Zarahn E, Celnik PA, Krakauer JW (2009) Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. Proc Natl Acad Sci U S A 106:1590-1595.

Wagenmakers E-J, Van Der Maas HLJ, Grasman RPPP (2007) An EZ-diffusion model for response time and accuracy. Psychonomic Bulletin & Review 14:3-22.

Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of personality and social psychology 54:1063.

Wittkuhn L, Eppinger B, Bartsch LM, Thurm F, Korb FM, Li SC (2018) Repetitive transcranial magnetic stimulation over dorsolateral prefrontal cortex modulates value-based learning during sequential decision-making. Neuroimage 167:384-395.