Jessica Oschwald*, Sabrina Guye, Franziskus Liem, Philippe Rast, Sherry Willis, Christina Röcke, Lutz Jäncke, Mike Martin and Susan Mérillat

Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change

https://doi.org/10.1515/revneuro-2018-0096
Received September 17, 2018; accepted March 2, 2019; previously published online June 5, 2019

Abstract: Little is still known about the neuroanatomical substrates related to changes in specific cognitive abilities in the course of healthy aging, and the existing evidence is predominantly based on cross-sectional studies. However, to understand the intricate dynamics between developmental changes in brain structure and changes in cognitive ability, longitudinal studies are needed. In the present article, we review the current longitudinal evidence on correlated changes between magnetic resonance imaging-derived measures of brain structure (e.g. gray matter/white matter volume, cortical thickness), and laboratory-based measures of fluid cognitive ability (e.g. intelligence, memory, processing speed) in healthy older adults. To theoretically embed the discussion, we refer to the revised Scaffolding Theory of Aging and Cognition. We found 31 eligible articles, with sample sizes ranging from n = 25 to n = 731 (median n = 104), and participant age ranging from 19 to 103. Several of these studies report positive correlated changes for specific regions and specific cognitive abilities (e.g. between structures of the medial temporal lobe and episodic memory). However, the number of studies presenting converging evidence is small, and the large methodological variability between studies precludes general conclusions. Methodological and theoretical limitations are discussed. Clearly, more empirical evidence is needed to advance the field. Therefore, we provide guidance for future researchers by presenting ideas to stimulate theory and methods for development.

Keywords: brain structure; change; cognitive ability; correlated change; healthy aging; longitudinal.

***Corresponding author: Jessica Oschwald,** University Research Priority Program 'Dynamics of Healthy Aging', University of Zurich, Andreasstrasse 15, CH-8050 Zurich, Switzerland, Phone: +41 44 634 52 19, e-mail: jessica.oschwald@gmail.com
Sabrina Guye, Franziskus Liem, Christina Röcke and Susan Mérillat: University Research Priority Program 'Dynamics of Healthy Aging', University of Zurich, Andreasstrasse 15, CH-8050 Zurich, Switzerland
Philippe Rast: Department of Psychology, University of California, Davis, 1 Shields Avenue, Davis, CA 95616, USA
Sherry Willis: Seattle Longitudinal Study, Department of Psychiatry and Behavioral Sciences, University of Washington, 2500 Sixth Ave N., Apt. 1, Seattle, WA 98195, USA
Lutz Jäncke: University Research Priority Program 'Dynamics of Healthy Aging', University of Zurich, Andreasstrasse 15, CH-8050 Zurich, Switzerland; and Department of Psychology (Neuropsychology Unit), University of Zurich, Binzmuehlestrasse 14, CH-8050 Zurich, Switzerland
Mike Martin: University Research Priority Program 'Dynamics of Healthy Aging', University of Zurich, Andreasstrasse 15, CH-8050 Zurich, Switzerland; and Department of Psychology (Gerontopsychology Unit), University of Zurich, Binzmuehlestrasse 14, CH-8050 Zurich, Switzerland

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). With an extended lifespan, it is increasingly important to understand how these additional years of life can be spent in good health. To foster research in this matter, the WHO recently announced the 'Decade of Healthy Aging' from 2020 until 2030, defining healthy aging as the 'process of developing and maintaining the functional ability that enables well-being in older age' (WHO, 2015). Cognitive health is of high importance for aging healthily (Lawton et al., 1999), with a substantial impact on tasks of independent living (Salthouse, 2012), such as medication adherence (Insel et al., 2006), telephone use, financial management, or nutritional choices (Gregory et al., 2009). For the present article, we therefore limit our definition of healthy aging to the cognitive domain. Specifically, we refer to aging processes that occur in the absence of pathological cognitive impairments, as previous literature has not yet reached a consensus on the definition of healthy cognitive aging.

Previous research has identified cognitive frailty as one of the most important threats for well being in healthy aging with an enormous impact on the decision to discount hypothetical years of life (Lawton et al., 1999)

and linked cognitive frailty to the degradation of neural mechanisms. Specifically, a vast number of studies have focused on the relationship between indicators of brain function derived from task-related functional magnetic resonance imaging (MRI) studies and task-based indicators of cognitive ability (e.g. episodic memory, processing speed, working memory) in healthy aging individuals (see Grady, 2012, for an overview). Much less is known, however, about the association between measures of brain structure and cognitive ability, although brain structure represents the hardware on which brain function is implemented (the neurobiological relevance of these structural measures is discussed in Box 1). The current article aims to shed more light on this relationship.

The association between brain structure and cognitive ability can be illuminated from different perspectives, depending on the research question one is interested in. Whereas the choice of the research question is mainly driven by theoretical considerations, the type of data, and methods available constrain whether one is able to address it (see Boker and Martin, 2018 for an in-depth discussion). As follows, we will discuss four research questions that are of theoretical relevance regarding the relation between brain structure and cognitive ability in healthy aging, and relate them to the data and methods needed to answer them, using Catell's (1988) data box. Figure 1 shows how the three dimensions of the data box (persons × variables × measurement occasion) are related to the different possible research questions.

Box 1: Neurobiological foundation of age-related change in brain structure and methodological advances in neuroimaging.

Recent reviews of cross-sectional brain imaging studies summarizing data from many subjects as well as large-scale longitudinal brain imaging studies report mean percentage changes per year for GM and WM ranging between 0.5% and 0.8% (Fjell and Walhovd, 2010; Ritchie et al., 2015b). The age-related GM loss may result from several neuroanatomical changes comprising loss of neuropil (unmyelinated axons, dendrites, and glial cells), shrinking of neural bodies, changes to the dendritic morphology (e.g. decline in the number of dendritic spines, shortening of dendritic shafts, and reduction of dendritic branching), or a decrease in synaptic density, probably indicating a loss in the number of synapses (Pannese, 2011; Juraska and Lowry, 2012). Furthermore, age-related degradation of WM may result from axonal degeneration, myelin changes (e.g. demyelination, deformation of the morphological structure), or other changes, such as glial scars or accumulation of cellular debris (Juraska and Lowry, 2012; Bennett and Madden, 2014). Overall, the cellular foundation of age-related neural changes as captured by MRI is still poorly understood, as the current knowledge is mostly based on animal research and post-mortem studies with humans. Nevertheless, several authors argue that the age-related brain tissue loss might be one of the reasons why cognitive functions decline on average with ongoing age. One technique that has become very popular due to its ease of use and its potential to study human brain tissue microstructure *in vivo* is diffusion-weighted MRI (DW-MRI). Tensor-derived diffusion indicators (see Table 1) are now frequently used to study WM microstructural changes in aging. However, these measures are difficult to interpret with regard to their biological basis, as many factors of the complex WM architecture (e.g. crossing fibers, glial cells) can modulate diffusion properties (Jones, 2010; Concha, 2014). While most of the current *in vivo* literature on neuroanatomical aging relies on T1-weighted and/or tensor-derived diffusion measures, several new promising brain imaging techniques are emerging, which might provide a more detailed view into the macroanatomical and microanatomical age-related changes.

One alternative to tensor-derived diffusion measures is neurite orientation dispersion, which allows the *in vivo* estimation of the microstructural characteristics of axons and dendrites (Zhang et al., 2012). Furthermore, Myelin Water Fraction as modeled based on the T2 relaxation properties of water captured between myelin sheets was shown to provide more specific estimates of myelin content than the tensor-based diffusion parameters (Arshad et al., 2016). Particularly promising are recent advances in quantitative MRI, which are computational methods that allow for the derivation of voxel-wise quantitative maps of MRI biomarkers, reflecting specific microstructural tissue properties, such as iron, myelin content, or axonal fiber orientation (Draganski et al., 2011; Weiskopf et al., 2015). In addition, scanners with ultrahigh field strengths of 7 Tesla are now available for practical use. While these scanners allow an increased signal-to-noise ratio and thus very high spatial resolution of brain images, their applicability is limited by a number of challenges, for example, an increased sensitivity to motion artifacts, inhomogeneities in the magnetic and radiofrequency field, and an increased specific absorption rate (Barisano et al., 2019).

Besides advances in MRI techniques, novel methods to process and quantify brain-imaging data hold promise for the study of brain structure-cognition associations in the future. For example, network connectivity and graph analysis methods allow the inference of information about organizational properties of structural brain networks based on structural MRI and DW-MRI data (see Bullmore and Sporns, 2009). These methods are particularly relevant for the cognitive neurosciences, as they can map network properties that are probably more reflective of the complexity of the underlying cognitive abilities than single structural brain measures. Another interesting development is the use of machine learning techniques to predict individual's biological age on the basis of structural brain imaging data (i.e. brain age) (Gaser et al., 2013; Valizadeh et al., 2017; Cole et al., 2018). Brain age prediction can provide important insights into potential biomarkers associated with premature brain aging and neurocognitive disorders. For example, higher brain age than chronological age is associated with increased mortality risk (Cole et al., 2018) and cognitive impairment (Liem et al., 2017). Brain age is found to be a superior predictor of later dementia conversion compared to common cognitive tests or CSF-derived biomarkers (Gaser et al., 2013).

Table 1: Glossary of neuroanatomical measures.

Neuroanatomical measure	Description
Gray matter (GM)	
Volume-based	
GM volume	Total GM volume consists of the neuropil (neuronal bodies and dendrites), glial cells, axons, and vasculature (Zatorre et al., 2012). Cortical GM volume is the product of cortical thickness and surface area (see surface-based measures), whereas subcortical volumes refer to GM volumes in compartmental structures (e.g. hippocampus)
GM density	GM density reflects a proportional measure of GM concentration at a singular voxel, which is standardized on the average GM concentration at neighboring voxels. GM density can be measured with voxel-based morphometry (Ashburner and Friston, 2000)
Surface-based	
Cortical thickness	The thickness of the cortex is measured as the distance between GM and WM surface. Cortical thickness is suggested as the primary driver of plasticity in GM volume (Hogstrom et al., 2013)
Surface area	Measure of the surface of WM at the boundary to cortical GM
White matter (WM)	
Volume-based	
WM volume	Total WM volume consists of myelinated and unmyelinated axons and glial cells (Walhovd et al., 2014). Axonal fiber bundles are arranged as pathways connecting intrahemispheric (association), cortical with subcortical regions (projection), and the cortical hemispheres with each other (commissural)
WM hyperintensity (WMH) volume	Volume of regions of increased signal intensity in WM of structural magnetic resonance imaging (MRI) images. WMH are linked to demyelination and axonal degradation due to small vessel disease (Prins and Scheltens, 2015)
Diffusion-based	
Fractional anisotropy (FA)	Index for the directedness of water diffusion independent of the rate of diffusion. Values range between 0 (low directedness, isotropic diffusion) and 1 (high directedness, anisotropic diffusion). As the microstructural properties of dense fiber tracts constrict water diffusion along one direction, high FA values are often interpreted as a measure for the intactness of WM fibers. FA is modulated by the packing density of axons, axon diameter, and myelin (Beaulieu, 2002)
Diffusivity	Mean diffusivity (MD) describes the average rate of diffusion independent of direction and is thus a summary measure. In contrast, axial diffusivity (AD) describes the rate of diffusion parallel, and radial diffusivity (RD) perpendicular to the main axis. AD has been suggested to be modulated by axon integrity, while RD is more sensitive to myelin (but see Box 1; Song et al., 2002)
Cerebrospinal fluid (CSF)	
Ventricular volume	Ventricular volume is typically used as a proxy of CSF, which fills the ventricular compartments of the brain and the subarachnoidal space. Enlarged ventricles result in increased CSF volume
Global measures	
Whole brain volume	
Total brain volume (TBV)	Sum of GM and WM volume, whereas the brainstem is typically excluded
Intracranial volume (ICV)	Sum of brain tissue (GM and WM), CSF, brain membranes, and volume between skull and membranes
Normalized brain volume (NBV)	TBV adjusted for an estimate of head size (e.g. ICV)

Research question type 1 (cross-sectional/univariate, see panel A): 'Do people differ in specific measures of brain structure or cognitive ability?'

To answer this question, cross-sectional data of several participants measured in indicators of brain structure (e.g. whole brain volume) or cognitive ability (e.g. working memory performance) are needed. As illustrated by the gray shaded cubes, this type of data varies along the dimension persons and is fixed along the dimensions variables × measurement occasions. Brain structure serves as an example for the selected variable here; however, cognitive ability could be used interchangeably. The measure of interest is the mean and the variance between persons (interindividual differences) in a measure of brain structure (or cognitive ability). Methodologically, this variance component can then be related to predictors by using, for example, regression analysis. A relevant predictor if one is interested in healthy aging is chronological age: if participants are sampled at different ages, the between-person variance also contains information about age-differences. For example, after controlling for height and sex, a study found age-related differences in a sample of participants aged from 18 to 77 years in several regional brain volumes with a specific vulnerability of the prefrontal cortex, such

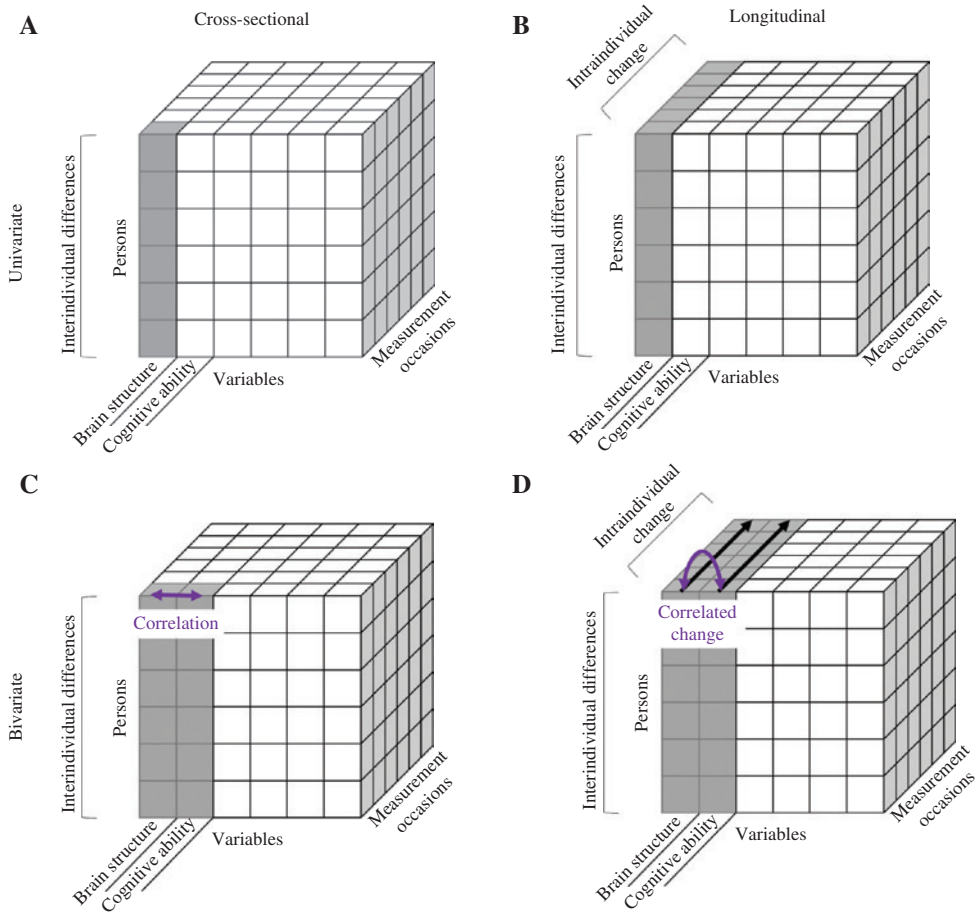

Figure 1: Four different research questions on the relation between brain structure and cognitive ability, as illustrated with Cattell's (1988) data box.
Panel A: research question type 1 refers to interindividual differences in a measure of brain structure or cognitive ability assessed at one measurement occasion. Panel B: research question type 2 refers to intraindividual changes in a measure of brain structure or cognitive ability assessed across several measurement occasions. Panel C: research question type 3 refers to the bivariate association between interindividual differences in a measure of brain structure and interindividual differences in a measure of cognitive ability assessed at one measurement occasion (correlation). Panel D: research question type 4 refers to the bivariate association between intraindividual change in a measure of brain structure and intraindividual change in a measure of cognitive ability assessed across several measurement occasions (correlated change).

that older adults had smaller prefrontal gray matter (GM) volumes than younger adults (Raz et al., 1997).

It is well documented that both brain structure and cognitive ability are not stationary, but subject to dynamic changes over the lifespan (Deary, 2001; Hedden and Gabrieli, 2004; Fjell and Walhovd, 2010; Salthouse, 2010). As the cross-sectional design contains information about interindividual differences between persons of different chronological ages, it is not a viable basis for the inference of change processes across time (Lindenberger et al., 2011). Furthermore, cohort differences are a common problem in cross-sectional designs, masking the effects of true change within individuals (Sliwinski et al., 2010). For example, several recent studies report a reduction of the prevalence of Alzheimer's disease (AD) over the past decade when comparing older adults of different cohorts, but similar age range (e.g. 75-year-olds in 2000 vs. 2012) (Larson et al., 2013; Matthews et al., 2016; Langa et al., 2017), suggesting cohort differences in age-related brain and cognitive changes. Although less problematic, cohort differences can also confound longitudinal estimates of change, especially when these studies include a wide age range (Hofer and Sliwinski, 2001). While longitudinal studies are faced with their own limitations, such as attrition (e.g. Lindenberger et al., 2002) or practice effects due to repeated cognitive testing (e.g. Salthouse et al., 2004), longitudinal measurements are necessary to make valid inferences

on developmental change (Raz and Lindenberger, 2011), leading to the next type of question:

Research question type 2 (longitudinal/univariate, see panel B): 'Do measures of brain structure or cognitive ability change over time within persons?'
To answer this question, longitudinal data of several people repeatedly measured in indicators of brain structure (or cognitive ability) are needed. In contrast to the previous example, the observations shaded in gray now also vary along the dimension measurement occasions (besides the persons dimension). The dimension variables is still held constant to one level, in this example it is again a measure of brain structure, but cognitive ability can be used as well. This type of data represents interindividual differences in how the values of one person in brain structure (or cognitive ability) change across measurement occasions within this individual (intraindividual change). Methodologically, such data can be analyzed in two ways. First, one can compute a difference score for each person to represent intraindividual change between two measurement occasions. This procedure reduces the dimension 'measurement occasion' to one value, allowing the application of the same methods as in research question type 1. The resulting outcome carries information about changes in interindividual differences. A second possibility is to use sophisticated methods for the analysis of change. With these methods, it is possible to analyze intraindividual change trajectories, and interindividual differences in intraindividual change, with the advantage of retaining all values along the dimension measurement occasion in the analysis, as well as all persons, regardless of how much longitudinal information they provide (see section 'The benefits of longitudinal designs' for more detail).

Taken together, these two types of data complement each other to generate valuable insights into how brain structure and cognitive ability each develop across the lifespan. As the cross-sectional design is less time and cost intensive, it has the advantage of informing about age-differences across a wide age range. In contrast, the feasibility of a longitudinal study covering the entire adult lifespan is highly unlikely, especially when studying brain aging, as MRI scanners would need to endure over a period of around 60 or 70 years. Even if this would be possible, technological advances would most likely result in the collected data being outdated and no longer meaningful. Longitudinal data are needed, however, to examine developmental change processes, both in the individual domains as well as on their associated changes. The next two questions focus on the association of brain structure and cognitive ability:

Research question type 3 (cross-sectional/bivariate, see panel C): 'Do people with more intact structural brain features demonstrate higher (or lower) levels of specific cognitive abilities'?
This question builds upon research question type 1, with the only difference that now two variables are included: brain structure and cognitive ability. Thus, the measured values vary along the dimensions persons × variables, and measurement occasions is held constant. The association of interest can be calculated via the correlation between these two dimensions, as indicated by the purple arrow.

As implied by the question above, cross-sectional data convey information about how specific structural brain characteristics and levels of cognitive ability are related in the population. However, another perspective that we deem specifically important and that has not received enough attention yet focuses on how change processes in brain structure and cognitive ability are differentially related within individuals:

Research question type 4 (longitudinal/bivariate, see panel D): 'Are changes in measures of brain structure differentially associated with changes in measures of cognitive ability within persons?'
This question is based upon research question type 2, with the only difference that two variables are assessed longitudinally and the two trajectories are related to one another. In the data box, the measured values vary along all dimensions: persons × variables × measurement occasions. The association of interest is the correlation between changes in these two dimensions as shown by the purple arrow, which is henceforth defined as correlated change. Building on the methods described in research question type 2, correlated change can be computed by correlating either the difference scores, or the intraindividual change slopes between two variables with each other across people. Different types of correlated changes can be distinguished depending on the sampling intensity and timing of measurements in each domain (see section 'A theoretical framework on longitudinal brain-cognition-environment interactions' for a detailed explanation). With a high number of repeated measurement occasions in both domains, intraindividual change slopes could also be correlated entirely within individuals. However, to date, only a small number of longitudinal studies have investigated correlated changes in brain structure and cognition, with typically low sampling intensity. This is not surprising, due to the financial and time-consuming expenses of conducting longitudinal MRI studies. Correlated changes between brain structure and cognitive ability are of major interest, however, as they

can provide new insights into the intricate developmental dynamics and interactions between these two domains as people age. This knowledge is especially important for the development of personalized interventions to promote better health and well-being in old age. Therefore, the present article will review the literature on correlated brain-cognition changes (section 'Literature review') in a broader context of a current theoretical model on brain-cognition-environment relations, which will be presented in the next section. Furthermore, the present article will discuss methodological limitations of the reviewed studies and present ideas for method development and application (section 'Methodological limitations and the need for method development'). Of specific relevance to the field of cognitive neuroscience, statistical methods for handling longitudinal neuroimaging and cognitive data are presented and explained (section 'The benefits of longitudinal designs'), and issues related to the handling of big data are discussed (section 'Handling and profiting from big data'). Finally, we present several trends and ideas for the development of theories on correlated changes between brain structure and cognitive ability in the future (section 'Theoretical limitations and the need for theory development').

A theoretical framework on longitudinal brain-cognition-environment interactions

One of the most pressing questions in the cognitive neuroscience of aging is to explain why some healthy aging individuals experience drastic age-related cognitive decline while others can maintain their levels of cognitive ability. Accordingly, several theoretical concepts have emerged that revolve around the idea that aging individuals may differ with regard to compensatory resources that support the maintenance of cognitive performance in the face of age-related brain degeneration (Reuter-Lorenz and Park, 2010; Park and Festini, 2016). For example, the different theoretical ideas of reserve assume that people differ in either neural capacity (brain reserve: e.g. number of neurons or brain size) or cognitive processing mechanisms (cognitive reserve: e.g. mental flexibility, strategy use) that allow them to cope with pathological brain damage, and thus stave off detrimental impacts on cognitive ability (for a detailed explanation see Stern, 2002, 2009). One multifactorial theoretical model that includes the dynamic interrelations between environmental variables, brain structure and function, and cognitive ability is the revised Scaffolding Theory of Aging and Cognition (STAC-r) model (Reuter-Lorenz and Park, 2014). STAC-r is a regulatory model, which assumes that compensatory mechanisms, termed compensatory scaffolding, can directly regulate the impact that brain structure or function changes exert on cognitive ability. The exact mechanisms of compensatory scaffolding are not clearly established, however, the authors suggest scaffolding to reflect a form of positive brain plasticity (Reuter-Lorenz and Park, 2014). For example, from functional imaging studies, it is known that healthy older as compared to younger adults recruit different brain regions (e.g. functional over-recruitment of prefrontal regions, or bilateral overactivation) during demanding cognitive tasks (see Eyler et al., 2011, for a review). These distinct functional activation patterns are interpreted as compensatory, if they are also related to better memory performance. Furthermore, structural brain reorganization, such as (to a limited extent) neurogenesis, synaptic, or axonal changes (Zatorre et al., 2012), or the use of different cognitive strategies (Stern, 2002, 2009), may potentially also serve a compensatory function. To illustrate STAC-r as a framework for the current literature review, Figure 2 displays the key parts of the model. Please note that we introduced a small adaptation to tailor the model to the specific focus of this review: the broad term 'cognitive function' of the original model is changed in Figure 2 to the more specific term 'cognitive ability,' as we will refer only to basic cognitive abilities as assessed by cognitive tests in the laboratory and not to cognitive functioning, for example, during daily activities like solving a crossword puzzle (Verhaeghen et al., 2012). The factors of specific interest in the present article are highlighted in white.

According to STAC-r, different pathways can be distinguished through which brain structure might be related to cognitive ability. Brain structure can affect cognitive ability via a direct pathway. For example, brain atrophy in old age might be linked to declines in cognitive ability. This is the hypothesis typically posited in the literature. Additionally, the model proposes an indirect pathway, in which the relation between brain structure and cognitive ability is shaped via compensatory scaffolding. For example, during a difficult cognitive task, additional brain networks might be recruited to compensate for age-related structural alterations in the primary network. First attempts were made to study this indirect path between brain structure, function, and cognitive ability longitudinally. These studies suggest a link between age-related structural brain reductions and increased functional activation in healthy older adults (Hakun et al., 2015; Fjell et al., 2016; Pudas et al., 2018; Vidal-Piñeiro et al., 2018).

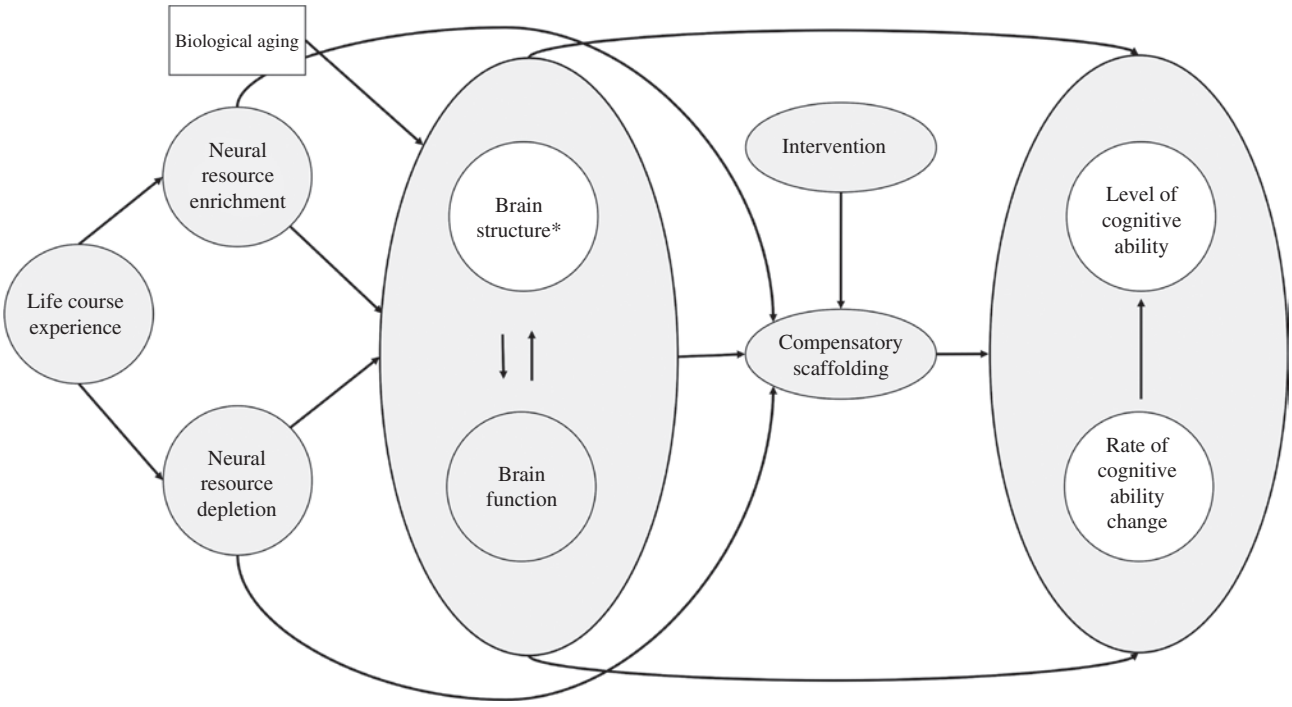

Figure 2: Scaffolding Theory of Aging and Cognition (STAC-r) model adapted from Reuter-Lorenz and Park (2014).
*Under brain structure we subsume both structural brain properties and rate of brain structure change.

However, with regard to the association with cognitive ability, it is still unclear whether these findings reflect compensation or rather an age-related loss of efficiency (Pudas et al., 2018). Arguing from the theoretical basis proposed by STAC-r, as long as compensatory scaffolding mechanisms function properly, the relationships between rate of brain structure and cognitive ability change would be expected to be weakened or even close to zero in healthy older adults, as compensatory scaffolding can buffer the immediate impact of brain structure deficits within individuals. Hence, the indirect pathway between brain structure and cognitive ability is especially relevant when focusing on the population of healthy older adults. The STAC-r model further proposes that compensatory scaffolding is impacted by a variety of factors. Besides changes in brain structure or function, neurally enriching or depleting experiences, termed life-course experiences, are assumed to stimulate scaffolding across the lifespan (Reuter-Lorenz and Park, 2014). Neurally enriching factors relate to activities or behaviors which positively stimulate brain plasticity, such as education, physical exercise, or multilingualism, while neurally depleting factors denote activities or influences that have a detrimental impact on the brain, such as high blood pressure, smoking, or stress (see Hertzog et al., 2009). As depicted in Figure 2, life-course experiences can either directly influence brain plasticity or stimulate compensatory scaffolding and thus potentially attenuate or delay cognitive decline. Also, interventions, for example, in the form of cognitive training or neurofeedback, can directly trigger compensatory scaffolding (Reuter-Lorenz and Park, 2014).

Using STAC-r as a theoretical framework, important implications can be derived for the present purpose. As compensatory scaffolding is assumed to modify the impact of detrimental structural brain alterations on cognitive ability, the strength of the concurrent association of brain structure and cognitive ability is expected to be weak in healthy aging individuals. Due to the variety of factors influencing the capacity of compensatory scaffolding, however, large interindividual and intraindividual variability in the strength of brain structure-cognition relations can be expected. In the following sections, we will review the current literature regarding the four research questions discussed in the introduction (see Figure 1), using the STAC-r model as a theoretical framework.

To introduce the terminology that will be used in this context, Figure 3 shows the possible cross-sectional and longitudinal cross-domain relations between brain structure and cognitive ability. As the literature so far has mainly investigated the direct pathway between brain structure and cognitive ability, scaffolding is not included in this figure. However, as the indirect pathway is considered as specifically relevant in the context of healthy aging, future theoretical development is encouraged to

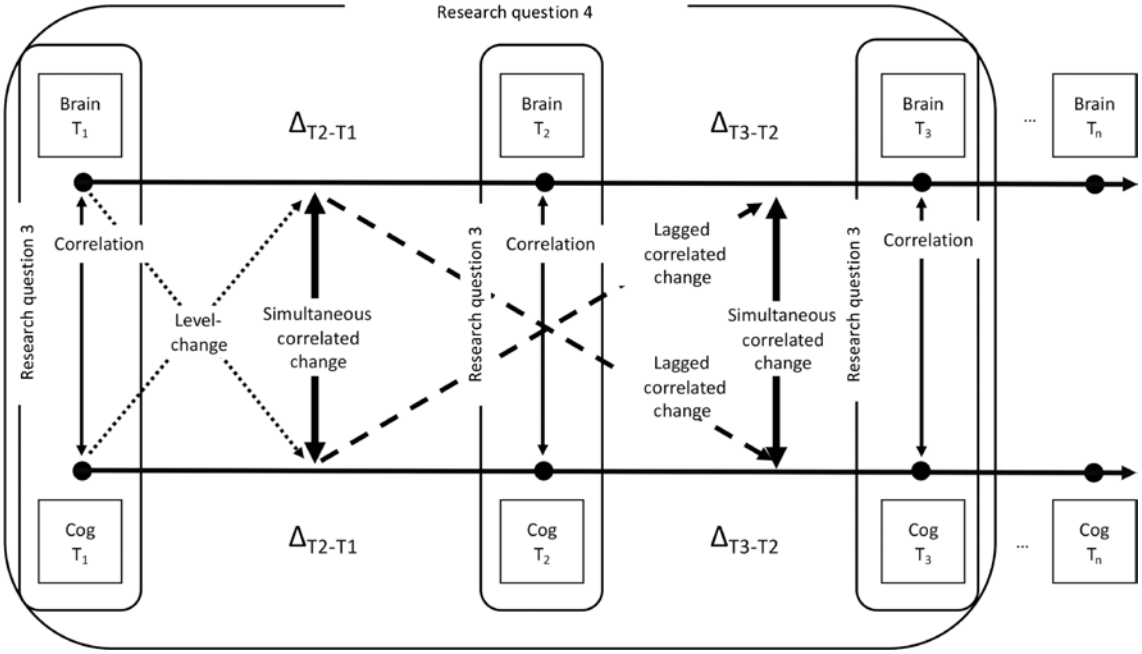

Figure 3: Potential cross-sectional (research question 3: correlation) and longitudinal (research question 4: level-change, simultaneous, and lagged correlated change) relations between brain structure (= Brain) and cognitive ability (= Cog). T = time/measurement occasion. $\Delta_{Tn+1\text{-}Tn}$ represents developmental change between two measurement occasions. Square shapes represent observed measures of a domain at a specific measurement occasion.

move into this direction (see section 'Theoretical limitations and the need for theory development' for an extended discussion).

In this figure, measures of brain structure properties and levels of cognitive ability (square shapes) and respective changes ($\Delta_{Tn+1-Tn)}$ in these variables are depicted between subsequent measurement occasions. Linking this figure to the multivariate research questions presented in the introduction, research question type 3 can be answered by looking at the cross-sectional correlation between specific structural brain features and levels of cognitive ability at a fixed measurement occasion. Regarding research question type 4 of longitudinal relations between brain structure and cognitive ability, several associations can be conceptually distinguished.

First, level-change associations might be observed, referring to any relationship between a cross-sectional measure – hereafter termed as level – of either structural brain properties or cognitive ability and longitudinal changes in the respective other domain. The term level is used in a statistical sense here to distinguish baseline assessments from longitudinal changes. For example, people with more intact structural brain features at baseline might be less likely to show age-related cognitive decline than people with lower levels of healthy brain tissue (in the sense of brain maintenance; see Nyberg et al., 2012). The reverse directionality is also plausible – higher levels of cognitive ability might protect from premature brain aging (in the sense of a cognitive reserve; see Stern, 2009; Barulli and Stern, 2013). In addition, both directions might be observed, such that bidirectional influences are at play. However, level-change associations are only quasi-longitudinal, as at least for one of the two variables, information on intraindividual change processes is lacking. It is impossible to know whether an individual with a seemingly more atrophied brain at baseline actually experienced intraindividual declines before the study period, as this longitudinal information is missing. As such, level-change associations only provide partial insights into change relations between the two domains.

Second, correlated change relationships between brain structure and cognitive ability might be observed. We refer to correlated change relationships to describe any temporal relationship between changes in both structural brain measures and cognitive measures. Furthermore, any correlated change relationship can either occur simultaneously or in a time-lagged fashion, such that changes in one variable over a certain time correlate with changes in the other variable at a later time-period. Simultaneous correlated changes between structural brain features and a specific cognitive ability carry information

about the association of changes that occur within the same observational time frame. Conceptually, such parallel cross-domain associations might either reflect directed relationships between the two domains that occur within the studied time frame, or the influence of a third variable on both developmental trajectories, in the sense of a common cause (Baltes and Lindenberger, 1997; Lindenberger and Ghisletta, 2009), which is, however, impossible to disentangle with this type of data.

The advantage of investigating lagged correlated changes is that they can yield more insights into the temporal dynamics of the association between changes in measures of brain structure and cognitive ability, allowing to investigate leading and lagging relationships (Grimm et al., 2012). Especially as scaffolding networks might be able to compensate for accumulating brain damage only until a certain threshold, such as posited by the theory of brain reserve (Satz et al., 2011), it is to be expected that detrimental changes in structural brain measures might significantly impact cognitive performance only after a certain time lag. As can be seen from looking at Figure 3, more than two measurement occasions are needed to study lagged correlated change associations.

Literature review

In the following sections, we will first summarize the existing literature on age-differences and changes (research questions type 1 and 2) in selected measures of brain structure and cognitive ability across the adult lifespan, with a focus on healthy old age, which – according to traditional conceptions – is defined as age 60 and older (e.g. Baltes and Smith, 2003). Second, cross-sectional and longitudinal associations between the two domains (research question type 3 and 4) will be discussed in the context of the STAC-r model (see Figure 1), presenting a systematic literature review of the available evidence on longitudinal correlated changes between brain structure and cognitive ability.

Age-differences and changes in brain structure in adult development

The brain undergoes substantial structural changes throughout the lifespan (Lockhart and DeCarli, 2014). With the advent of MRI, it has become possible to observe these changes *in vivo*. The MRI-derived parameters commonly used to describe aging can be roughly divided in measures of GM, white matter (WM) tissue, and cerebrospinal fluid (CSF), which together constitute whole brain volume (see Table 1 for detailed explanations). GM indices include volume-based measures, such as GM volume or density, and surface-based measures, such as cortical thickness or surface area. WM indices consist of volume-based measures, such as WM volume and WM hyperintensity (WMH) volume, and measures of WM anisotropy, and diffusivity, which yield information about the restrictedness of water diffusion in the WM tissue of the brain, and thus indirectly, about WM microstructural properties. Regarding whole brain volume and CSF, different measures are distinguished (see Table 1). As follows, we will give an overview over age-differences and changes in these structural brain indices in healthy aging.

GM

The GM volume gradually declines across the adult lifespan (Hedman et al., 2012). However, the onset, and the shape (e.g. linear, quadratic) of tissue loss are dependent on the brain region under study (Ziegler et al., 2012). Cross-sectional estimates of age-differences suggest that the implied decline of GM volume typically follows a last-in-first-out pattern, with anterior brain regions (e.g. prefrontal cortex) being the latest to mature and the first to show age-related deficits, and posterior regions that mature early in development (e.g. visual, auditory cortex) being less vulnerable to GM atrophy (Sowell et al., 2004). This pattern of structural brain differences across age is confirmed by a longitudinal study, with the exception of structures of the medial temporal regions (e.g. hippocampus, amygdala), which showed moderate reductions in children and young adults, but declined substantially in older adults (Tamnes et al., 2013). Regarding the shape of GM change across the adult lifespan, an age-heterogeneous longitudinal study (age range 23–87 years) reported nonlinear (implied) declines over age for GM volume in most areas of the cortex, with accelerating declines in temporal and occipital, and decelerating declines in prefrontal and anterior cingulate regions (Storsve et al., 2014). As cortical thickness and surface area are the two constituent measures of GM volume, those measures were also investigated. Interestingly, larger and nonlinear changes were found for cortical thickness, while surface area showed smaller and predominantly linear curves across most regions of the cortex. The authors interpreted this as evidence that cortical thickness contributes more strongly to GM volume changes in old age than surface area. Measuring middle-aged to older adults for up to five measurement occasions, Rast et al. (2017) reported nonlinear cortical thinning over 8 years in five lobar composites across the cortex, but

with decelerating changes, which were most pronounced for older adults in frontal, temporal, and cingulate cortices. With regard to subcortical structures, cross-sectional evidence across multiple sites shows a pattern of age-differences indicative of predominantly nonlinear decline trajectories across chronological age (e.g. hippocampus, caudate), but also of linear decline slopes for some structures (e.g. thalamus, accumbens; Walhovd et al., 2011; Ziegler et al., 2012; Fjell et al., 2013). These findings need to be interpreted with caution, however, as the automated reconstruction of subcortical structures is shown to be of variable reliability, depending, for example, on segmentation choices or the size of the structures (i.e. lower reliability for small structures, Morey et al., 2010).

WM

Cross-sectional estimates of age-differences (Westlye et al., 2010; Liu et al., 2016) and longitudinal estimates of change (e.g. Hedman et al., 2012) suggest that WM volume follows a nonlinear developmental pattern across the adult lifespan, with (implied) increases up to around age 50, and accelerated age-differences or declines thereafter. Similar to GM volume, the onset of WM volume decline is region-specific. The largest age-differences and declines were found in the frontal cortex, succeeded by the temporal (Bartzokis et al., 2001; Raz et al., 2005) and parietal cortices (Resnick et al., 2003), whereas occipital regions remain relatively spared (Raz et al., 2005).

Besides volumetric deficits, WM degradation in the course of healthy aging manifests itself also as age-differences and declines in microstructural properties of WM fiber tracts, and as an accumulation of WMH with increasing age (see Bennett and Madden, 2014 for a review). The former can be estimated with diffusion-weighted MRI (DW-MRI), a MRI method sensitive to the diffusion of water molecules in the brain (Jones, 2010). Compared to non-restricted diffusion of water molecules (i.e. in a glass of water), diffusion is highly directed in WM with the fibers acting as natural boundaries limiting the diffusion movement in certain directions. Indices derived from a tensor model fitted at each voxel reflect, for example, the degree of directedness of diffusion (fractional anisotropy, FA), or the mean rate of diffusion of a tissue (mean diffusivity, MD), independent of directionality (see Table 1; Beaulieu, 2002). Findings from cross-sectional studies show lower FA and higher MD in older as compared to younger adults (e.g. Cox et al., 2016), which is often interpreted as age-related deficits in the integrity of WM tracts. However, this interpretation is criticized (Jones, 2010), due to the lack of specificity of these measures with regard to their neurobiological foundation (see also Box 1). The few existing longitudinal studies demonstrate changes in WM microstructure with increasing age that are indicative of WM degradation, with prefrontal fiber systems being especially vulnerable to degradation (e.g. Barrick et al., 2010; Sullivan et al., 2010; Teipel et al., 2010). In addition, some longitudinal studies are suggestive of a similar last-in-first out pattern of change as seen in brain volumetric indices (e.g. Bender et al., 2016b; Storsve et al., 2016). However, exceptions to this trend is noted, with larger declines in posterior than frontal regions (Salat et al., 2005), and some support also exists for a superior-inferior gradient of WM aging (e.g. Sexton et al., 2014). Also, studies of WM microstructure in healthy aging increasingly differentiate between diffusion parallel (axial diffusivity; AD) and perpendicular (radial diffusivity; RD) to the main axis (see Table 1). Besides the changes in diffusion properties, increasing age is associated with a higher amount of WMHs, with cross-sectional estimates indicating a linear trajectory (e.g. Birdsill et al., 2014). WMHs can be detected and extracted from T2-weighted MRI images (particularly from pulse sequences; Fluid-Attenuated Inversion Recovery) using manual or (semi)-automated approaches (see Wardlaw et al., 2015) and are linked to pathological changes in vascular functions (Bennett and Madden, 2014). In a longitudinal study with healthy elderly, WMH volume increase was most pronounced in anterior regions of deep WMH (Sachdev et al., 2007).

CSF

Cross-sectional estimates suggest that CSF volume is larger in older compared to younger adults, and the CSF-filled ventricles appear to expand quadratically over the lifespan, with relative stability up to middle adulthood, and accelerated expansion thereafter (DeCarli et al., 2005; Carmichael et al., 2007; Fjell et al., 2013). Ventricular expansion is often used as a nonspecific proxy for global structural brain differences and changes and is shown a sensitive biomarker for AD progression (Madsen et al., 2013).

Whole brain volume

Especially in earlier publications, authors used variables reflecting combinations of tissue classes in order to make conclusions about differences and changes in whole brain volume. Depending on whether CSF is included in these measures or not one can dissociate total brain volume

(TBV) from intracranial volume (ICV); (for more details, see Table 1). Some authors also use a measure of TBV that is normalized for some estimate of overall head size (e.g. ICV), hereafter termed as normalized brain volume (NBV). NBV is a widely used index for brain atrophy, as overall head size as measured by ICV remains relatively stable across the lifespan and thus serves as a good measure to reduce between-subject differences with regard to maximum healthy brain size (e.g. Whitwell et al., 2001). A meta-analysis of 22 longitudinal studies implies a gradual decline in whole brain volume of 0.2% per year around the age of 35, and accelerated declines around age 60 (0.5% per year; Hedman et al., 2012). As whole brain volume includes both GM and WM, this estimated trajectory reflects a combination of the latter indices. Consequently, whole brain volume is a rather crude estimate of structural brain changes.

Summary

The mean trends reported above show that age exerts a stronger influence on brain structure in older than younger adults, which is reflected in the dominance of age-differences and decline in healthy old age. Interestingly, the average onset and the shape of age-related structural brain degradation varies depending on the type of tissue and the brain region under investigation, implying regional differences in structural brain aging that are shared among individuals. Roughly, GM atrophy onset is estimated at earlier ages, while WM remains relatively stable until old age. Moreover, a mean trend towards higher vulnerability of anterior, late developing regions as opposed to posterior, early developing regions is reported by several studies of WM and GM aging. A premise of STAC-r is that life course experiences of various kinds shape brain structure besides the mere influence of passing time. Assuming that the brain remains plastic up into higher ages, variability between healthy aging individuals with regard to the onset and shape of brain structure change can be expected (i.e. including maintenance and growth as potential trajectories). The finding of predominantly nonlinear average trajectories for many brain structures (e.g. cortical thickness, subcortical GM, WM) lends some support to this hypothesis.

Age-differences and changes in cognitive ability in adult development

To date, research in the field of cognitive aging has provided ample support for the multidirectional development of cognitive abilities across the lifespan (Baltes, 1987). A prominent theory is the division of cognitive abilities into fluid and crystallized intelligence proposed half a century ago by Cattell (1963). Fluid intelligence describes the ability of reasoning and novel problem solving, and is often discussed as a higher-order factor of fluid cognitive abilities. Crystallized intelligence refers to well-practiced abilities and knowledge accumulated across the lifespan. In general, fluid processing mechanisms, (e.g. perceptual speed, working memory) gradually decline with age, whereas well-practiced crystallized abilities, such as vocabulary, knowledge, and autobiographical memory show patterns of increase, and stability well into older adulthood (Hedden and Gabrieli, 2004; Salthouse, 2010). However, recent evidence also suggests that the broad division into fluid and crystallized abilities falls somewhat short on the complexity and heterogeneity of developmental patterns of cognitive domains (Hartshorne and Germine, 2015). Therefore, we will summarize the evidence of age-differences and changes separately for a selected set of specific cognitive abilities relevant in the scope of the current review.

Memory

Roughly, memory can be divided in retrospective and prospective memory (Baddeley et al., 2009). First, retrospective memory refers to the memory for information acquired in the past and can be further distinguished into short-term and long-term memory. Regarding the memory for short-term information, age-differences and changes are mainly observed for working memory, which is discussed separately below. Long-term memory can be divided into explicit memory (involving episodic and semantic memory) and implicit memory (Schacter, 1987). Episodic memory refers to the recollection of events experienced in the past and is especially vulnerable to aging (Tulving, 1972). Tasks testing episodic memory require participants to memorize a set of stimuli (e.g. words) and later, to recall them (free recall) or decide whether they have encountered the stimulus before (recognition). Older adults show more difficulties with recall than recognition of previously memorized information (Craik and McDowd, 1987). Whereas age-related differences suggest an early onset of episodic memory decline in young adulthood (Salthouse, 2003), longitudinal evidence does not support age-related declines before age 60 (Rönnlund et al., 2005; Schaie, 2005). Semantic memory describes the memory for factual knowledge (e.g. vocabulary) and comprehension (Tulving, 1972). Due to life-long accumulation of

knowledge, longitudinal evidence suggests that semantic memory increases or remains stable at least until age 55 (Rönnlund et al., 2005), with late-life declines smaller in size than for episodic memory. Longitudinal findings from the Berlin Aging Study demonstrated stability in verbal knowledge even up to the age of 90 (Park et al., 2002). Implicit memory refers to the unconscious influence of previously acquired information on present performance. A meta-analysis of the cross-sectional literature showed that the deficit that older participants show in implicit memory performance as compared to younger participants is much smaller compared to the age-differences in episodic memory (Light et al., 2000). Furthermore, a longitudinal study did not show implicit memory declines in healthy older participants over 3 years, suggesting relative stability of implicit memory in healthy aging (Fleischman et al., 2004).

Second, prospective memory is needed to remember and enact a previously made plan in the future. In the laboratory, prospective memory is tested by abstract tasks, for example, remembering to ask for a pen at 9 AM (time-based) or whenever the investigator mentions a code word (event-based). Prospective memory is highly relevant in everyday situations, for example, when one needs to remember taking medication every morning. Evidence from cross-sectional studies shows that older adults perform worse in laboratory-based prospective memory tasks as compared to young adults (Henry et al., 2004; Kliegel et al., 2016), especially in strategically more demanding tasks (i.e. specified task order, see Ihle et al., 2013; nonfocal task cues, see Kliegel et al., 2008). However, outside of the laboratory, older adults show similar or even superior performance in naturalistic prospective memory tasks. This paradoxical finding has been related to differences between the two settings, such as higher motivation, more flexibility for self-management, and less engagement in distracting activities in older adults when tested in their everyday life (Schnitzspahn et al., 2011).

Executive functions

Executive functions (EF) are higher-order abilities needed to pursue complex tasks of planning, organization, and goal-directed behavior (Burgess, 1997). Besides a general component, EF are composed of a set of basic abilities involving the inhibition of prepotent responses, shifting between mental representations, and updating of representations held in working memory (Miyake et al., 2000). As updating tasks tap into working memory ability (Schmiedek et al., 2009), we will discuss this line of research in the section on working memory. Overall, absolute age-differences can be found in tasks requiring executive control compared to tasks involving only little control demands (Verhaeghen, 2011). Regarding inhibition and shifting, cross-sectional (Healey et al., 2014; Treitz et al., 2007) and longitudinal studies have shown age-differences and declines (Goh et al., 2012; Van der Elst et al., 2013; Adólfsdóttir et al., 2017). However, a meta-analysis could not support age-related deficits specific to inhibition, as compared to a baseline condition with the inhibitory control aspect removed (Verhaeghen, 2011), suggesting age-differences in more basal processes. For shifting, specific age-related deficits of older adults were only found for global shifting (Wasylyshyn et al., 2011), a measure for monitoring ability in dual-task as opposed to single-task situations.

Attention

Attention involves the capacity-limited ability to direct one's focus to selected stimuli in the environment (Jäncke, 2017). Important aspects of attention are sustained attention (maintain focus and vigilance over a prolonged time), selective attention (focus on one stimulus while ignoring irrelevant information), and divided attention (focus on two stimuli at the same time) (Drag and Bieliauskas, 2010). While older adults typically do not differ from younger adults regarding sustained attention (e.g. Berardi et al., 2001), selective and divided attention seem to be more sensitive to aging. First, selective attention requires the inhibition of distracting information. Similar to the literature on inhibition, negative age-differences are reported for selective attention (e.g. Plude and Hoyer, 1986; Brink and McDowd, 1999). However, recent evidence suggests that these deficits are limited to specific modalities (e.g. auditory task with visual distraction), and it is yet unclear whether this modality-dependency can be explained with age-differences in inhibition (Van Gerven and Guerreiro, 2016). Second, divided attention is assessed with the performance in dual-task situations, and task-switching experiments that are also used to assess the shifting factor of EF (Verhaeghen and Cerella, 2002). The results from a meta-analysis indicate negative age-differences in dual-task situations for older adults beyond age-related slowing (Verhaeghen et al., 2003). Regarding the performance in task-switching experiments, age-differences are restricted to a global disadvantage of managing dual-task situations (compare to results of EF shifting).

Working memory

Working memory describes a limited-capacity system that is involved in the simultaneous storage, and processing or updating of information (Cowan, 1995; Baddeley, 1998; Oberauer, 2009). It is typically tested with tasks that require participants to hold a certain number of stimuli in working memory, while simultaneously updating or manipulating information. For example, the reading span task requires participants to read sentences, answer related questions, and then recall the last word of each sentence (Daneman and Carpenter, 1980). Working memory is discussed to be highly related to fluid intelligence, however, it still has independent explanatory value (Salthouse and Pink, 2008). Age-related differences are reported as early as from young adulthood (20 years), following a linear trend (Brockmole and Logie, 2013). Longitudinal evidence has suggested declines in working memory capacity in middle-aged to older adults over the time span of 3 years (Hultsch et al., 1992).

Processing speed

Processing speed refers to the speed with which information is processed and can be divided in measures of psychomotor speed and perceptual speed. While psychomotor speed refers to the speeded performance in very basic motor task (e.g. finger tapping), perceptual speed tasks additionally include varying amounts of executive control (e.g. copying symbols or substituting digits with symbols) (Cepeda et al., 2013). As the distinction between these two types of speed measure is often neglected in the literature, we will hereafter refer to processing speed as a broader construct. Processing speed is a core component of higher-order cognitive abilities and thus suggested to drive age-related changes in other fluid cognitive domains (Robitaille et al., 2013). The rationale is that if simple processing steps take up more time due to age-related slowing, the remaining time for more complex operations is consequently limited. Furthermore, the slowing of basic mental operations might lead to the loss of information (e.g. through mechanisms of decay) by the time it is required for higher-order operations (see Salthouse, 1996). Cross-sectional studies indicate that age-deficits in processing speed can already be found in early adulthood, implying an early onset of cognitive slowing (Salthouse, 2010). While longitudinal findings of the Seattle Longitudinal study suggest a much later onset of decline around the age 60, processing speed is still one of the earliest fluid cognitive abilities to decline (Schaie, 2005).

Summary

On average, age-differences and declines with old age are found for episodic and prospective memory, EF (although driven partly by basal processes), selective and divided attention, working memory, and processing speed, whereas implicit, semantic memory, and sustained attention show relative stability into older age. From the perspective of STAC-r, such multidirectional changes might reflect differences in the extent to which these cognitive domains rely on cognitive processes which can be compensated by strategy use or experience when biological resources are not sufficient anymore. For example, vocabulary knowledge (semantic memory) might be more strongly influenced by experience (i.e. education, frequent social interactions or reading the newspaper) and thus more adept to compensatory maintenance than the ability to complete a task as fast as possible (processing speed). The latter might thus rely more on a youthful brain structure and function. Regarding the onset of age-related differences and declines, cross-sectional studies typically estimate an earlier onset than longitudinal studies. This discrepancy can be explained in part by methodological limitations inherent to the respective study designs (e.g. practice effects in longitudinal studies, Salthouse, 2014; cohort-effects in cross-sectional studies, Schaie, 2005). As scaffolding is a regulatory process that occurs within individuals over time, only longitudinal studies can directly capture this process. This might also explain why longitudinal studies report stability of cognitive ability into much higher ages. Specifically, accelerated cognitive declines (e.g. of episodic memory, processing speed) observed in longitudinal studies could reflect a turning point when compensatory mechanisms start to lose their functionality (e.g. due to degradation of the frontal cortex).

Associations between changes in brain structure and cognitive ability in healthy aging

So far, the relation between brain structure and cognitive ability was mainly investigated by cross-sectional studies. This literature has previously been reviewed elsewhere (see Kaup et al., 2011; Salthouse, 2011), which is why we refer to these works for an in-depth discussion. In brief, cross-sectional studies typically correlate a measure of brain structure and a measure of cognitive ability while controlling for age (see research question 3). Several of

these studies show a trend toward a positive brain-cognition correlation (but see Salat et al., 2002; Van Petten et al., 2004), suggesting that people with larger brain volumes, a thicker cortex, or better WM health (i.e. less WMH load, more intact WM microstructure) on average perform better in a variety of cognitive tasks, independent of their age. However, the results remain largely inconclusive with regard to the association of specific brain regions and single cognitive domains, given vast methodological differences between studies in terms of brain structure proxies, the selection of cognitive tasks, or varying sample sizes. Furthermore, the focus has often not been on (healthy) aging, but more generally on brain-cognition relations. Although other studies have tested more specific hypotheses with regards to aging (e.g. investigating brain structure as a mediator of age-effects on cognitive performance), these studies often fall short in contrasting their findings with alternate theories (Salthouse, 2011). Moreover, mediational analyses are not sufficient to disentangle directional relationships (Hofer et al., 2006; Lindenberger et al., 2011). As is discussed already in the introduction, longitudinal studies are necessary to draw inferences about the interrelation of change trajectories in brain structure and cognition over time. Therefore, we will comprehensively review this literature in the following sections.

Literature search and inclusion criteria

To review the longitudinal literature on brain structure-cognition relations in healthy aging, we searched Pubmed using the terms ('brain structure' OR 'brain volume' OR 'white matter' OR 'grey matter' OR 'gray matter' OR 'cerebrospinal fluid' OR 'CSF' OR 'ventricle' OR 'ventricular' OR 'cortex' OR 'cortical' OR 'diffusion' OR 'hippocampus' OR 'hippocampal') AND ('cognition' OR 'cognitive' OR 'speed' OR 'memory' OR 'executive functions' OR 'EF' OR 'intelligence' OR 'attention') AND ('old age' OR 'aging' OR 'ageing' OR 'elderly') AND ('change' OR 'changes' OR 'trajectory' OR 'trajectories') AND ('longitudinal' OR 'over time' OR 'follow-up'). Figure 4 shows a flow-diagram of our search procedure, adapted from the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines (Moher et al., 2009). The search was conducted on 13 February 2019, and yielded 1180 results. In addition, we identified four articles from other sources (i.e. reference lists of the screened articles) that seemed relevant, resulting in 1184 articles. We screened the titles and abstracts and included only those articles that:

(1) studied older participants over 60 years of age (articles spanning a wider age range were also included if part of the sample was older than 60 years),
(2) studied cognitively healthy participants (full sample or a subset),
(3) reported two or more measurement occasions of both a measure of brain structure and cognitive ability,
(4) reported information regarding the statistical analysis of correlated changes between brain structure and cognitive ability,
(5) and were prospective observation studies.

We retained 268 articles for closer examination that fulfilled these criteria or that were not providing sufficiently clear information in the title/abstract to be excluded. In the next step, we screened the full text of these articles for eligibility and excluded 238 articles (for reasons see Figure 4), resulting in 31 relevant articles, which will be reviewed in the following sections. The results for GM (see Table 2), WM (see Table 3), and whole brain volume and CSF (see Table 4) are reported separately (see column brain-cognition relations). Whole brain volume and CSF are reported in the same table, as both measures reflect direct or indirect (in the case of CSF) estimates of global changes in brain tissue. If an article included results for multiple indicators (e.g. both GM and WM), the results for these indicators are listed in separate tables, thus leading to overlapping samples between tables. In addition, if an article reported results for both cognitively healthy and pathological participants, only the results for the healthy subgroup are reported. If some initially healthy participants converted to mild cognitive impairment or dementia during the time of the study, we only included the respective article if it reported results without the cognitively impaired participants. Due to a lack of consensus in the field on a definition of healthy aging in general and cut-offs for cognitive health in specific, we decided to leave the decision on exclusion criteria for healthy cognitive aging up to the researchers (see column 'Healthy cognitive aging' in Tables 2–4), while, however, discussing this variability as a methodological limitation (see section 'Methodological limitations'). In the scope of this literature review, we limited our search to brain structure measures that are commonly reported in the literature (see Table 1). For the cognitive ability measures, we did not include measures from screening instruments for the detection of cognitive impairment [e.g. mini mental state examination (MMSE)], unless they were part of a composite score with

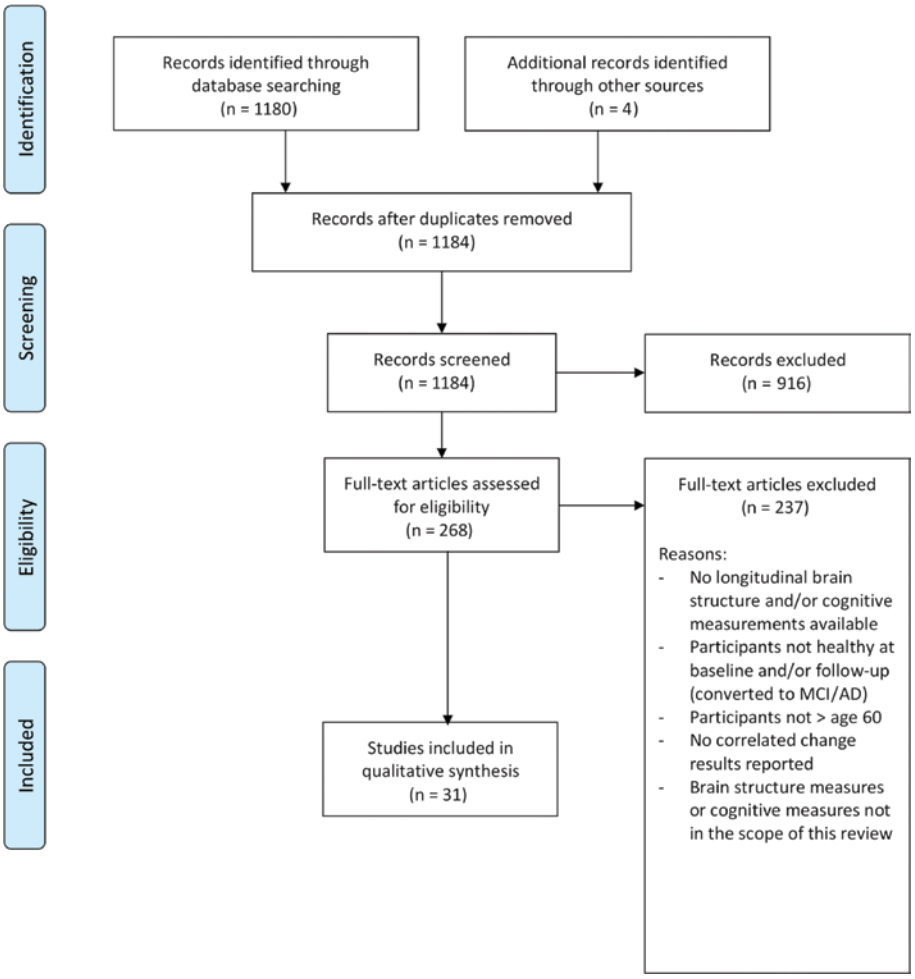

Figure 4: Preferred reporting items for systematic reviews and meta-analyses flow-chart of the literature search procedure.

other neuropsychological and/or psychometric cognitive tasks. In addition, in Tables 2–4, we list only structural brain and cognitive measures that were also considered for the analysis of brain-cognition relations. All reviewed studies used MRI to measure brain structure. Only one of the reviewed studies reported lagged correlated changes. For reasons of simplicity, we thus refer to correlated changes when reporting results on simultaneous correlated changes and will specifically highlight the discussion of lagged correlated changes.

Due to the similarities in study design and measures investigated, we also give an overview of the literature on cognitive training in healthy older adults (see Box 2). In addition to delivering a cognitive intervention, these studies also included cognitive and neuroimaging assessments at least at pretest and post-test. We limit our discussion to cognitive training studies that included a control group (active or passive) and administered a substantive amount of training sessions (at least 10).

Results

The results discussed in the following sections can be interpreted as follows: (1) positive level-change correlations suggest that higher levels of brain structure (or cognitive ability) are associated with a more positive change (i.e. less decline) in cognitive ability (or brain structure) and vice versa, (2) positive change-change correlations suggest that a more positive change (i.e. increase or less decline) in brain structure (or cognitive ability) is associated with a more positive change in cognitive ability (or brain structure) and vice versa. To avoid misinterpretations, the results are presented such that higher values in cognitive tasks reflect better performance.

GM and cognitive ability

In total, 18 studies investigated longitudinal associations between measures of GM structure and cognitive ability

Table 2: Overview of prospective observation studies investigating level change or correlated change associations of GM and cognitive abilities.

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, −=negative)
GM										
Cohen et al. (2001)	25 female (~50–70[a])	2	2	None[b]	Correlation (change $ratio_{brain/cog}$)	None	GM volume	Hippocampus	**Episodic memory** Buschke-fuld test (free and delayed recall, consistency) Wechsler memory scale-revised (verbal, visual, general, attention, and delay index) **Verbal fluency** Letter, category, supermarket items **Visuospatial processing** Rey figure test (copy, recall). Block design test	**Level-change** Not reported **Correlated change** No effect
Raz et al. (2007)	46 (45–77)	2	5–6	MMSE ≥ 26, BIMC ≥ 85% correct, CES-D ≤ 15	Repeated-measures regression analysis[c,d]	Age	GM volume, adjusted for ICV	Average cortical volume[e]	**Fluid intelligence** CFIT (subtest 3)	**Level-change** Not reported **Correlated change** No effect
Raz et al. (2008)	87 (20–77)	2	5–6	See Raz et al. (2007)	Bivariate LCS (latent change $score_{brain/cog}$)	Age Sex VR	GM volume, adjusted for ICV	LPFC OFC Inferior temporal gyrus Fusiform gyrus VC Hippocampus Entorhinal cortex Caudate Cerebellum	**Fluid intelligence** CFIT (subtests 1 and 3)	**Level-change:** Fluid intelligence T1/T2 + Δentorhinal cortex[f] **Correlated change** Not reported[g]
Leow et al. (2009), ADNI	40 normal controls (75.27 ± 5.33)	2	1	MMSE ≥ 28, CDR = 0	Regression analysis (structural change map_{brain})[h]	None	GM volume	Temporal lobe	**Episodic memory** Logical memory (immediate, delayed recall) AVLT (delayed recognition)	**Level-change** No effect **Correlated change** No effect
Persson et al. (2012), Betula	26/15 with VR (55–79)	2	$6_{brain}/10_{cog}$	MMSE ≥ 24	Regression analysis (difference $score_{brain/cog}$)	Age APOE-ε4	GM volume, adjusted for body size	Hippocampus	**Episodic memory** Composite of five tasks (immediate/ delayed recall of commands with and without enactment, auditory word list)	**Level-change** ΔEpisodic memory + right hippocampus T2 **Correlated change** ΔRight hippocampus + Δepisodic memory
Fjell et al. (2014), ADNI	94 normal controls (55–90)[i]	2	1	MMSE ≥ 24 CDR = 0 Logical Memory Test ≥ 3–9	Correlation (change $ratio_{brain/cog}$)	None[j]	Cortical thickness	Entorhinal cortex	**Episodic memory** AVLT (learning, delayed recognition)	**Level-change** ΔEntorhinal + AVLT delayed T2 **Correlated change** ΔEntorhinal + ΔAVLT learning

Table 2 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, –=negative)
Daugherty et al. (2015)	125 (19–77)	2	2	MMSE ≥ 26 CES-D < 16	Bivariate $LCS_{brain/cog}$	Age Iron content (baseline and change)	GM volume, adjusted for ICV	Hippocampus Caudate Putamen	**Episodic memory** Logical memory (immediate, delayed recall) **Working memory** Listening span Verbal N-back	**Level-change** Not reported **Correlated change** No effect
Mak et al. (2015b)	33 normal controls (76.7 ± 5.3)	2	1	No history of dementia or CAMCOG score < 80	Correlation (change $ratio_{brain}$)[h]	Age Sex	Cortical thickness	Whole brain	**Episodic memory** Brief BVMT **Global cognition** CAMCOG	**Level-change** Not reported **Correlated change** No effect
Ritchie et al. (2015b), Lothian Birth Cohort	657 (Mean age 73)	2	3	MMSE ≥ 24	Multivariate $LCS^{k}_{brain/cog}$	Age Sex	GM volume	-	**Fluid intelligence** Matrix reasoning Block design Digit span backward Letter-number sequencing **Episodic memory** Logical memory (immediate, delayed recall) Verbal paired associates, Spatial span (forward, backward) **Processing speed** Digit-symbol Symbol search Simple-/four-choice reaction time task Inspection time task	**Level-change** No effect **Correlated change** ΔGM volume + Δspeed, Δfluid intelligence
Fjell et al. (2016), Cognition and plasticity through the lifespan	56 (63–86)	2	3.1	MMSE ≥ 26, BDI ≤ 16 IQ ≥ 85	Stepwise regression analysis (difference $score_{brain}$/ change $ratio_{cog}$)	Age Sex Head motion time interval Δfunctional connectivity (caudate, hippocampus)	GM volume	Hippocampus Caudate	**Episodic memory** CVLT immediate recall	**Level-change** Not reported **Correlated change** No effects

Table 2 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, –=negative)
Möller et al. (2016)	34 normal controls (60.6 ± 6)	2	2	None[b-]	Correlation (change $ratio_{brain/cog}$)	None	Cortical thickness	Left and right lobar compartments (frontal, parietal, temporal, occipital), Cingulate Insula	**Episodic memory** AVLT (immediate, delayed recall, recognition) Visual association test A **EF** Digit span backwards TMT B Letter fluency Stroop color-word **Attention** TMT A Digit span forward Letter digit substitution **Visuospatial functioning** Number location **Language** Picture naming Category fluency	**Level-change** Left and right occipital T1 + Δepisodic memory Right occipital T1 + Δlanguage **Correlated change** ΔRight cingulate + Δepisodic memory ΔRight occipital – ΔEF
Persson et al. (2016)	167 (19–79)	2	2	MMSE ≥ 26, CES-D ≤ 15	Bivariate $LCS_{brain/cog}$	Age SES VR APOE-ε4	GM volume, adjusted for ICV	LPFC, Hippocampus Parahippocampal gyrus Cerebellar hemispheres VC	**Fluid intelligence** CFIT (four subtests) **Episodic memory** Name-picture associative memory, story recall (both immediate/ delayed recall) **Vocabulary** two vocabulary lists (selecting synonyms)	**Level-change** Hippocampus T1, parahippocampal gyrus T1, cerebellar hemispheres T1 + Δfluid intelligence Episodic memory T1 + ΔLPFC **Correlated Change** No effect[l]
Anblagan et al. (2018), Lothian Birth Cohort	655 (72.5 ± 0.7)	2	3	MMSE ≥ 24	Multivariate $LCS_{brain/cog}$	Age Sex	GM volume FA Diffusivity	Hippocampus (volume unadjusted, adjusted for ICV, FA, MD, longitudinal relaxation time T1)	**Working memory** Spatial span (forward, backward) Digit span backward Letter-number sequencing **Verbal memory** Logical memory (immediate, delayed recall) Verbal paired associates, **Processing speed** Digit-symbol substitution Symbol search Four-choice reaction time task Inspection time task	**Level-change** MD T1 – Δverbal memory, Δspeed, Δworking memory **Correlated change** ΔMD + Δworking memory

Table 2 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, −=negative)
Gorbach et al. (2017), Betula	264 (55–80)	$2_{brain}/3–4_{cog}$	$4_{brain}/15_{cog}$	None[b]	Correlation (change $ratio_{brain}$/ regression slope $estimates_{cog}$)	Age VR	GM volume	LPFC (two subregions) Rostral anterior cingulate cortex Hippocampus Parahippocampal gyrus Inferior temporal cortex Superior parietal cortex Precuneus Putamen Lateral occipital cortex	**Fluid intelligence** Block design **Episodic memory** Composite of five tasks (immediate/ delayed recall of commands with and without enactment, auditory word list) **Processing speed** Letter-digit substitution Letter comparison Figure comparison **Verbal fluency** Letter fluency task (three conditions)	**Level-change** Not reported **Correlated change** ΔHippocampus + Δepisodic memory
Hohman et al. (2017), ADNI	379 normal controls (73.9 ± 5.9)	>2[m]	4	MMSE ≥ 24 CDR = 0 Logical memory test ≥ 3–9	MLM analysis[n]	Age CSF biomarkers (Tau, Ab-42)	GM volume, adjusted for ICV	Default mode network ROIs (DMN) Frontal cortex (FC) Medial temporal cortex (MTL) Sensorimotor areas (SENS)	**Episodic memory** AVLT (learning, delayed recognition), MMSE Logical memory Alzheimer's disease assessment scale (cognitive subscale) **EF** TMT (A and B) Digit span backward Digit symbol Category fluency (animals/vegetables) Clock drawing	**Level-change** Hippocampus T1 + ΔEF, Δepisodic memory **Correlated Change** No effect[o]
Leong et al. (2017), Singapore Longitudinal Aging Brain Study	111 (56.1–83.1)	2–4	1.5–2	MMSE ≥ 26	Correlation (MLM slope $estimates_{brain/cog}$)	Age Sex Education VR BMI ICV	GM volume	Total GM Lobar GM (frontal, parietal, temporal, occipital) Hippocampus	**Global cognition** Composite of all abilities below **EF** Category fluency (three categories) Design fluency TMT B **Processing speed** TMT A Symbol search Symbol-digit modalities **Visuospatial memory** Visual paired associates test (immediate, delayed recall) **Verbal memory** AVLT (immediate, delayed recall) **Attention** Digit span Spatial span (forward and backward)	**Level-change** Not reported **Correlated change** ΔTotal GM, Δlobar GM (frontal, parietal, temporal), Δhippocampus) + Δglobal cognition ΔOccipital GM − Δglobal cognition ΔHippocampus + Δverbal memory, ΔEF

Table 2 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, –=negative)
Sala-Llonch et al. (2017)	89 (64.74–89.79)	2	2	None[b]	Regression analysis (change $ratio_{brain/cog}$)	Age Sex	Cortical thickness	-	**Episodic memory** Word list memory task **Verbal Fluency** Animal Naming Test (ANT) **EF** TMT B	**Level-change** Not reported **Correlated change** ΔCortical thickness (right hemispheric clusters covering supramarginal, postcentral and inferior parietal cortices) + ΔANT
Yuan et al. (2018)	73 (51–85)	1–4	~1–5 (max 8 years in the study)	MMSE ≥ 26, CES-D < 16	MLM analysis	Age	Cortical thickness GM volume	Total cortical thickness[p] Cortical GM volume[p]	**Fluid intelligence** CFIT (four subtests) **Crystallized intelligence** Vocabulary Tests (two versions)	**Level-change** Fluid intelligence T1/T2 + Δcortical GM volume Cristallized intelligence T1/T2 – Δcortical GM volume, Δtotal cortical thickness **Correlated change** Not reported[g]

brain, brain measurements; cog, cognitive measurements; ADNI, Alzheimer's disease neuroimaging initiative; MMSE, mini-mental-state-examination; BIMC, blessed information-memory-concentration test; CES-D, Center for Epidemiologic Studies Depression Scale; CDR, clinical dementia rating; LCS, latent change score model; MLM, multilevel model; change ratio, a measure of change proportionalized on the baseline level and in some cases also on the time interval (for exact formula consult original publications); difference score, value at T1 – value at T2; ICV, intracranial volume; SES, socioeconomic status; VR, vascular risk (for details consult original publications); APOE-ε4, apolipoprotein E allele 4; BMI, body mass index; ROIs, regions of interest; LPFC, lateral prefrontal cortex; OFC, orbitofrontal cortex; VC, primary visual cortex; TMT, trail making test; AVLT, Rey auditory verbal learning test; CFIT, Catell Culture Fair Test; CVLT, California verbal learning test; EF, executive functions; T, measurement occasion (e.g. T1 = measurement occasion 1); Δ, change; GM, gray matter.

[a]Age reported separately for two subgroups of APOE-ε4 negative (n = 9, M age = 60.6 ± 10.2 years) and APOE-ε4 positive (n = 16, M age = 55.1 ± 6.0 years) participants. [b]No details provided on screening instruments and/or cut-offs used (for description of healthy aging criteria consult original publication). [c]Total WMH volume was included as a predictor in the same model (cf. Table 3). [d]Change measure was not clearly described. [e]Unclear, which cortical regions were included in this variable. [f]No longer significant after sex and VR were included. [g]A lack of interindividual differences in intraindividual change in the cognitive variables precluded correlated change analyses. [h]Change measure for the cognitive variables not clearly describe. [i]Age range of the full sample, as age range of the normal subsample was not reported. [j]Age had no significant relation with any of the brain/cognitive performance measures, therefore it was not included as a covariate. [k]All cognitive variables were included in the same statistical model. [l]Of the cognitive measures, only intelligence showed significant interindividual differences in intraindividual change and was used for the correlated change analyses. [m]Change was modeled over age (instead of time; consult original publication for details). [n]Number of measurement occasions not clearly indicated. [o]The correlated change analysis was conducted using a latent measure of global brain changes that comprised the single composites. [p]Measures of interest are modeled across 19 regions, using region as crossed random effect (for details consult original publication).

Table 3: Overview of prospective observation studies investigating level-change or correlated change associations of WM and cognitive abilities.

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+ = positive, – = negative)
WM										
Schmidt et al. (1999), Austrian Stroke Prevention Study	273 (60 ± 6.1)	2	3	None[a]	Repeated measures regression analysis (difference $score_{brain}$)	Age Education (years)	WMH volume	-	**Episodic memory** Lern- und Gedächtnistest (immediate, delayed recall) **Reasoning** WCST **Attention and speed** Alters-Konzentrations-Test TMT B Digit span Complex reaction time test	**Level-change** Not reported **Correlated change** No effect
Schmidt et al. (2005), Austrian Stroke Prevention Study	329 (50–75)	2–3	3	See Schmidt et al. (1999)	Repeated measures regression analysis (difference $score_{brain}$)	Age Sex Education (years) VR Diabetes Cardiac disease Smoking Cholesterol	WMH volume	-	**Episodic memory** Lern- und Gedächtnistest (immediate, delayed recall) **Reasoning** WCST **Attention and speed** Alters-Konzentrations-Test TMT B Digit span Complex reaction time test	**Level-change** Not reported **Correlated change** ΔGlobal WMH volume – Δepisodic memory, Δreasoning[b]
Raz et al. (2007)	46 (45–77)	2	5–6	MMSE ≥ 26, BIMC ≥ 85% correct, CES-D ≤ 15	Repeated-measures regression analysis[c,d]	Age	WMH volume	Lobar WMH (frontal, parietal, temporal, occipital) Global WMH	**Fluid intelligence** CFIT (subtest 3)	**Level-change** Not reported **Correlated change** ΔGlobal WMH – Δfluid intelligence
Raz et al. (2008)	87 (20–77)	2	5–6	See Raz et al. (2007)	Bivariate $LCS_{brain/cog}$	Age Sex VR	WM volume, adjusted for ICV	Prefrontal WM Inferior parietal WM	**Fluid intelligence** CFIT (subtests 1 and 3)	**Level-change** No effect **Correlated change** Not reported[e]
Silbert et al. (2008), Oregon Brain Aging Study	104 (65–99)	3–13	0.9–5.5 (max 18 years in the study)	MMSE ≥ 24, CDR = 0	Regression analysis (slope $coefficient_{brain/cog}$)	Age ICV	WMH volume	Total WMH Periventricular WMH Subcortical WMH	**Episodic memory** Logical memory (delayed story recall)	**Level-change** No effect **Correlated change** ΔSubcortical WMH – Δepisodic memory

Table 3 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+ = positive, – = negative)
Charlton et al. (2010), GENIE	106 (50–90)	2	2	None[a]	Correlation (change $ratio_{brain/cog}$)	Age	FA Diffusivity WMH volume	Total FA, MD distribution parameters (peak height frequency and intensity, skewness, kurtosis, median) Total WMH	**Working memory** Digit span backwards Letter-number sequencing	**Level-change** No effect **Correlated Change** ΔMD (peak height frequency, skewness, kurtosis) –Δworking memory
Lövdén et al. (2014), SNAC-K	563/70 with brain data (81–103)	2	3	None[a]	Bivariate $LCS_{brain/cog}$	Age Sex Education -(years)	FA Diffusivity	FA of: Cingulum cingulate gyrus Superior longitudinal fasciculus FA and MD of: Corticospinal tract Forceps major Forceps minor	**Processing speed** Digit cancellation, Pattern comparison **Episodic memory** Noun recall/recognition **Semantic memory** Vocabulary General knowledge **Verbal fluency** Letter and category fluency (2 tasks each)	**Level-change** Not reported **Correlated change** ΔFA corticospinal tract + Δprocessing peed ΔMD corticospinal tract – Δprocessing speed
Ritchie et al. (2015b), Lothian Birth Cohort	657 (M age 73)	2	3	MMSE ≥ 24	Multivariate $LCS^{f}_{brain/cog}$	Age Sex	WM volume WMH volume	Normal appearing WM volume (NAWM) Total WMH (WMH)	**Fluid intelligence** Matrix reasoning Block design Digit span backward Letter-number sequencing **Episodic memory** Logical memory (immediate, delayed recall) Verbal paired associates, Spatial span (forward, backward) **Processing speed** Digit-symbol substitution Symbol search Simple-/four-choice reaction time task Inspection time task	**Level-change** Episodic memory T1, processing speed T1 + ΔNAWM Processing speed T1 – ΔWMH **Correlated change** ΔWMH – Δfluid intelligence, Δepisodic memory, Δprocessing speed

Table 3 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+ = positive, − = negative)
Ritchie et al. (2015a), Lothian birth cohort	731 (72.68 ± 0.72)	2	3	MMSE ≥ 24	Multivariate $LCS^{I}_{brain/cog}$	Age Sex	FA	Corpus callosum (genu, splenium) Rostral cingulum cingulate gyri Arcuate fasciculus Uncinate fasciculus Inferior longitudinal fasciculus Anterior thalamic radiation	see Ritchie et al. (2015b)	**Level-change** FA T1 + ΔprocessingSpeed **Correlated change** ΔFA + Δfluid intelligence
Bender et al. (2016)	96 (19–78)	2	2	MMSE ≥ 26, CES-D ≤ 15	Bivariate $LCS_{brain/cog}$	Age VR	FA Diffusivity	FA and RD of: Dorsal cingulum bundle Anterior limb of internal capsule Forceps minor Inferior frontal-occipital fasciculus Uncinate fasciculus	**Fluid intelligence** CFIT (subtest 3B) Letter sets test **Episodic memory** Picture-name associations Word pair recognition	**Level-change** No effects **Correlated Change** ΔFA – Δ Episodic memory ΔRD + Δ Episodic memory
Fjell et al. (2016), Cognition and plasticity through the lifespan	56 (63–86)	2	3.1	MMSE ≥ 26, BDI ≤ 16 IQ ≥ 85	Stepwise regression analysis (difference $score_{brain}$/change $ratio_{cog}$)	Age Sex Head motion time interval Functional connectivity changes (caudate, hippocampus)	FA Diffusivity	FA, MD of: Cingulate gyrus Cingulum angular bundle Superior longitudinal fasciculus temporal part Uncinate	**Episodic memory** CVLT immediate recall	**Level-change** Not reported **Correlated change** ΔCingulate gyrus MD – ΔCVLT immediate recall
Köhncke et al. (2016), SNAC-K	442/70 with brain data (85.1 ± 4.5)	2	3	See Lövdén et al. (2014)	Bivariate $LCS_{brain/cog}$	Age Sex Education (years)	FA Diffusivity WMH volume	FA and MD of: Corticospinal tract Total WMH	**Processing speed** Digit cancellation, Pattern comparison	**Level-change** Not reported **Correlated change** ΔFA corticospinal tract + ΔprocessingPeed ΔMD corticospinal tract – ΔprocessingSpeed

Table 3 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+ = positive, – = negative)
Persson et al. (2016)	167 (19–79)	2	2	MMSE ≥ 26, CES-D ≤ 15	Bivariate $LCS_{brain/cog}$	Age SES VR APOE-ε4	WM volume, adjusted for ICV	Prefrontal WM (PFw)	**Fluid intelligence** CFIT (four subtests) **Episodic memory** Name-picture associative memory, story recall(both tests with immediate/ delayed recall) **Vocabulary** 2 vocabulary lists (selecting synonyms)	**Level-change** PFw T1 + Δfluid intelligence **Correlated change** ΔPFw + Δfluid intelligence[g]
Fjell et al. (2017), Cognition and plasticity through the lifespan	119 (23–86)	2	3.3	See Fjell et al. (2016)	Correlation (change $ratio_{brain}$)[h]	Age Head motion Time interval WM hypointensities	FA Diffusivity	FA, MD, RD, AD of: Corpus callosum Forceps major/minor Superior longitudinal fasciculus temporal (SLFT) and parietal (SLFP) Inferior longitudinal fasciculus (ILF) Cingulum: angular bundle and cingulate gyrus Corticospinal tract Anterior thalamic radiation Uncinate fasciculus	**EF** Stroop shifting	**Level-Change** Not reported **Correlated change** ΔILF MD – Δstroop shifting ΔSLFT MD – Δstroop shifting[i]
Gorbach et al. (2017), Betula	264 (55–80)	$2_{brain}/3–4_{cog}$	$4_{brain}/15_{cog}$	None[a]	Correlation (change $ratio_{brain}$/ regression slope $estimates_{cog}$)	Age VR	FA WMH volume	FA of corpus callosum (genu, body, splenium) Total WMH volume	**Fluid intelligence** Block design **Episodic memory** Composite of five tasks (immediate/delayed recall of commands with and without enactment, auditory word list) **Processing speed** Letter-digit substitution, Letter comparison, Figure comparison **Verbal fluency** Letter fluency task (three conditions)	**Level-change** Not reported **Correlated change** No effects

Table 3 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+ = positive, − = negative)
Leong et al. (2017), Singapore Longitudinal Aging Brain Study	111 (56.1–83.1)	2–4	1.5–2	MMSE ≥ 26	Correlation (MLM slope estimates$_{brain/cog}$)	Age Sex Education VR BMI ICV	WM volume	Total WM Lobar WM (frontal, parietal, temporal, occipital)	**Global cognition** Composite of all abilities below **EF** Category fluency (three categories) Design fluency TMT B **Processing speed** TMT A Symbol search Symbol-digit modalities **Visuospatial memory** Visual paired associates test (immediate, delayed recall) **Verbal memory** AVLT (immediate, delayed recall) **Attention** Digit span Spatial span (forward and backward)	**Level-change** Not reported **Correlated change** ΔTotal WM, Δparietal, Δfrontal WM + Δverbal memory
Moon et al. (2017), MAPT	109[j] (74.8 ± 4.1)[j]	$2_{brain}/5_{cog}$	3	MMSE ≥ 24 CDR = 0[k]	MLM analysis (difference score$_{brain}$)	Age Sex Education (level) VR ICV APOE-ε4 CDR status	WMH volume	-	**Episodic memory** Free and cued selective reminding test (FCSRT; free/ total recall) **Processing speed and EF** TMT A TMT B Controlled oral word association test (COWAT) **Global cognition** FCSRT MMSE Digit-symbol substitution Category fluency	**Level-change** WMH T1 − ΔTMT A **Correlated change** ΔTotal WMH − ΔTMT A, ΔTMT B

Table 3 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging: screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+ = positive, – = negative)
Song et al. (2018), DLBS	55 (55–89)	2	4	MMSE ≥ 26	Regression analysis (annual change$_{brain}$ [d])	Age Sex	Diffusivity	RD of fornix crus	**Fluid intelligence** Letter sets RAPM Stockings of Cambridge task **Crystallized intelligence** Advanced vocabulary scale **Episodic memory** Free recall (immediate, delayed) Hopkins verbal learning (delayed recognition) Verbal recognition memory (free recall) **Processing speed** Digit-symbol substitution Digit comparison	**Level-change** No effect **Correlated change** No effect

brain, brain measurements; cog, cognitive measurements; GENIE, St GEorge's Neuropsychology and Imaging in the Elderly; SNAC-K, Swedish National study on Aging and Care in Kungsholmen; DLBS, Dallas Lifespan Brain Study; MAPT, Multidomain Alzheimer Preventive Trial; HAM-D, Hamilton Depression Rating Scale; MMSE, mini-mental-state-examination; BDI, Beck Depression Inventory; BIMC, blessed information-memory-concentration test; CES-D, Center for Epidemiologic Studies Depression Scale; CDR, clinical dementia rating; LCS, latent change score model; MLM, multilevel model; change ratio, a measure of change proportionalized on the baseline level and in some cases also on the time interval (for exact formula consult original publications); difference score, value at T1 – value at T2; SES, socio-economic status; VR, vascular risk (for details refer consult original publications); APOE-ε4, apolipoprotein E allele 4; BMI, body mass index; FA, fractional anisotropy; MD, mean diffusivity; RD, radial diffusivity; AD, axial diffusivity; WMH, WM hyperintensities; ICV, intracranial volume; LPFC, lateral prefrontal cortex; TMT, trail making test; AVLT, Rey auditory verbal learning test; CFIT, Catell Culture Fair Test; CVLT, California Verbal Learning Test; EF, executive functions; T, measurement occasion (e.g. T1 = measurement occasion 1); Δ, change; WM, white matter.

[a]No information provided on screening instruments and/or cut-offs used (for description of healthy aging criteria consult original publication). [b]None of these effects were significant anymore after including change in whole brain volume as a predictor (cf. Table 4). [c]GM predictors were included in the same model (cf. Table 2). [d]Change measure not clearly described. [e]A lack of interindividual differences in intraindividual change in the cognitive variables precluded correlated change analyses. [f]All cognitive variables were included in the same statistical model. [g]Of the cognitive measures, only intelligence showed significant interindividual differences in intraindividual change and was used for the correlated change analyses. [h]Change measure for the cognitive variables not clearly described. [i]Not significant after controlling for age. [j]Sample size and age range of the full sample. The results reported here are from a stratified subsample of cognitively healthy participants for which this information was not indicated. [k]All subjects reported limitations in one instrumental activity of daily living and memory complaints.

Table 4: Overview of prospective observation studies investigating level-change or correlated change associations of whole brain volume/CSF and cognitive abilities.

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging:screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, −=negative)
Whole brain										
Schmidt et al. (2005), Austrian Stroke Prevention Study	329 (50–75)	2–3	3	None[a]	Repeated measures regression analysis (change $ratio_{brain}$)	Age Sex Education (years) VR Diabetes Cardiac disease Smoking Cholesterol	Whole brain volume (NBV)	-	**Episodic memory** Lern- und Gedächtnistest (immediate, delayed recall) **Reasoning** WCST **Attention and speed** Alters-Konzentrations-Test TMT B Digit span Complex reaction time test	**Level-change** Not reported **Correlated change** ΔNBV + Δepisodic memory, Δattention and speed
Charlton et al. (2010), GENIE	106 (50–90)	2	2	None[a]	Correlation (change $ratio_{brain/cog}$)	Age	Whole brain volume (NBV)	-	**Working memory** Digit span backwards, Letter-number sequencing	**Level-change** No effect **Correlated change** No effect
Grimm et al. (2012), BLSA	149 (56–86)	6.8 ± 2.8	1	None[a]	Bivariate $LCS^{b}_{brain/cog}$	None	Ventricular volume (CSF)	Lateral ventricles	**Episodic memory** CLVT	**Level-change** No effect (in final model) **Correlated change (lagged)** ΔLateral ventricle volume (previous) – ΔCLVT (subsequent)
Mak et al. (2015a)	33 normal controls (76.7 ± 5.3)	2	1	No history of dementia or CAMCOG score < 80	Correlation (change $ratio_{brain/cog}$)	Age Sex	Whole brain volume (TBV)	-	**Global cognition** CAMCOG	**Level-change** No effect **Correlated change** No effect
Ritchie et al. (2015b), Lothian Birth Cohort	657 (M age 73)	2	3	MMSE ≥ 24	Multivariate $LCS^{c}_{brain/cog}$	Age, Sex	Whole brain volume (TBV)	-	**Fluid intelligence** Matrix reasoning Block design Digit span backward Letter-number sequencing **Episodic memory** Logical memory (immediate, delayed recall) Verbal paired associates Spatial span (forward, backward) **Processing speed** Digit-symbol Symbol search Simple-/four-choice reaction time task Inspection time task	**Level-change** Episodic memory T1 + ΔTBV **Correlated change** No effect

Table 4 (continued)

Brain measure/ author, study	Sample size at baseline (age range in years)	Measurement occasions	Time intervals in years	Healthy cognitive aging:screening instruments at baseline	Statistical method (change measure)	Covariates	Neuroanatomical measure (cf. Table 1)	Brain regions	Cognitive variables	Brain-cognition relations (+=positive, −=negative)
Leong et al. (2017), Singapore Longitudinal Aging Brain Study	111 (56.1–83.1)	2–4	1.5–2	MMSE ≥ 26	Correlation (MLM slope estimates$_{brain/cog}$)	Age Sex Education VR BMI ICV	Whole brain volume (TBV) Ventricular volume (CSF)	-	**Global cognition** Composite of all abilities below **EF** Category fluency (three categories) Design fluency TMT B **Processing speed** TMT A Symbol search Symbol-digit modalities **Visuospatial memory** Visual paired associates test (immediate, delayed recall) **Verbal memory** AVLT (immediate, delayed recall) **Attention** Digit span Spatial span (forward and backward)	**Level-change** Not reported **Correlated change** ΔTBV + Δglobal cognition, Δverbal memory ΔTotal ventricular volume–Δverbal memory, ΔEF

brain, brain measurements; cog, cognitive measurements; GENIE, St GEorge's Neuropsychology and Imaging in the Elderly; LCS, latent change score model; MLM, multilevel model; change ratio, a measure of change proportionalized on the baseline level and in some cases also on the time interval (for exact formula consult original publications); difference score, value at T1 – value at T2; CSF, cerebrospinal fluid; ICV, intracranial volume; TBV, total brain volume; NBV, normalized brain volume; TMT, trail making test; AVLT, Rey auditory verbal learning test; EF, executive functions; MMSE, mini-mental-state-examination; CAMCOG, Cambridge Cognitive Examination; VR, vascular risk (for details refer consult original publications); APOE-ε4, apolipoprotein E allele 4; BMI, body mass index; Δ, change.

[a]No information provided on screening instruments and/or cut-offs used (for description of healthy aging criteria consult original publication). [b]Change was modeled over age (instead of time; consult original publication for details). [c]All cognitive variables were included in the same model.

Box 2: Evidence from cognitive training studies.

According to STAC-r, interventions such as cognitive training can stimulate compensatory scaffolding directly. Consequently, experimental training studies using neuroimaging provide the unique opportunity to investigate the indirect pathway between structural brain and cognitive aging, deepening the understanding of how compensatory mechanisms on the level of structural brain changes may reverse or attenuate age-related cognitive decline. In recent years, there has been an increasing interest in cognitive training interventions, because they constitute a potentially powerful, safe, and economical approach to prevent age-related cognitive decline. Compared to nonexperimental longitudinal studies, combined training, and neuroimaging studies provide strong causal inference on the influence of cognitive exercising on brain structure. To gain insight into the neuroanatomical underpinnings associated with training-related cognitive changes in older adults, a number of neuroimaging studies were conducted to reveal alterations in GM and WM, respectively (see Valkanova et al., 2014; ten Brinke et al., 2017, for reviews).

Cortical increases in GM in areas that are associated with the trained cognitive ability are found after an 8-week strategy memory training (i.e. the method of loci; Engvig et al., 2010) and 12-week computer-based multi-domain training targeting memory, attention, response speed, EF, and language (Lampit et al., 2015). Interestingly, the observed structural changes seem to be stronger at the beginning of training than in later stages of training. In the study of Lampit et al. (2015), more than half of the increase in GM occurred within the first 3 weeks of training, whereas the following 9 weeks of training resulted in relatively smaller increase. These findings are in line with the recently proposed expansion-renormalization model (Wenger et al., 2017). According to this model, learning- or training-related neuroanatomical changes are characterized by three stages: expansion, selection, and renormalization. Whereas at the beginning of a training intervention brain tissue expands (potentially by changes to synapses, glial cells, or vasculature and to a limited extent via the generation of additional neurons), brain tissue starts to return to the normal, baseline level when the cognitive process can be optimally performed. However, as the cellular mechanisms underlying GM tissue changes in humans are still not sufficiently understood and studied, this model requires further validation. In addition, these promising findings are contrasted by studies that found no differences in GM tissue between the experimental and a control group after an 8-week attention and distractibility training (Mozolic et al., 2010), a 26-week multidomain COGPACK training (Suo et al., 2016), and a 16-week spatial navigation training (Wenger et al., 2012). However, Wenger et al. (2012) found a trend towards cortical thinning in the control group, which was not observed in the navigation group.

A small number of studies have also investigated training-related changes in WM. They consistently found that the experimental group exhibited an increase in FA compared to the control groups (both passive and active) in related brain areas (Lövdén et al., 2010; Engvig et al., 2012; Chapman et al., 2015; de Lange et al., 2017). Further, training-related changes in MD have also been reported (Lövdén et al., 2010, 2012; de Lange et al., 2017). One study has also investigated long-term changes in WM 12 months after completion of a cognitive training intervention compared to an active control (Cao et al., 2016). The authors found an overall trend that in the multidomain training group, AD decreased while FA, MD, and RD remained stable. In the control group, however, FA decreased, while MD and RD increased.

Only few studies have, however, directly investigated the association between the size of training gains and the size of training-induced structural brain changes. Some of them reported a positive relationship between training-induced cognitive improvement and neuroanatomical change indices, indicating that individuals who showed the largest improvements during training also showed the strongest changes in GM (Engvig et al., 2010; Lampit et al., 2015), and WM (Engvig et al., 2012; de Lange et al., 2017). But again, other studies found no correlation between cognitive training performance and training-induced changes in GM (in younger adults; Wenger et al., 2012) and WM (Lövdén et al., 2010, 2012). So far, only one study has investigated the effect of repeated phases of cognitive training on neural plasticity and training-gains (de Lange et al., 2018). The authors could show that age-related WM microstructural decline over the study period of 40 weeks was attenuated during phases of memory training, supporting a mitigating effect of cognitive training on brain aging. In contrast, memory performance was less dependent on continued training, showing stability after an initial training-induced gain.

In conclusion, so far only a small number of cognitive intervention studies have investigated training-induced structural brain changes, finding either increases or stability in brain structure, as compared to a control group. In addition, if reported, associations between structural brain changes and cognitive training-gains were either positive or nonsignificant. However, large methodological differences between studies limit the generalizability of these findings. In general, the field of cognitive intervention studies is still emerging and further research is required to determine what type of cognitive training and in which dosage (i.e. intensity, frequency, and duration) is required to achieve maximum training gains and structural brain changes. Besides cognitive interventions, studies administering physical interventions have also reported training-induced structural brain and cognitive performance changes, however, this literature goes beyond the scope of the present literature and is reviewed elsewhere (e.g. Brehmer et al., 2014; Mandolesi et al., 2018). While most of the existing training studies rely on a group design, a promising trend for future research is the development of individually targeted interventions, based on neuroanatomical predispositions (Park et al., 2018) and nonbiological factors (Guye et al., 2016) that can help individuals to best maintain their health and well-being far into old age.

(see Table 2). Among these articles, 10 reported level-change associations, of which eight showed a significant result (Raz et al., 2008; Persson et al., 2012; Fjell et al., 2014; Möller et al., 2016; Persson et al., 2016; Hohman et al., 2017; Anblagan et al., 2018; Yuan et al., 2018) and two did not (Leow et al., 2009; Ritchie et al., 2015b). Moreover, 16 studies were able to compute results for correlated changes, of which eight were also significant (Persson et al., 2012; Fjell et al., 2014; Ritchie et al., 2015b; Möller et al., 2016; Gorbach et al., 2017; Leong et al., 2017;

Sala-Llonch et al., 2017; Anblagan et al., 2018) and eight were nonsignificant (Cohen et al., 2001; Raz et al., 2007; Leow et al., 2009; Daugherty et al., 2015; Mak et al., 2015b; Fjell et al., 2016; Persson et al., 2016; Hohman et al., 2017). However, as not all studies did report results for both level-change and correlated change associations, the proportion of significant results might be biased. In addition, the potential threat of a publication bias needs to be taken into account. Studies with nonsignificant associations are less likely to get published (Ioannidis et al., 2014), and it is possible that the studies reported here are a selective subset of the literature reporting significant results.

Intelligence

Five studies observed level-change or correlated change associations of GM with a measure of intelligence (Raz et al., 2008; Ritchie et al., 2015b; Persson et al., 2016; Leong et al., 2017; Yuan et al., 2018). Two very well-powered studies used latent change score models (LCS; McArdle and Hamagami, 2001; McArdle, 2009) to assess correlated changes between latent measures of GM volume and latent measures of cognitive ability (Ritchie et al., 2015b: n = 657; Persson et al., 2016: n = 167). The LCS model is estimated in the structural equation modeling (SEM) framework and allows the estimation of a latent change score between two subsequent measurement occasions, thus separating true change (at least in part) from measurement error. Furthermore, using the LCS model, it is possible to separate interindividual differences from intraindividual change (for a further explanation of this model, see section 'The benefits of longitudinal designs' on latent change models). Ritchie et al. (2015b) found significant correlated changes between GM volume and fluid intelligence, such that steeper declines in GM volume were associated with steeper declines in fluid intelligence over a period of 3 years. Furthermore, Persson et al. (2016) reported that participants with lower baseline GM volume in cerebellar hemispheres, parahippocampal gyrus, and hippocampus showed larger declines in fluid intelligence over 2 years. Also using the LCS model, Raz et al. (2008) found that steeper changes in entorhinal cortex volume were associated with lower levels of fluid intelligence in a sample of young and older adults (age range 20–77). Moreover, Yuan et al. (2018) reported differential effects of fluid and crystallized intelligence on GM aging: while participants with higher fluid intelligence levels demonstrated reduced cortical GM volume shrinkage over a time span of around 5 years, participants with higher levels of crystallized intelligence showed steeper declines in cortical GM volume and total cortical thickness. The authors were, however, unable to compute correlated change, due to a lack of interindividual differences in the cognitive variables. Finally, using a composite of global cognition across five fluid cognitive tasks (i.e. EF, processing speed, verbal, and visuospatial memory), Leong et al. (2017) reported positive correlated changes of this measure with total GM volume (specifically frontal, parietal, and temporal lobar volumes) and hippocampal volume. In contrast, the authors reported negative correlated changes between lobar GM in the occipital cortex and global cognition, indicating that on average, increases in occipital GM were associated with decreases in global cognition.

Memory

The most consistent evidence was found for correlated changes between episodic memory and GM volume or GM thickness in medial temporal regions (Persson et al., 2012; Fjell et al., 2014; Gorbach et al., 2017; Hohman et al., 2017; Leong et al., 2017; Anblagan et al., 2018). With one exception (n = 26: Persson et al., 2012), these studies included large (e.g. n = 111: Leong et al., 2017) to very large (e.g. n = 655: Anblagan et al., 2018) sample sizes.

Two studies could show positive longitudinal associations between hippocampal volume and episodic memory, such that older adults with larger baseline hippocampal volumes (Hohman et al., 2017), or less hippocampal atrophy over time (Leong et al., 2017) showed less declines in episodic memory over a period of 4–8 years, respectively. Furthermore, Persson et al. (2012) reported that changes in episodic memory were positively correlated with changes in the right hippocampus, as well as hippocampal volume at follow-up (after 6–10 years) in a sample of middle-aged to older healthy participants. Similarly, Gorbach et al. (2017) found positive correlations between 15-year changes in an episodic memory composite of five tasks and simultaneous 4-year changes (towards the end of the same testing period) in GM volume of the hippocampus in a sample of healthy middle-aged to older adults. Notably, this effect was driven by the participants aged over 65. To reduce practice effects, the authors used slightly different versions of the episodic memory tasks across measurement occasions (i.e. by switching item lists between tasks or changing the item order). Also, one study found that baseline hippocampal microstructure as measured with MD (more conventionally used to detect WM changes; see Table 1), was associated with changes in verbal episodic memory (Anblagan et al., 2018), such

that participants with higher MD showed steeper subsequent declines in episodic memory performance. In contrast, correlated changes between hippocampal volume and verbal episodic memory did not survive correction for multiple comparisons. The authors speculate that higher MD values might reflect an age-related increase in water content in hippocampal tissue that could be a precursor for age-related pathological changes that influence cognitive abilities before brain atrophy can be observed (Anblagan et al., 2018).

Positive correlated changes with episodic memory were also reported for cortical thickness in the entorhinal cortex (Fjell et al., 2014) and in right hemispheric regions (Sala-Llonch et al., 2014; Möller et al., 2016). Finally, one study reported a significant association of higher baseline episodic memory performance with reduced 2-year GM volume declines in the lateral prefrontal cortex in a sample including younger and older adults (age span of 19–79 years; Persson et al., 2016).

EF

Four studies reported level-change associations for a measure of GM and EF: in a large sample of healthy controls of the Alzheimer's disease and neuroimaging initiative (ADNI; n = 379), baseline hippocampal volume was positively related to changes in EF over 4 years (Hohman et al., 2017), suggesting that older adults with lower hippocampal volume showed steeper declines in EF. Similarly, Leong et al. (2017) reported positive correlated changes between hippocampal volume and EF over 8 years of follow-up. In addition, in one study, 2-year changes in cortical thickness of the right occipital cortex were negatively correlated with simultaneous performance changes in EF tasks (Möller et al., 2016).

Working memory

Only one study found longitudinal relationships between hippocampal MD and working memory (Anblagan et al., 2018): unlike the level-change association, which went into the expected direction, correlated changes between MD and working memory were positive, suggesting that increases in MD (usually interpreted as disruptions in WM microstructure) were related to improvements in working memory. However, due to the lack of specificity of MD, other influences (e.g. crossing fibers) could potentially have led to a local increase in MD (Zatorre et al., 2012).

Processing speed

In two well-powered studies using data from the Lothian Birth cohort, more intact hippocampal GM at baseline (i.e. lower MD; Anblagan et al., 2018), and less decline in whole brain GM volume (Ritchie et al., 2015b) were associated with reduced declines in a latent measure of processing speed over 3 years.

However, nine studies did not show any longitudinal relationships between GM and cognition regarding correlated change (Cohen et al., 2001; Raz et al., 2007; Daugherty et al., 2015; Mak et al., 2015b; Fjell et al., 2016; Persson et al., 2016; Hohman et al., 2017), level-change (Ritchie et al., 2015b) or both (Leow et al., 2009). Several of these studies had comparatively small sample sizes (between n = 25 and 56) and might have thus not been sufficiently powered to detect any significant effects (Cohen et al., 2001; Raz et al., 2007; Leow et al., 2009; Mak et al., 2015b; Fjell et al., 2016).

WM and cognitive ability

In total, 18 studies investigated longitudinal associations between measures of WM structure and cognitive ability (see Table 3). Out of these, nine studies reported level-change associations, of which four were significant (Ritchie et al., 2015a,b; Persson et al., 2016; Moon et al., 2017), and five were not (Raz et al., 2008; Silbert et al., 2008; Charlton et al., 2010; Bender et al., 2016a; Song et al., 2018). Moreover, 17 studies reported correlated change associations, of which 14 were significant (Schmidt et al., 2005; Raz et al., 2007; Silbert et al., 2008; Charlton et al., 2010; Lövdén et al., 2014; Ritchie et al., 2015a,b; Bender et al., 2016a; Köhncke et al., 2016; Persson et al., 2016; Fjell et al., 2016, 2017; Leong et al., 2017; Moon et al., 2017), and three were not (Schmidt et al., 1999; Gorbach et al., 2017; Song et al., 2018). Again, many studies did not report results for both level-change and correlated change, therefore the proportion of significant results should be interpreted with caution.

Intelligence

Several studies found positive level-change or correlated change associations between global measures of WM health and some measure of fluid intelligence, such that lower baseline WM volumes or faster degradation of WM were associated with larger declines in intelligence performance. Specifically, accumulation of global WMH

volume over time (Schmidt et al., 2005; Raz et al., 2007; Ritchie et al., 2015b), decrease in global FA (Ritchie et al., 2015a), and both lower baseline prefrontal WM volume and atrophy therein (Persson et al., 2016) were associated with declines in intelligence over time. These studies all covered a time span between 2 and 6 years. Furthermore, several of these studies used LCS models to investigate interindividual differences in intraindividual changes (Ritchie et al., 2015a,b; Persson et al., 2016), and included large to very large sample sizes given the standards in the field (n = 731: Ritchie et al., 2015a; n = 657: Ritchie et al., 2015b; n = 167: Persson et al., 2016).

Memory

Some evidence also exists for correlated change associations between WM health and measures of episodic memory (Schmidt et al., 2005; Silbert et al., 2008; Ritchie et al., 2015b; Fjell et al., 2016; Bender et al., 2016a; Leong et al., 2017). In one study, larger declines in global and lobar (i.e. parietal and frontal) WM volume were related to steeper declines in a composite reflecting verbal episodic memory performance (Leong et al., 2017). Furthermore, three studies reported negative correlated changes between WMH and episodic memory (global WMH: Schmidt et al., 2005; subcortical WMH: Silbert et al., 2008; global WMH: Ritchie et al., 2015b), which can be interpreted as a positive relationship between changes in WM health and episodic memory. Also, one study reported a relationship between changes in WM microstructure (increases in MD of the cingulate gyrus) typically understood as WM deterioration, and declines in episodic memory (Fjell et al., 2016). However, Bender et al. (2016a) observed the opposite relationship. In their study, changes in WM microstructure (decreases in FA, increases in RD), which are commonly interpreted as WM degradation were correlated with improvements in episodic memory. Given the uncertainty regarding the cellular mechanisms of change in diffusion properties, however, it is also possible that the pattern of WM changes reflects a form of plastic reorganization (Bender et al., 2016a).

EF

Two studies reported evidence for level-change or correlated change associations between WM and EF (Fjell et al., 2017; Moon et al., 2017). Fjell et al. (2017) found a negative correlation between MD changes in the inferior and superior longitudinal fasciculi (averaged across hemispheres) with performance changes in a shifting condition of the stroop task across a period of 3 years. However, the result for the superior longitudinal fasciculus did not survive control for age. This means that increases in MD in the inferior longitudinal fasciculus were related to declines in inhibitory control, independent of participant age. In a subsample of cognitively healthy elderly, Moon et al. (2017) reported a negative association between WMH progression and steeper 3-year declines in EF, measured as performance in the Trail Making Test B. This relationship was not found for the entire sample, which contained also participants with impaired cognition (i.e. clinical dementia rating scale > 0).

Working memory

One study reported negative correlated changes between a measure of WM microstructure (i.e. MD; higher values reflect lower integrity) and a composite of two working memory tasks (Charlton et al., 2010). These tasks are elsewhere interpreted as measures of fluid intelligence (Ritchie et al., 2015a,b).

Processing speed

Overall, five studies found level-change or correlated change associations between WM health and processing speed. Notably, these studies were very well-powered, with sample sizes above n = 400 (with the exception of Moon et al., 2017). Based on participants from the same sample of healthy older adults, two studies reported positive correlated changes over 2 years between indices of WM microstructure (decreases in FA, increases in MD) of the corticospinal tract and processing speed (Lövdén et al., 2014; Köhncke et al., 2016), indicating that older adults with less intact WM microstructure in the corticospinal tract show steeper declines in processing speed. Furthermore, lower baseline global FA (Ritchie et al., 2015a), higher global WMH at baseline, and higher WMH increases were associated with steeper declines in processing speed (Ritchie et al., 2015b; Moon et al., 2017).

Finally, seven studies did not show any longitudinal relationships between WM and cognition regarding correlated changes (Schmidt et al., 1999; Gorbach et al., 2017), level-change associations (Raz et al., 2008; Silbert et al., 2008; Charlton et al., 2010; Bender et al., 2016a), or both (Song et al., 2018). The sample sizes of these studies were very heterogeneous, with some including smaller (e.g. n = 84: Charlton et al., 2010; n = 55: Song et al., 2018), and others large samples of more than 200 participants (Schmidt et al., 1999, 2005; Gorbach et al., 2017).

Whole brain volume, CSF, and cognitive ability

Overall, six studies reported longitudinal associations between measures of whole brain volume or CSF, and cognitive ability (see Table 4). Four of these studies reported level-change associations, of which one showed a significant result (Ritchie et al., 2015b), and three did not (Charlton et al., 2010; Grimm et al., 2012; Mak et al., 2015a). Moreover, all six studies reported results for correlated changes, of which three were also significant (Schmidt et al., 2005; Grimm et al., 2012; Leong et al., 2017), and three were nonsignificant (Charlton et al., 2010; Mak et al., 2015a; Ritchie et al., 2015b).

Memory

Three studies found positive level-change (Ritchie et al., 2015b) or correlated change associations (Schmidt et al., 2005; Leong et al., 2017) of a measure of episodic memory and of a measure of whole brain volume (see Table 1), in the direction that participants with lower levels of episodic memory performance at baseline, or decreases therein, showed on average steeper decline in whole brain volume. These studies were well to very well-powered (n = 329: Schmidt et al., 2005; n = 657: Ritchie et al., 2015b; n = 111: Leong et al., 2017).

With regard to CSF, two studies found negative correlated change associations between ventricular volume and episodic memory performance, such that ventricular enlargement (CSF increases) was related to simultaneous or subsequent declines in memory performance (Grimm et al., 2012; Leong et al., 2017). Specifically, we would like to highlight the study by Grimm et al. (2012), which was the only one to model lagged correlated changes. The authors assessed 149 participants in a measure of CSF (lateral ventricle volume) and episodic memory with seven repeated measurements over a maximum period of 10 years. They estimated change across participant age (60–90 years) in a bivariate dual change score model (a variant of a LCS model). Conducting a series of model comparisons, they concluded that a final model where previous changes in CSF led to subsequent changes in episodic memory reflected the data best, whereas the other directionality (cognitive changes leading to brain changes) lowered model fit substantially. We would like to emphasize this study, as it provides the methodological tools to explore directional hypotheses in the study of dynamic within-person associations between changes in brain structure and cognitive ability. As the expansion of CSF in the ventricles indicates a loss of brain volume, these results, when taken together, mirror the findings for whole brain volume.

Other

In addition, Schmidt et al. (2005) reported positive correlated changes of NBV with a composite of four cognitive tasks that tap into attention and processing speed. Another study reported a negative association of ventricular expansion with executive function declines (Leong et al., 2017).

However, two studies did not find any significant association regarding level-change and correlated change associations (Charlton et al., 2010; Mak et al., 2017). These studies had smaller sample sizes compared to the studies reported above that found significant brain structure-cognition relations (i.e. n = 106: Charlton et al., 2010; n = 33: Mak et al., 2017), thus having less power to detect any significant effects.

Summary

In conclusion, so far only a small number of studies have investigated level-change or correlated change associations between measures of brain structure and cognitive ability. Several of these studies report positive associations, indicating that declines (or increases) in structural brain intactness are related to simultaneous losses (or gains) in cognitive performance. However, others have found the opposite relation, such that brain structure was negatively associated with cognitive ability, or showed no significant correlation at all. In light of the scarcity of evidence, the positive association between structural brain properties of medial temporal regions (specifically the hippocampus) and global brain metrics with episodic memory has received comparatively more attention. As the hippocampus is involved in neurogenesis up into old age (Lillard and Erisir, 2011), this structure is highly relevant for brain plasticity, and potentially also for compensatory scaffolding (see also Park and Reuter-Lorenz, 2009). We also found some support for level-change or correlated changes of global GM and WM indices with fluid intelligence, and a few studies showed brain structure relations with EF and processing speed, however, these findings are more heterogeneous. Generally, the interpretation of significant correlated change associations between structural brain indices and cognitive abilities poses a challenge, as it is impossible to disentangle whether age-related neuroanatomical changes are causally linked to simultaneous cognitive

declines, or whether a (positive) relationship between these variables reflects a common, underlying causal mechanism. Based on STAC-r, we expected associations between brain structure and cognitive ability to be weak especially in healthy older adults as they should be able to compensate for age-related brain atrophy and thus maintain cognitive performance. Unfortunately, it was not possible to derive a fully comparable measure of the strength of the correlated change relationships, due to large methodological differences between studies. Even if it was possible to gain an effect size estimate, it would be impossible to judge at the present moment if weak brain structure-cognition relationships resulted from methodological limitations of the reviewed studies or actually from intraindividual compensatory resources as predicted by STAC-r. We therefore conclude that the current limitations of the reviewed literature prohibit general conclusions on correlated change relationships in healthy aging. In the following sections, we will discuss these limitations in-depth, and provide ideas for the advancement of methods and theories in the field.

Methodological limitations and the need for methods development

The concerted findings from prospective observation studies provide evidence for intertwined changes between WM and GM and whole structural brain correlates and cognitive abilities in healthy aging individuals (see Tables 2–4). Furthermore, the results gained from cognitive intervention studies provide causal insights into the relation between brain structure and cognitive ability changes. However, taken together, the results are far from being consistent with respect to the brain regions and cognitive measures that are associated with each other, or even regarding the directionality of the relation between brain structure and cognition (Bender et al., 2016a; Leong et al., 2017; Anblagan et al., 2018). In the following sections, we will discuss methodological limitations of the reviewed studies and potential avenues for methodological advancement.

Methodological limitations

Reliability, shape, and dynamics of change

The majority of the reviewed studies assessed only two measurement occasions. However, two-occasion studies are limited with regard to the reliability of the change estimate (Willett, 1989). A line drawn through two observations will always fit perfectly and hence measurement error will be ignored (King et al., 2018). Consequently, the reliability of the change estimate in two-occasion studies is highly dependent on the accuracy of the individual brain or cognitive measures observed at each occasion (see the sections on reliability and validity of structural brain measures/cognitive ability measures below). Regarding the shape of change, two-occasion studies allow only the estimation of linear change. Ideally, at least four occasions are required to estimate nonlinear trajectories of change (King et al., 2018). As already touched upon in the beginning of this article, cross-sectional and longitudinal evidence supports nonlinear age-related changes in brain structure and cognitive ability. Furthermore, only with more than two measurement occasions it is possible to study time-lagged relations between changes in two variables, and thus test directional hypotheses of the dynamics between structural brain and cognitive changes. It is reasonable to assume that changes in two developmental variables are not perfectly synchronized, but rather follow a lagged pattern, potentially even with bidirectional relations (Salthouse, 2011). Only few studies included three or more measurement occasions (Schmidt et al., 2005; Silbert et al., 2008; Grimm et al., 2012; Gorbach et al., 2017; Leong et al., 2017; Moon et al., 2017; Yuan et al., 2018) that would allow the estimation of more complex change dynamics. Of these studies, all assumed linear curves to estimate correlated change, and only two studies specifically tested first if a nonlinear trajectory fit the data better (which it did not; Hohman et al., 2017; Leong et al., 2017). Furthermore, only one study estimated lagged change relations, using, however, a rough measure of global brain structure changes (lateral ventricle size; Grimm et al., 2012).

Choice of age-range

The samples used in the reviewed studies differed substantially regarding the age ranges covered. Several studies selected a lifespan approach, covering a broad age range from young or middle adulthood to old age. Such lifespan samples are useful for research, as they can inform about development across a longer time span that would otherwise be almost impossible to gather from following one cohort across the entire lifespan. However, one needs to be careful when comparing results of age-heterogeneous studies to those from studies with more narrow age ranges. In age-heterogeneous studies, average change is composed of both within-person change and between-person age-differences. This can be problematic if these

two types of age effects do not converge. Age convergence can be formally tested for (Sliwinski et al., 2010), however, it is often not realistic to find age convergence in studies with broad age ranges, as very different developmental mechanisms might be at work in adults of different ages. For example, the chance of experiencing declines in structural brain features and cognitive performance is much higher around age 70 than age 50, and negligibly small for adults in their twenties. Thus, if these two sources of variance are not properly distinguished, estimated rates of change can be confounded with an increased probability of change with age at the between-person level (Hofer and Sliwinski, 2001). Different methodological approaches exist to include between-person age differences better into the analysis. For example, participant's baseline age can be included as a predictor to control for between-person age differences (see Sliwinski et al., 2010 for an extended discussion).

Definition of healthy aging

As already touched upon in the introduction of this article, healthy aging is not a well-defined term. The current definition of healthy aging provided by the WHO places importance on the interplay between a person's resources (i.e. intrinsic capacity) and the living context on the micro to macro level that is necessary to retain satisfactory levels of well-being through the successful pursuit of one's personal goals. This definition of healthy aging is more inclusive than the more general definition of health as a state of 'complete physical, mental, and social well-being and not merely the absence of disease or infirmity' (WHO, 1946). The recent definition from the First World Report on Ageing and Health (WHO, 2015) states that healthy aging is a process that applies to individuals at varying levels of functional capacity and health.

Given that the current review focuses on the association between brain and cognition, and that cognitive health is an important predictor of well-being in old age, we understand healthy aging from a cognitive ability point of view. Thus, when using the term healthy aging, we mean aging in the absence of clinically relevant cognitive impairment. All of the studies reviewed here tested for cognitive impairment using one or more common screening instruments for dementia or psychiatric illness related to cognitive impairment (i.e. depression) or described their participants as cognitively normal or dementia-free without closer information on specific instruments or cut-offs used (see column 'Healthy cognitive aging' in Tables 2–4). However, they showed vast differences regarding the exact protocols used to determine cognitive health (e.g. medical screening by trained experts vs. brief health interview), and the level of detail they provided in describing these protocols, which is why we direct interested readers to the original publications for more information. Many studies used the same instruments to screen for cognitive impairment (e.g. MMSE; Folstein et al., 1975), lending to some comparability. However, the cut-off values for healthy cognitive aging varied substantially between studies (e.g. from a minimum score of 24 up to a score of 28 in the MMSE). Moreover, many studies applied additional exclusion criteria beyond cognitive health [e.g. diabetes, cardiovascular disease, or vascular risk (VR)], leading to substantial differences in the overall health status between study samples. For example, besides screening for a range of medical conditions (including neurological disorders such as AD), Möller et al. (2016) excluded participants with cerebrovascular disease. In contrast, Raz et al., specifically included a subsample of participants with medically treated hypertension, as they were interested in the effect of VR on brain and cognitive health (Raz et al., 2007, 2008). The choice of criteria influences the prevalence of people categorized as healthy agers substantially (McLaughlin et al., 2012; Rodriguez-Laso et al., 2018). This discrepancy in the use of screening instruments and cut-off criteria underlines the urgent need for a clear consensus on a definition of healthy aging. Finally, on a conceptual level, it is still a matter of debate whether cognitive decline is a normal aspect of healthy aging and disease represents a qualitatively different state of the brain or whether it simply reflects the starting point on a continuum to later disease progression (i.e. dementia) – and age serves as a proxy for pathological changes.

Choice of time interval

The time intervals between measurement occasions varied between studies, ranging from 1 year (Leow et al., 2009; Grimm et al., 2012; Fjell et al., 2014; Mak et al., 2015a,b) up to 15 years (Gorbach et al., 2017). Also, for some studies the time intervals were different for the MRI and cognitive measurements (e.g. Fjell et al., 2014; Gorbach et al., 2017). Depending on the width of the time window under investigation, it is possible that changes are driven by different developmental processes or external influences (Hofer and Piccinin, 2009). When planning a longitudinal study, researchers need to be aware that different neuroanatomical substrates (e.g. volumetric or surface-based measures, WMH, WM microstructure), or different regional structures most likely vary with regard to the timing and duration of

age-related change processes. Thus, especially if neuroanatomical measurements are also combined with cognitive performance assessments, there will probably not be any ideal time interval that is suited to capture all of these change processes. Importantly, the choice of the length of the time intervals substantially influences the magnitude of the parameters estimated in traditional statistical models for the analysis of longitudinal data (Voelkle and Oud, 2013). Unfortunately, not many guidelines exist that help researchers in choosing the optimal spacing between measurements when planning a longitudinal study of aging (Dormann and Griffin, 2015).

Choice of covariates

Many authors included one or more covariates into their statistical analysis, to control for the potentially confounding influence of third variables on the relationship between changes in brain structure and cognitive performance (see column 'Covariates' in Table 2). Notably, the selection of covariates was heterogeneous between studies, further preventing direct comparisons of the level-change and correlated change relations. For example, while some authors controlled for age only, others additionally included other covariates (e.g. VR, apolipoprotein E allele 4) to control for the impact of neurally depleting factors. A source of heterogeneity was also introduced by different approaches to control for brain size. While some authors decided to adjust the raw brain volumes with a measure of maximum healthy brain size such as ICV (e.g. Persson et al., 2016), others included it as a covariate (e.g. Leong et al., 2017). At the present moment, no clear consensus is reached on one approach in the literature, as many factors play a role (e.g. the measure of brain size: Jäncke et al., 2015), and it is yet unclear how the correction with a global measure of brain size impacts findings in longitudinal developmental studies (Mills and Tamnes, 2014). We acknowledge that the choice of covariates is not a trivial matter, and often neglected in the discussion of results. Furthermore, covariates are often selected retrospectively, based on the variables that are available in a dataset, and not necessarily based on the most relevant potential confounds. This is especially an issue when using large, publicly available datasets that might not be designed for the research question at hand. Even in an ideal scenario where all potential confounds are observed, the decision on which variables to include into an analysis still remains subjective. Importantly, the choice of covariates can substantially alter the results of an analysis, as is nicely demonstrated in a recent study by Silberzahn et al. (2018). They asked 29 teams of researchers to independently analyze the same dataset and answer the question, whether soccer referees are more likely to give red cards to dark skinned players. The conclusions the researchers made varied immensely, in part because of their choice of covariates. Especially with regard to observational studies, the inclusion of certain types of covariates can impose a threat to causal conclusions (see Rohrer, 2018). To enable future meta-analytical comparisons, we thus advocate the transparent reporting of the results both with and without the included covariates. In addition, future studies might consider including time-varying covariates in their analyses, if theoretical reasons exist that they impact brain and/or cognitive performance differentially over time. For example, high body mass index or hypertension is associated to higher risks of developing dementia if experienced in midlife (Kennelly et al., 2009; Kivimäki et al., 2018).

Statistical method

Another source of variation between studies relates to the statistical methods used to model correlated change associations. In most cases, the investigators decided to compute two change scores (raw difference score or a change ratio) and either run a standard correlation between them, or include one score as a predictor and the other as a dependent variable in a regression model (see column 'Statistical method' in Tables 2–4). In contrast to a raw difference score, a change ratio is typically standardized with regard to baseline values of brain structure or cognitive ability, thus reflecting a relative difference. As many authors used different methods to compute change ratios (e.g. annual percentage change vs. the proportion of level at T2 to level at T1), this lack of consensus introduces another source of noise complicating the comparability of previous findings. Irrelevant of the computation, however, change scores include not only variation due to change within individuals over time but they are also confounded by variation stemming from between-person differences. Importantly, the main interest of every investigation of brain structure-cognition relations is to make assertions on how the two variables of interest are related to each other within individuals. Thus, appropriate statistical methods for the analysis of longitudinal change are necessary that are able to isolate these different portions of variance. Such methods are, for example, variants of latent growth curve models (i.e. random coefficient or multilevel models) (Raudenbush and Bryk, 2002), or latent change models (McArdle and Hamagami, 2001; McArdle, 2009), which allow to estimate interindividual

(i.e. between-person) differences in intraindividual (i.e. within-person) change. Specifically, a variant of latent change modeling, the bivariate LCS model estimated in a SEM framework is well-suited to address questions of correlated change associations between two variables (Kievit et al., 2018; for more details on the method see section 'The benefits of longitudinal designs'). Several of the reviewed publications have already adopted this approach (Raz et al., 2008; Grimm et al., 2012; Lövdén et al., 2014; Daugherty et al., 2015; Ritchie et al., 2015a,b; Köhncke et al., 2016; Bender et al., 2016a; Persson et al., 2016; Anblagan et al., 2018).

Power to detect change

The sample sizes studied were highly diverse between studies. Especially, to detect correlated change, sufficient statistical power is necessary. While this is not usually a concern for moderately sized longitudinal studies covering a few years, power can be an issue for short-term longitudinal studies with few measurement occasions, few participants and small effect sizes (cf., Rast and Hofer, 2014). Generally, in order to investigate developmental change in both, cognitive abilities and brain structure, studies will need to cover years rather than months to provide robust estimates of (correlated) change (von Oertzen and Brandmaier, 2013).

Problem of multiple comparisons

Especially in the field of neuroimaging, many statistical tests are often conducted concurrently (e.g. for voxel-wise comparisons across the brain). Specifically, as in the case of the studies reviewed here, when testing hypotheses about correlated changes between different regions of the brain and multiple cognitive abilities, the number of simultaneous hypothesis tests is high, leading to an increased risk of making a type I error (Lindquist and Mejia, 2015). Some authors solved this problem by applying a correction for multiple comparisons (Lövdén et al., 2014; Persson et al., 2014, 2016; Ritchie et al., 2015b; Fjell et al., 2017; Gorbach et al., 2017). Such corrections typically lower the threshold of the p-value, which lowers the risk of false positive results, however, on the downside also leads to a reduction of the statistical power to find the effect of interest (e.g. correlated change). Different methods for dealing with multiple comparisons exist that are designed to keep the loss in statistical power to a minimum (see Lindquist and Mejia, 2015 for an overview). Another possibility in the current context is to use advanced multivariate statistical methods that are able to run multiple hypothesis tests in one model (see section 'The benefits of longitudinal designs'). In any case, we consider it of high importance to report the applied procedure transparently, and ideally results both with and without corrections for multiple comparisons, as was done, for example, by Ritchie et al. (2015b).

Reuse of data

Several of the reviewed publications use the same sample, or a subsample of the larger participant pool, to address different research questions. Conducting large-scale longitudinal studies requires a lot of time and resources and we strongly agree that it is important to pool efforts and use the collected data in a sustainable and efficient manner. However, it would be desirable to openly communicate this matter and to explicitly address the statistical consequences of using data from the same sample to answer different research questions. We therefore advocate the transparent documentation of the publications that have used data from the same longitudinal database, for example, via an open science platform, such as the Open Science Framework (Foster and Deardorff, 2017).

Reliability and validity of structural brain measures

Another limitation relates to the reliability and validity of indices derived from brain imaging. First, the protocols used to process longitudinal brain imaging data vary greatly between research groups. Whereas some groups use manual or semi-automated methods to delineate anatomical regions of interest, others rely on fully automated procedures. Some use default settings and others additionally apply fine-tuning to such default protocol parameters. This variety introduces unwanted noise into any efforts of replication (Mills and Tamnes, 2014).

An important topic for longitudinal investigations is the concern of retest-reliability. Generally, the measurement of compartmental volumes, surface area and cortical thickness with automated methods is reliable across repeated testing (Vijayakumar et al., 2017), with regional scan-rescan reliabilities ranging between 0.8 and 0.9 (Liem et al., 2015). However, factors such as the measured structure (Morey et al., 2010), choice of segmentation software and protocol (Jovicich et al., 2013; Heinen et al., 2016), or magnetic field strength (Heinen et al., 2016) can impact the reliability. With regard to the assessment of brain structure-cognition relationships, Dickerson et al. (2008) reported reliable estimation of cortical thickness

correlates of cognitive performance across different sessions, scanners, and field strengths in a group of healthy older adults. In comparison, measurements from DW-MRI are less robust, with estimated scan-rescan reliabilities between 0.5 and 0.8 in dependence of preprocessing choices and WM measure of interest (cf., Madhyastha et al., 2014). As DW-MRI registers signal-loss due to the movement of water-molecules, the sensitivity to detect changes is specifically affected by head motion artifacts (Yendiki et al., 2014). Even though to a lesser extent, head motion has also been reported to bias estimates of cortical thickness and GM compartmental volumes (Reuter et al., 2015; Alexander-Bloch et al., 2016; Pardoe et al., 2016; Savalia et al., 2017). Especially in longitudinal studies, artefacts such as head motion or changes in scanning systems across occasions (e.g. from field strength 1.5–3T) can increase measurement error and substantially reduce the sensitivity to detect change. To diminish the impact of measurement error when aligning multiple occasion imaging data in three-dimensional space, longitudinal imaging pipelines are now emerging that are designed to re-align the brain images within participants over time (e.g. Reuter et al., 2012; Yendiki et al., 2016). However, only limited evidence is available on how these processing streams perform in the case of long-term longitudinal studies (Willis et al., 2013). Simulation studies would help to shed light on the conditions under which longitudinal processing streams perform optimally or result in an underestimation of change.

Regarding the validity of the brain structure indices with regard to the underlying biological basis, all studies suffer from the same limitation that structural brain measurements from MRI are only estimates of the underlying cellular structure (see Box 1). Furthermore, most of the reviewed studies correlated individual brain measures (either local or global structural indices) with one or more cognitive measures. However, cognitive abilities are most likely based on distinct and distributed brain networks. Thus, a single regional brain measure may capture only a fraction of the variance of the underlying cognitive ability of interest. Another limitation is that single structural brain measurements are often biased by noise. One solution to reduce measurement error and thus to increase the validity of the assessed construct is the use of latent variables to capture the shared variance across multiple brain measures (Kievit et al., 2018). Using such a latent approach, it is also possible to separate more brain-wide effects of aging, which are shared among different structural brain measures from measurement-specific changes (for an example, see Lövdén et al., 2014; Ritchie et al., 2015a,b; Bender et al., 2016a).

Quality control procedures

The reporting and application of quality control procedures was very heterogeneous across the studies reviewed here. This is not surprising, as no consensus exists on standardized procedures for the control of head motion (but see: Yendiki et al., 2016), or quality of the MR images in general (see Vijayakumar et al., 2017, for a review). While there exist many metrics that quantify image quality and can be calculated with tools like QAP (Quality Assessment Protocol, Shehzad et al., 2015) or MRIQC (MRI Quality Control tool, Esteban et al., 2017), determining which metrics provide a good judgment of image quality is nontrivial. Rosen et al. (2018) have recently proposed the Euler number as a metric for the quality of surface reconstruction. Importantly for the fields of development and aging, they demonstrated in a young sample that scan quality mediated the relationship between age and cortical thickness. Furthermore, Esteban et al. (2017) proposed an automatic prediction of an image quality label. However, further work is needed in order to derive quality control standards for those measures.

Reliability and validity of cognitive ability measures

Regarding the assessment of cognitive ability, similar challenges can be discussed that threaten the reliability and validity of repeated assessments. It is well known that in longitudinal studies, practice effects can lead to the underestimation of age-related decline and thus negatively impact reliability (Hertzog and Nesselroade, 2003). One intuitive solution to deal with this problem is to use longer time intervals. However, this might not match the theoretical question at hand. In addition, Salthouse et al. (2004) showed that very long time intervals are required (between 7 and 13 years) until practice effects are no longer observable. Other possibilities are the inclusion of a new subsample at each wave to estimate the performance gains due to repeated testing, or the administration of parallel task versions. Also, estimates of long-term longitudinal change in a cognitive test can be compared to performance gains in a control condition where a subset of the same sample or a different group of individuals complete the same test repeatedly over a short time interval (Tucker-Drob and Salthouse, 2008). Furthermore, practice effects can be explicitly included in the statistical model, if the time intervals between measurements are not confounded with age (see Ferrer et al., 2004).

Regarding the validity of the cognitive measurements, the reviewed studies differ substantially with regard to the assessment of the ability of interest. Whereas some investigators tested associations between brain structure indices and single tasks, others created composites, or latent variables of multiple tasks. The advantage of using multiple tasks is that the problem of task-impurity can be lowered (Miyake and Friedman, 2012). Specifically, estimating latent factors of cognitive abilities with multiple indicators has the advantage of extracting their shared variance, which is free from task-specific measurement error. Another source of variance between studies was the choice of the type of tasks that were used as indicators of a respective cognitive domain. Due to the aforementioned problem of task-impurity, it is often not possible to clearly assign a cognitive task to a specific cognitive ability. For example, whereas one study used the digit span backwards and the letter-number sequencing task as indices for a latent factor of fluid intelligence (Ritchie et al., 2015a), another study combined the same tasks into a composite of working memory (Charlton et al., 2010). As many tasks assessing working memory resemble those measuring some aspect of fluid intelligence (Salthouse and Pink, 2008), both methods are defendable, however. Especially tasks developed in the field of neuropsychology, as were used by many of the reviewed studies, are designed to assess multiple aspects of cognitive abilities (Snyder et al., 2015). For example, the verbal fluency task (i.e. list words of a semantic category or a given starting letter), taps into verbal ability, as well as aspects of executive function (shifting between word clusters, inhibiting non-relevant words).

In general, we note that the lack of common standards for the assessment of cognitive ability in older age is a major reason for inconsistencies in the results of brain-behavior correlations.

Method development and application

To address the methodological limitations discussed above, we distinguish two different topics relevant for the advancement of future method development and application. First, the benefits of longitudinal designs and longitudinal statistical methods are discussed (see section 'The benefits of longitudinal designs'). Second, with more and more large-scale longitudinal studies emerging, it is of utmost importance to reflect on how to best handle and profit from big data (see section 'Handling and profiting from big data').

The benefits of longitudinal designs

As outlined above, the currently best way to extract information on development and correlated change over time is by relying on longitudinal studies. Only data from longitudinal research designs offer the possibility to partition within- from between-person variance and only longitudinal methods provide unbiased parameter estimates under repeated sampling. While these designs are still rare in neuroscience, an increasing number of ongoing studies are now reaching the stage in which actual longitudinal inference can be drawn. For example, Rast et al. (2017) used five waves covering 8 years from the Seattle Longitudinal Study (SLS) to characterize and identify change in cortical thickness in midlife and adulthood. While long-term longitudinal studies are not yet the norm in the field of developmental neuroscience the path in this direction is set and it is helpful to consider some of the advantages and caveats inherent in planning longitudinal studies.

Study design and power to detect change

Attention to study design (i.e. number and temporal spacing of assessments) and measurement-related issues (i.e. reliability, number of indicators, measurement modeling) are fundamental to life course and lifespan developmental research and will have direct influence on the type and quality of results obtained from a research study. Given the extensive costs associated with longitudinal research, especially when neuroimaging is involved, we need to be able to make informed decisions about our designs beforehand in order to get sufficient statistical power with minimal requirements. As such it is imperative to understand what design elements increase statistical power while keeping participant burden, sample size, and measurement occasions minimal without compromising the quality of the data.

Longitudinal studies vary in a number of elements (Lerner et al., 2009; von Oertzen and Brandmaier, 2013; Rast and Hofer, 2014) such as differences in samples (e.g. age homogeneous vs. age heterogeneous; representativeness), number of occasions (e.g. few, many), spacing between assessments (e.g. widely spaced panel designs; single session repeated testing experiment), and whether new samples of individuals are obtained at subsequent measurement occasions (e.g. sequential designs). Moreover, depending on the questions of interest, individuals can be drawn from different populations of birth cohorts, cultures, and nations, born at different historical periods, with short-term or long-term intervals between

assessments, and with measures that are time-invariant or time varying within individuals. These features can be combined in a number of ways to create study designs that are particularly suitable for answering research questions that vary in scope from regarding population change across birth cohorts to focusing on the dynamics of short-term within-person processes.

Once the type and expected temporal trajectory of the change process of interest is identified, the longitudinal study needs to be designed accordingly. Assuming that the effect sizes of the phenomenon (e.g. covariance among cortical thinning and cognitive decline) are given in the population and cannot be modified experimentally, researchers typically only have control over the duration of the study, the number and spacing of measurement occasions, and the number of participants that enter the study. These decisions can be optimized in terms of detecting individual differences in change and correlations among change processes. Rast and Hofer (2014), for example, illustrate the interplay among study duration, number of measurement occasions and interval lengths among measurements for statistical power to detect (co) variances of rates of change. Importantly, they showed how statistical power differentially reacts to changes of the study design – these changes can be exploited to optimize the study design. For example, power can be maximized if measurement occasions are spaced out unequally towards the beginning and the end of a longitudinal study. While these decisions can optimize power in general, they should not replace a tailored analysis to investigate what the exact data requirements are for a specific research question and what precautions need to be taken, especially to get the most out of studies in early stages. While Monte Carlo simulations are the most flexible tool as they can recreate the conditions, which our modeling assumption ultimately will be based on, they can also be rather complex. To mitigate the complexity of designing a power analysis and to obtain a 'feeling' of how study ingredients are interrelated, Brandmaier et al. (2015) developed the LIFESPAN tool for that specific purpose, which builds on the notion of power equivalence to analytically, and immediately, derive power for different parameter and design combinations (von Oertzen, 2010).

Statistical models

A variety of statistical approaches can be used for the estimation of change and the interaction of change processes in brain structures and cognitive abilities. From our perspective, regression models that have often been used in the past do not adequately capture the between-person variability that is present in datasets of healthy older adults. The multilevel approach and the SEM approach are well suited to model this important aspect. Both approaches contain different techniques, which focus on different modeling aspects. For example, a distinction might be drawn according to the number of dependent variables present in the model (univariate vs. multivariate). Also, models can be divided into 'static' and 'dynamic' models for change. While static models capture the change process in terms of a given functional form, such as a slope or a nonlinear trajectory, dynamic models follow the tradition of dynamical systems where the focus is on the dynamics underlying the change process itself. While the focus of multilevel approaches is mainly on the measurement model, that is, on the link between observed and latent variables, the SEM approach focuses on the structural models that links the latent constructs to additional covariates or to each other. There is no clear distinction between the two different classes of models, as they can be made equivalent for the most part, but they tend to be applied in different fields and as a result, they have approached longitudinal data from different perspectives.

Univariate multilevel models

A rather simple representation of such an approach and a common analytic method for the analysis of longitudinal data is the technique of latent growth curve modeling (i.e. random coefficient or multilevel modeling). Repeated measurement designs yield at least two levels of analysis: the Level 1 model summarizes individual level outcome data at three or more occasions in terms of 'true' initial level of performance (intercept), slope (improvement or rate of change), and error (residual) parameters. The Level 2 model estimates fixed (i.e. average), and random (i.e. individually varying) interindividual and intraindividual differences and can include predictors of individual/group differences in Level 1 parameters (i.e. intercept, slope). Detailed descriptions of these methods are available elsewhere (McArdle and Epstein, 1987; McArdle, 1988; McArdle and Hamagami, 1992; Snijders and Bosker, 1999; Verbeke and Molenberghs, 2000; Raudenbush and Bryk, 2002; Ferrer and McArdle, 2003; Singer and Willett, 2003). Conceptually, growth curve analysis involves estimating within-individual regressions of change or performance over time and on expected predictors of these individual regression parameters.

Multivariate multilevel models (MMLM)

One step towards higher complexity is to use more than one dependent variable. For the particular case of brain-behavior relationships, one automatically has multiple dependent variables of interest. In addition, within the domain of longitudinal structural brain data several levels of dependency need to be addressed in a statistical model. Measurements of GM or WM can be obtained for certain parcels nested within larger areas, nested within both hemispheres and obtained within individuals who were measured repeatedly over time. These data points are highly dependent on each other and any attempt to take this hierarchical structure into account will yield better parameter estimates in terms of biased and standard errors (e.g. Verbeke and Davidian, 2009). Multilevel or mixed-effects models are optimally suited to account for this dependency. Moreover, they make full use of the available data as they do not require that all participants have the same number of visits, or require all participants to be measured at common time points (cf., Raudenbush, 2001). To complicate matters, brain related data are also multivariate in the sense that GM or WM can be obtained from adjacent and correlated areas. To account for the multivariate nature of these data one may jointly examine the association structure in longitudinal change among different areas and composites within individuals. For example, Rast et al. (2017) modeled up to five dependent variables simultaneously over 8 years with a MMLM (MacCallum et al., 1997). In addition to estimating covariances among growth parameters (intercept and slope) the MMLM also accounts for covariances among all random effects between the different dependent variables. For example, in a model with random intercept and slope, the univariate MLM estimates the variance for both parameters (intercept variance and slope variance) as well as the covariance among the intercept and slope. In addition to these estimates, the multivariate MLM also accounts for the covariances among the dependent variables of the intercept and slope. That is, if five areas are included in the model as dependent variables, the MMLM estimates covariances for each of the random effect within and across these areas. In that case, the MMLM with random intercepts and slopes for each dependent variable estimates 45 different covariances and 10 variances. At the same time, the MMLM addresses seamlessly the issue of multiple comparisons, which typically arises in analysis of variance-type analyses. MMLMs do not necessitate pairwise comparisons and post-hoc alpha-value corrections because group mean comparisons are obtained via according coding patterns (e.g. dummy coding) that enter the model as predictors. With this approach, predictors compete for explained variance and significance tests of each predictor will be stringent. While the fit of the model to the data will increase with more parameters entering the model, the significance for each single parameter decreases. Further, multilevel techniques shrink group-level variances toward the mean, which inherently reduces the number of statistically significant comparisons, thus reducing the risk of type I errors in multiple comparisons (cf., Gelman et al., 2012). These classes of models can be expanded to address nonlinear trajectories or non-Gaussian processes. Moreover, they can be expanded to include submodels for the within-person variance structure. This is especially useful if within-person variability is the focus.

A note on the metric of change

In these models, a level and a slope parameter are generally specified for change relative to a particular time metric. The selection of the time metric is not trivial as it results in different models that reflect different assumptions about the underlying process of change. A common choice is to define the level as the initial point of measurement in a longitudinal data set whereas the slope parameter captures the rate of change over time in study. Especially in models with higher order terms or interactions (including all models with predictors of slope variance) the choice of the centering method influences the interpretation of the parameters (cf., Biesanz et al., 2004). The intercept or initial level should be carefully chosen to reflect the hypotheses tested in this context, especially when time-varying covariates are used as predictors of change (e.g. Curran and Bauer, 2011; Hamaker and Grasman, 2015; Wang and Maxwell, 2015).

While age heterogeneity of the sample is not a problem for growth models in particular, it is necessary to consider it in all cross-sectional or longitudinal models. Unlike traditional single-cohort longitudinal designs, individuals may vary considerably in age (and birth cohort membership) at each wave in the study, and the range of these between-person, cross-sectional, age differences tends to exceed the range of within-person, longitudinal, age changes over the course of data collection. A common technique to obtain 'longitudinal' data from studies with few (e.g. two) measurement occasions, is to combine the longitudinal with the cross-sectional age information by indexing change via the age of the study participants (e.g. Grimm et al., 2012). These age-heterogeneous samples were seen as an opportunity to virtually 'accelerate' longitudinal designs (e.g. McArdle and Bell, 2000; Mehta

and West, 2000) by representing time as the different ages available in the study, rather than as time as the inception of the study. While this enables one to model growth trajectories that exceed the individual time-in-study span over the full age range they also bear the risk of confounding between-person differences with within-person change. That is, without the continued inclusion of baseline age differences, the resulting model produces estimates that represent a mix of the cross-sectional and longitudinal effects.

For models using chronological age as the time basis with a focus on random effects, it is important to keep in mind the variance component shrinkage due to the extrapolation beyond individual data (i.e. random effects will be estimated closer to the population mean) (e.g. Raudenbush and Bryk, 2002, Ch. 5). With greater age heterogeneity than study duration, the population mean will be dominated by the cross-sectional information. In addition, the confounding of between-person age differences and within-person age changes in longitudinal models muddles the potential for inference to increasingly selective, and thus conditional, 'aging' populations. Selectivity of participants must be accounted for as between-person sampling will be based on the proportion of the population who are alive (population mortality selection) and healthy but will also dropout from the study due to health and mortality causes. Such inference to aging populations must, therefore, be conditional on survival, and may be more directly obtained using between-person age differences and survival age (or time-to-death) as conditional predictors in a time-in-study longitudinal model (e.g. Johansson et al., 2004; Hoffmann, 2012).

Latent change models

Multilevel models can also be specified in the SEM framework, which has the advantage of incorporating additional models that operate at the latent level. For instance, instead of averaging dependent measures from several cognitive tasks into one composite score, a latent factor for the respective cognitive ability can be specified, thus estimating the shared variance among tasks while attenuating the effect of error variance. This also means that SEMs can be expanded to higher orders to include, for example, mechanisms of change. This can be in the form of latent growth models that define a linear (or nonlinear) model for lower order factors or in the form of LCS models that define the observed change from an occasion to the next as a sum of higher order factors. Moreover, SEMs easily lend themselves to measurement invariance (MI) testing. By constraining different elements of the measurement model, MI defines conditions under which meaningful comparisons among groups or within individuals across time can be drawn.

While most models discussed so far serve the field well, they might be considered static models because they define change resulting from static element such as the slope in a linear model. For example, Figure 5 represents a simple univariate latent growth model for five time points in the typical structural equation path diagram notation. The squares represent manifest variables (x1 to x5) and the circles represent latent, unobserved variables, the intercept (I) and the slope (S). Double-headed arrows denote undirected relationships such as covariances and single-headed arrows represent directed relationships such as regression weights, or loadings. Both, the intercept and slope have associated variances that capture the individual differences therein. Moreover, the intercept and slope are allowed to covary.

Another approach is to capture the change trajectory through models that originate in dynamical systems where the current state of the system is defined by previous states of the system (e.g. Boker and Wenger, 2007). Notably, the LCS (McArdle and Hamagami, 2001; McArdle, 2009) model and its multivariate extension, the bivariate LCS have been applied successfully in recent years to longitudinal data. The LCS addresses change, from one time point to the next, from a SEM standpoint. That is, while one could index change from one time point t–1 to t on the observed data, the LCS indexes these changes at the latent level, thus separating measurement error from the true score. Once the differences among time points are

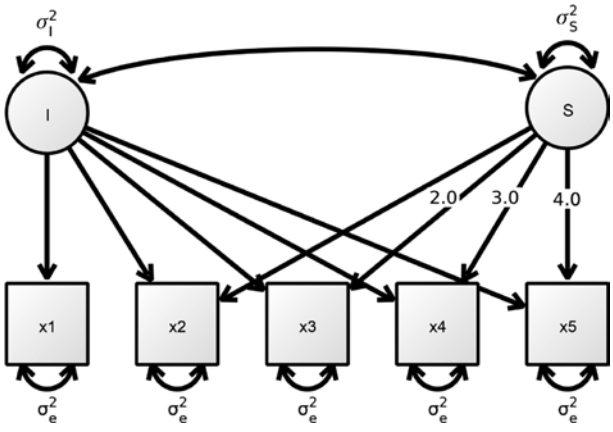

Figure 5: A univariate latent growth curve model. Circles represent latent variables while squares represent manifest variables. One headed arrows denote directed relationships and double headed arrows represent undirected relationships. Here, I is the latent intercept and S is the latent slope, each with their corresponding variances σ_I^2 and σ_S^2.

defined, the LCS focuses on the rate of change, rather than latent change itself. Figure 6 represents a basic LCS for one common factor over five time points. The manifest variables (x1 to x5) are again symbolized by squares. η1–η5 represent latent true scores, Δη0–Δη4 are the LCS and I and S define the latent intercept and slope, or rate of change. Unlabeled paths are fixed to 1. Here, the first change score is defined as Δη0. The change does not affect the prior score η1 but it does influence the second true score directly and it is an indirect part of all the other latent variables. The same holds for the following change scores. α paths represent constant change and β paths represent proportional change from the variable measured at the previous time point. The rate of change is passed into the latent change scores, typically with a constant weight of 1 for equally spaced time intervals. Essentially, the rate of change is defined as the difference between two latent variables divided by the length of the given discrete time interval among them. A constant rate of change would be nothing else than the first partial derivative of a linear function with respect to time–but with nonlinear changes, the rate of change will take different values for different measurement occasions.

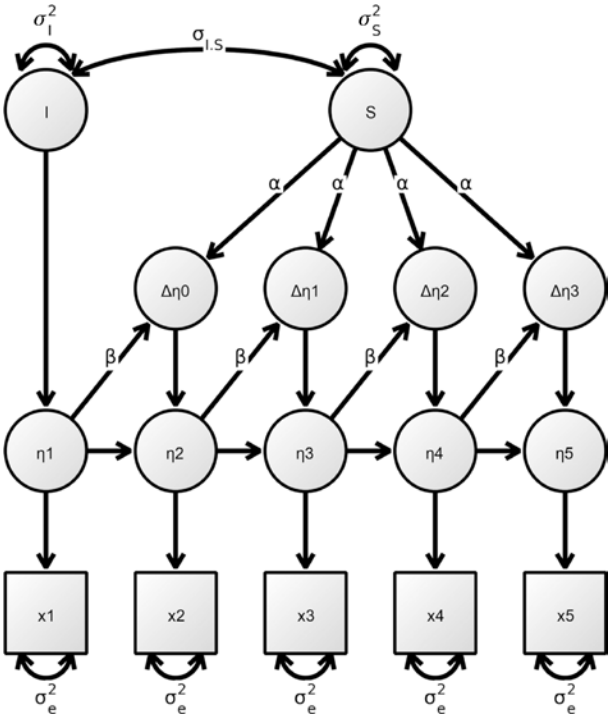

Figure 6: A univariate latent change score model. x1–x5 represent the observed variable measured at five time points, η1–η5 represent latent true scores, Δη0–Δη4 are the latent change scores and I and S define the latent intercept and slope α paths represent constant change and β paths represent proportional change from the variable measured at the previous time point.

While the constraint for discrete time seems rather restrictive, Voelkle et al. (2012) expanded the LCS to a continuous time model where this constraint is resolved. Importantly, LCS and continuous time models (and variants thereof) can be readily expanded to the multivariate case which makes them ideal candidates for modeling correlated changes in brain related and cognitive longitudinal data. Moreover, given that these models do not impose a functional form of change, they are very flexible and can handle a multitude of curvilinear trajectories in one or more processes and relate changes in one variable to changes in another. While the adoption of latent change models is slow, likely due to a rather high level technical sophistication in order to implement these models, new software developments now facilitate the use of LCS models (see Kievit et al., 2018 for a tutorial). For example, to depict the models in Figures 5 and 6, we used Ωnyx (von Oertzen et al., 2015), a freely and openly accessible graphical tool that provides an easy and intuitive approach to depicting SEM-based models.

Advanced models

Besides the models discussed above, other methodological advances have recently emerged and/or are being developed that are of particular relevance for aging and lifespan studies. For example, growth mixture models permit the identification of subpopulations that exhibit distinct multivariate patterns of change and are therefore well suited for exploratory analyses. Similar to latent class models, growth mixture models (e.g. Muthén, 2001) assume that the sample is composed of members from more than one population that exhibit distinct patterns of change. Using individual response patterns in a longitudinal setting with repeated measurements to define trajectories, growth mixture models (1) identify homogeneous groups of individuals or trajectory classes, (2) assign each participant a probability of belonging to a particular trajectory class, and (3) use class membership information to estimate the influence of individual characteristics on trajectory shape. A related, but less familiar method are SEM trees (Brandmaier and McArdle, 2013), which split the data into homogeneous subgroups based on a set of predictors and fit an SEM separately to each of these subgroups (see Jacobucci et al., 2017, for a comparison of the two methods). In contrast with other SEM-based statistical approaches, SEM trees are designed as an exploratory method. Especially in light of the many variables that influence aging, such approaches of data reduction gain more and more importance.

Ultimately, most of the above-discussed models can also be estimated in a Bayesian framework. The Bayesian

approach to inferential statistics has the advantage that it can overcome the typical problems surrounding the p-value in null hypothesis testing (e.g. corrections for multiple comparisons), and allows to incorporate previous knowledge about the expected effect into the analysis (Kruschke, 2010). For example, Bayes Factors provide an estimate of the strength of evidence both for the null hypothesis (i.e. absence of the effect), and the alternative hypothesis (i.e. presence of the effect), necessitating researchers to think about what they would actually consider plausible values for the alternative hypothesis (Dienes, 2014). Thus, using Bayesian estimation, it is possible to test specific theoretical predictions, and gain an estimate of the strength of the evidence for the presence or absence of the predicted effect. Recent software developments have provided novel tools to efficiently estimate Bayesian models of change both in the MLM (Bürkner, 2017) and the SEM framework (Asparouhov and Muthén, 2010; Merkle and Rosseel, 2016).

Overall, a wide range of longitudinal modeling tools now exist or are emerging that can capture multivariate change and correlated change with models that are able to accommodate a wide range of developmental questions in brain and cognitive aging.

Handling and profiting from big data

Making science more reproducible is a growing concern in cognitive neuroscience and beyond (Munafò et al., 2017). This can be achieved by making science more transparent, for instance, by openly sharing data sets and analysis tools. Additionally, the need for well powered brain imaging studies (Button et al., 2013; Nord et al., 2017) resulted in steadily growing sample sizes over the last two decades (Poldrack et al., 2017). Early data sharing initiatives pooled data from multiple sites to increase sample size (for instance the 1000 Functional Connectomes Project; Biswal et al., 2010).

More recently, data from large-scale projects, investigating hundreds to thousands of participants, have been made public. The mode of accessing the data varies from download without registration, to signing a data usage agreement to submitting a project proposal. Open brain-behavior data sets investigating the adult lifespan in a cross-sectional approach and providing a large variety of brain (functional, structural), and cognitive data include the Nathan Kline Institute-Rockland Sample (Nooner et al., 2012), the Cambridge Centre for Ageing Neuroscience study (Cam-CAN) (Shafto et al., 2014; Taylor et al., 2017), the UK Biobank (Palmer, 2007), and the Harvard Aging Brain Study (HABS) (Dagley et al., 2017). Some of those projects aim to also provide longitudinal data in the future (Cam-CAN and HABS). Additionally, the dementia-focused projects Open Access Series of Imaging Studies (Marcus et al., 2010) and the ADNI (Petersen et al., 2010) currently provide longitudinal anatomical brain data. Another effort to combine several lifespan cognitive neuroscience samples has recently been launched, the Lifebrain project (Lifebrain, n.d.).

Open data enables researchers to validate their results in independent data sets. However, the heterogeneity in data organization between projects can make it tedious to apply a processing pipeline to a new data set. The recently introduced brain imaging data structure (BIDS) initiative proposes a system to harmonize data organization and provides guidelines for the documentation of important imaging meta-data (Gorgolewski et al., 2016).

As analysis pipelines are complex and cannot be fully described in the text of a scientific paper, many researchers make analysis code repositories publicly available on websites like GitHub (GitHub, n.d.) or the Open Science Framework (Open Science Framework, n.d.). A more recent trend in simplifying software (re)use is to provide executable code as software containers, making the installation of dependencies obsolete and facilitating archiving of entire software environments for later (re) analysis. This approach is championed by the BIDS Apps project (Gorgolewski et al., 2017), which provides neuroimaging analysis pipelines that seamlessly can be applied to BIDS-formatted data sets, making it very efficient to run (1) established pipelines on newly acquired data, and (2) newly developed pipelines on available data sets.

As a consequence of increasing sample size and the increased availability of high performance computing resources, data processing is moving from local computers to clusters and cloud systems (Sherif et al., 2014; Vogelstein et al., 2016; Kiar et al., 2017). For instance, this can be in the form of OpenNeuro (OpenNeuro, n.d.), a neuroimaging analysis service that allows scientists to upload their raw data to a server, which executes standard analysis pipelines. Another innovative approach for collaborative neuroscience is followed by the Open Neuroimaging Laboratory (Open Neuroimaging Laboratory, n.d.), which allows scientists to conjointly work on publicly available data via the web.

Theoretical limitations and the need for theory development

The study of the relationship between the structure of the brain and observable cognitive performance is tapping

into one of the oldest problems of psychological science and philosophy: the relation of the mind and body. While initial conceptions proposed a dualistic theory of mind and body as separate and independent entities (Descartes, 1641/2013), modern-day approaches are moving towards more holistic theories (e.g. the principle of complementarity; Fahrenberg, 1979). Yet, still no final consensus has been reached on how to integrate these different levels of observation theoretically and methodologically. Generating and developing theories is a major challenge for researchers in this field, and it does not come as a surprise that the studies we reviewed here did not build their hypotheses on strong theoretical grounds. As a consequence, we believe that future efforts are needed to invest in the development of theories on the intersection between brain and behavior. Here, we outline several theoretical ideas and trends that we deem important in the context of studying correlated changes between brain structure and cognitive ability.

Refining and developing existing theories

While the STAC-r model provides a multifactorial framework for the complexity of cognitive aging (Reuter-Lorenz and Park, 2014), more specific theories are needed to derive concrete, testable hypotheses for individual components within this larger framework. Several well-established theories already exist that serve this function. For example, the theories of reserve propose that individuals differ with regard to their neural (brain reserve) or cognitive resources (cognitive reserve), such that some individuals are better able to compensate for age-related brain changes than others (Stern, 2002, 2009). Comparable to compensatory scaffolding in the STAC-r model, reserve is assumed to be malleable by life course experiences (Reuter-Lorenz and Park, 2014). Empirically, however, it is still a matter of debate how to best operationalize compensatory scaffolding and reserve capacity (Nilsson and Lövdén, 2018). For example, some authors assess cognitive reserve with enriching life course experiences (leisure time activities: Hertzog et al., 2009; education: Boots et al., 2015; occupational complexity: Serra et al., 2015) and others via levels of cognitive ability (e.g. intelligence: Barulli et al., 2013). We therefore argue that future efforts of theory development need to be directed towards refining and consolidating already existing theories, and towards developing a common consensus of how to operationalize the theoretical core mechanisms of interest and their interactions (see Cabeza et al., 2018; Stern et al., 2018 for promising advances in this direction). Besides working towards more consistent operationalizing, the conceptual integration of advantageous, and adverse life course experiences in theories on age-related structural brain and cognitive changes presents an important challenge in the near future (e.g. Köhncke et al., 2016).

One potentially fruitful approach to inform and complement established theories is the use of predictive methods (Yarkoni and Westfall, 2017; Bzdok and Ioannidis, 2019). The goal of predictive methods is to make use of large data sets and to identify patterns therein that most accurately predict individual behavior. For instance, machine learning approaches are used to detect those variables that best predict a relevant outcome in a subset of a sample, a process which then needs to be cross-validated in another subset of the sample (or in a new sample), to ultimately identify the algorithm that most accurately approximates an outcome of interest (e.g. change in cognitive abilities). The relative importance of individual predictors can be evaluated by comparing the predictive accuracy of different models with and without the predictors of interest (Yarkoni and Westfall, 2017). In the context of investigating correlated changes between brain structure and cognitive ability, such an approach may be helpful to determine the relative importance of changes in a number of structural brain measures (e.g. WM and GM volume, cortical thickness, WMH etc.) for the prediction of changes in certain cognitive abilities. Finally, traditional theory-guided statistical approaches can benefit from cross-validating models in independent datasets as it is usually done in predictive modeling (Yarkoni and Westfall, 2017).

Correlated change relationships across the lifespan

So far, studies on correlated change have focused more on the direct relation between brain structure and cognitive ability (direct path in STAC-r), ignoring the role of compensatory scaffolding networks as a moderator of brain structure-cognition relations (indirect path in STAC-r; see Figure 2). Related to testing the latter, future research would be welcome to further elaborate how correlated changes between structural brain and cognitive abilities change across the lifespan. A first step into this direction would be to investigate whether specific structural brain measures and cognitive abilities show increased (de)coupling over the adult lifespan, such that the strength of correlated changes between brain structure and cognitive ability either varies between persons of different ages or changes within persons with increasing age. According

to STAC-r, correlated changes between brain structure and cognitive ability should be higher in younger than (healthy) older adults, as scaffolding networks should gain more importance with increasing age, when brain damage starts to accumulate. Moreover, after a certain advanced age, when compensatory scaffolding networks are no longer functional to compensate for brain damage, change correlations might increase again. According to this rationale, larger brain structure-cognition correlations should also be found in individuals with preclinical pathological brain changes, who are most likely included in many aging samples. However, it is still unclear how much pathology can be accumulated before observable detrimental brain changes manifest in cognitive performance measures. So far, the hypothesis of developmental (de)coupling of brain structure and cognitive ability with aging has only been directly tested in cross-sectional studies (e.g. de Mooij et al., 2018).

Insight from imaging brain function

Another fruitful avenue for future research is the application of findings and theories derived from functional MRI studies to structural brain measures. The functional MRI literature has paid more tribute to the indirect path of the STAC-r model. For example, functional evidence found that older adults show different patterns of functional brain activation than younger adults when faced with a difficult cognitive task (Reuter-Lorenz and Park, 2010), suggestive of a compensatory reorganization of functional brain networks (comparable to scaffolding; but see Morcom and Johnson, 2015).

One prominent theory in the functional literature interprets these activation patterns as a sign of neural dedifferentiation in the sense that older adults show lower neural specificity than younger adults when performing distinct cognitive tasks (e.g. Cabeza, 2002; Park et al., 2004). This theory originates from behavioral findings relating the deterioration of fluid cognitive abilities with aging to sensory declines, which is interpreted as evidence for a common cause of cognitive aging (Baltes and Lindenberger, 1997). Only little research has, however, investigated whether this loss in neural specificity with aging can also be found in structural brain measures, and the results are difficult to reconcile: Cox et al. (2016) showed that single WM fiber pathways were more correlated with increasing age, suggesting similar age-related dedifferentiation as reported in functional MRI studies. In contrast, de Mooij et al. (2018) reported the opposite pattern of lower correlations between regional GM volumes and FA in a subset of WM pathways with increasing age, indicating age-related differentiation. Importantly, however, there is a lack of research studying neural (de)differentiation longitudinally.

Interdisciplinary contributions to theory advancement

To develop novel ideas or to refine existing theories that capture the manifold factors influencing brain and cognitive aging, interdisciplinary collaborations are gaining more and more importance. This pertains to closely related research fields, such as child and youth development, as well as to more distant disciplines. Given the multitude of collaboration opportunities, we restrict ourselves to providing some examples to substantiate our claim.

Aging research, for example, can clearly benefit from insights made in other domains of life span research (e.g. Goddings et al., 2014; Mills et al., 2014, 2016) given that theoretical approaches and methodology is partly overlapping. Moreover, recent studies suggest a link between early life influences, such as birth weight, parental education, or childhood cognitive ability, on cognitive and brain developmental processes in older adulthood (Karama et al., 2014; Walhovd et al., 2016). Also, the liaison with medicine seems very promising. For example, by merging epidemiological research with healthcare databases, a wide array of health-related information can be obtained and fed into models and analyses. The epidemiological UK Biobank project, which combines questionnaire, cognitive and neuroimaging data from 500 000 participants with biological samples (e.g. blood, saliva) and genomic data (e.g. genotyping) (Miller et al., 2016), follows this approach. By linking the newly collected data with participant's health records from the UK National Health Service it allows the long-term monitoring of the participants' health state. Most importantly, big, interdisciplinary databases, such as the UK Biobank, might enable researchers to detect biomarkers which can serve as early predictors of future pathologies. Future efforts will need to be increasingly directed towards determining how the resulting multimodal data can be meaningfully aggregated across multiple levels of analysis (e.g. from genes, to cognitive performance and brain properties, to the larger societal context) (Falk et al., 2013).

Third, aging research is benefiting from technological advances in various disciplines, such as geoinformatics or computer science in general. While life course experiences, for example, are traditionally assessed via self-report, new mobile technologies facilitate ambulatory

assessment. With small electronic devices, people's behaviors (e.g. physical activity, social engagement, mobility), or physiology (e.g. blood pressure, electrodermal activity) can be sampled with high density and in real-time as people go about their life (Conner and Mehl, 2015) and, in a further step, linked to processes of cognitive and brain aging (e.g. Seresinhe et al., 2015). Besides person-specific variables, also information about the broader contextual situation (e.g. weather conditions, air pollution, etc.) can be recorded via sensor technology and integrated into theoretical models. While many of these tools and devices are just emerging and still awaiting further validation, these developments hold great promise to gain new, ecologically valid insights into the daily processes influencing brain and cognitive aging.

Conclusion and outlook

In the present article, we have provided a broad overview over the literature on the association of the neural architecture and cognitive abilities in healthy old age. Specifically, we reviewed the existing longitudinal studies that investigated correlated changes between these domains over time, and discussed the present stance of the literature from a theoretical perspective, adopting the STAC-r model (Reuter-Lorenz and Park, 2014). While the overall evidence suggests a trend towards positive change-associations between measures of brain structure and cognitive ability in healthy aging, the number of longitudinal studies reviewed here is small, and the variability between them regarding the methods used (e.g. study design, statistical analysis) precludes meta-analytical comparisons of effect sizes. While some evidence supports correlated changes for specific regions and specific cognitive abilities (e.g. between structures of the medial temporal lobe and episodic memory), the number of studies reporting converging results is considerably small, and most of the reported change relations are very heterogeneous and far from conclusive. One reason for these mixed findings is certainly that large methodological differences exist between studies. However, we argue that from the perspective of an individual regulatory model such as STAC-r, weak relationships between brain and cognitive ability, and large interindividual variability in these relations are to be expected in healthy aging individuals, since many other influencing factors play a role that have an impact on the capacity of the brain to compensate. The inclusion of such factors into future research will impose a challenge, but also a chance to advance the study of aging. Moreover, it would be even more interesting for future research to focus on alternative metrics such as change trajectories in the strength of brain-behavior correlations throughout adult life. To capture such dynamic processes such as in the case of brain cognitive development, many more measurement occasions and complex statistical models are needed. Current longitudinal studies, such as the SLS, or the longitudinal healthy aging brain (LHAB) study are moving towards this direction, by including multiple repeated assessments of cognitive and MRI assessments (SLS; Schaie, 1996; Schaie and Willis, 2010; LHAB; Zöllig et al., 2011). In addition, in the moment of writing this review, further measurement occasions in these and other large-scale longitudinal studies are under way. The development towards open science and big data sharing will help to cover a broad spectrum of variables influencing structural brain and cognitive aging and will in the future, facilitate the development of individually targeted interventions to promote health and well-being.

Acknowledgments: This work was supported by the University Research Priority Program 'Dynamics of Healthy Aging.' Research reported in this publication was also in part supported by the National Institute on Aging of the National Institutes of Health under Award Number Funder Id: http://dx.doi.org/10.13039/100000049, R01AG050720 to P. Rast. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. During the work on their dissertation, J. Oschwald and S. Guye were pre-doctoral fellows 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).

Conflict of interest statement: All authors declare no conflicts of interest.

References

Adólfsdóttir, S., Wollschlaeger, D., Wehling, E., and Lundervold, A.J. (2017). Inhibition and switching in healthy aging: a longitudinal study. J. Int. Neuropsychol. Soc. *23*, 90–97.

Alexander-Bloch, A., Clasen, L., Stockman, M., Ronan, L., Lalonde, F., Giedd, J., and Raznahan, A. (2016). Subtle in-scanner motion biases automated measurement of brain anatomy from *in vivo* MRI. Hum. Brain Mapp. *37*, 2385–2397.

Anblagan, D., Valdés Hernández, M.C., Ritchie, S.J., Aribisala, B.S., Royle, N.A., Hamilton, I.F., Cox, S.R., Gow, A.J., Pattie, A., Corley, J., et al. (2018). Coupled changes in hippocampal

structure and cognitive ability in later life. Brain Behav. *8*, e00838.

Arshad, M., Stanley, J.A., and Raz, N. (2016). Adult age differences in subcortical myelin content are consistent with protracted myelination and unrelated to diffusion tensor imaging indices. Neuroimage *143*, 26–39.

Ashburner, J. and Friston, K.J. (2000). Voxel-based morphometry – the methods. Neuroimage *11*, 805–821.

Asparouhov, T. and Muthén, B. (2010). Bayesian analysis of latent variables models using Mplus. Retrieved from https://www.statmodel.com/download/BayesAdvantages18.pdf.

Baddeley, A. (1998). Working memory. Comptes Rendus l' Acad. Des Sci. Ser. III Sci. La Vie *321*, 167–173.

Baddeley, A., Eysenck, M.W., and Anderson, M.C. (2009). Memory (Hove, UK: Psychology Press).

Baltes, P.B. (1987). Theoretical propositions of life-span developmental psychology: on the dynamics between growth and decline. Dev. Psychol. *23*, 611–626.

Baltes, P.B. and Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? Psychol. Aging *12*, 12–21.

Baltes, P.B. and Smith, J.S. (2003). New frontiers in the future of aging: from successful aging of the young old to the dilemmas of the fourth age. Gerontology *49*, 123–135.

Barisano, G., Sepehrband, F., Ma, S., Jann, K., Cabeen, R., Wang, D.J., Toga, A.W., and Law, M. (2019). Clinical 7T MRI: are we there yet? A review about magnetic resonance imaging at ultra-high field. Br. J. Radiol. *92*, 20180492.

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.

Bartzokis, G., Beckson, M., Lu, P.H., Nuechterlein, K.H., Edwards, N., and Mintz, J. (2001). Age-related changes in frontal and temporal lobe volumes in men: a magnetic resonance imaging study. Arch. Gen. Psychiatry *58*, 461–465.

Barulli, D. and Stern, Y. (2013). Emerging concepts in cognitive reserve. Trends Cogn. Sci. *17*, 502–509.

Barulli, D.J., Rakitin, B.C., Lemaire, P., and Stern, Y. (2013). The influence of cognitive reserve on strategy selection in normal aging. J. Int. Neuropsychol. Soc. *19*, 841–844.

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system – a technical review. NMR Biomed. *15*, 435–455.

Bender, A.R., Prindle, J.J., Brandmaier, A.M., and Raz, N. (2016a). White matter and memory in healthy adults: coupled changes over 2 years. Neuroimage *131*, 193–204.

Bender, A.R., Völkle, M.C., and Raz, N. (2016b). Differential aging of cerebral white matter in middle-aged and older adults: a 7-year follow-up. Neuroimage *125*, 74–83.

Bennett, I.J. and Madden, D.J. (2014). Disconnected aging: cerebral white matter integrity and age-related differences in cognition. Neuroscience *276*, 187–205.

Berardi, A.M., Parasuraman, R., and Haxby, J.V. (2001). Overall vigilance and sustained attention decrement in healthy aging. Exp. Aging Res. *27*, 19–39.

Biesanz, J.C., Deeb-Sossa, N., Papadakis, A.A., Bollen, K.A., and Curran, P.J. (2004). The role of coding time in estimating and interpreting growth curve models. Psychol. Methods *9*, 30–52.

Birdsill, A.C., Koscik, R.L., Jonaitis, E.M., Johnson, S.C., Okonkwo, O.C., Hermann, B.P., LaRue, A., Sager, M.A., and Bendlin, B.B. (2014). Regional white matter hyperintensities: aging, Alzheimer's disease risk, and cognitive function. Neurobiol. Aging *35*, 769–776.

Biswal, B., Mennes, M., Zuo, X., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Bendlin, B.B. (2010). Toward discovery science of human brain function. Proc. Natl Acad. Sci. U.S.A. *107*, 4734–4739.

Boker, S.M. and Martin, M. (2018). A conversation between theory, methods, and data. Multivariate Behav. Res. *53*, 806–819.

Boker, S.M. and Wenger, M.J., eds. (2007). Data Analytic Techniques for Dynamical Systems (Notre Dame Series on Quantitative Methodology) (Mahwah, NJ: Lawrence Erlbaum Associates).

Boots, E.A., Schultz, S.A., Almeida, R.P., Oh, J.M., Koscik, R.L., Dowling, M.N., Gallagher, C.L., Carlsson, C.M., Rowley, H.A., Bendlin, B.B., et al. (2015). Occupational complexity and cognitive reserve in a middle-aged cohort at risk for Alzheimer's disease. Arch. Clin. Neuropsychol. *30*, 634–642.

Brandmaier, A.M. and McArdle, J.J. (2013). Structural equation model trees. Psychol. Methods *18*, 71–86.

Brandmaier, A.M., von Oertzen, T., Ghisletta, P., Hertzog, C., and Lindenberger, U. (2015). LIFESPAN: a tool for the computer-aided design of longitudinal studies. Front. Psychol. *6*, 272.

Brehmer, Y., Kalpouzos, G., Wenger, E., and Lövdén, M. (2014). Plasticity of brain and cognition in older adults. Psychol. Res. *78*, 790–802.

Brink, J.M. and McDowd, J.M. (1999). Aging and selective attention: an issue of complexity or multiple mechanisms? J. Gerontol. Ser. B Psychol. Sci. Soc. Sci. *54*, 30–33.

Brockmole, J.R. and Logie, R.H. (2013). Age-related change in visual working memory: a study of 55,753 participants aged 8–75. Front. Psychol. *4*, 12.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. *10*, 186–198.

Burgess, P.W. (1997). Theory and methodology in executive function research. In: Theory and Methodology of Frontal and Executive Function. P. Rabbitt, ed. (Hove, UK: Psychology Press). pp. 81–116.

Bürkner, P.-C. (2017). Advanced bayesian multilevel modeling with the R Package brms. Retrieved from http://arxiv.org/abs/1705.11123.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., and Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. *14*, 365–376.

Bzdok, D. and Ioannidis, J.P.A. (2019). Exploration, inference, and prediction in neuroscience and biomedicine. Trends Neurosci. *42*, 251–262.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. Psychol. Aging *17*, 85–100.

Cabeza, R., Albert, M., Belleville, S., Craik, F.I.M., Duarte, A., Grady, C.L., Lindenberger, U., Nyberg, L., Park, D.C., Reuter-Lorenz, P.A., et al. (2018). Maintenance, reserve and compensation: the cognitive neuroscience of healthy ageing. Nat. Rev. Neurosci. *19*, 701–710.

Cao, X., Yao, Y., Li, T., Cheng, Y., Feng, W., Shen, Y., Li, Q., Jiang, L., Wu, W., Wang, J., et al. (2016). The impact of cognitive training on cerebral white matter in community-dwelling elderly: 1-year prospective longitudinal diffusion tensor imaging study. Sci. Rep. *6*, 33212.

Carmichael, O.T., Kuller, L.H., Lopez, O.L., Thompson, P.M., Dutton, R.A., Lu, A., Lee, S.E., Lee, J.Y., Aizenstein, H.J., Meltzer, C.C., et al. (2007). Acceleration of cerebral ventricular expansion in the Cardiovascular Health Study. Neurobiol. Aging *28*, 1316–1321.

Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: acritical experiment. J. Educ. Psychol. *54*, 1–22.

Cattell, R.B. (1988). The data box. In: Handbook of Multivariate Experimental Psychology. Perspectives on Individual Differences. J.R. Nesselroade and R.B. Cattell, eds. (Boston, MA: Springer). pp. 69–130.

Cepeda, N.J., Blackwell, K.A., and Munakata, Y. (2013). Speed isn't everything: complex processing speed measures mask individual differences and developmental changes in executive control. Dev. Sci. *16*, 269–286.

Chapman, S.B., Aslan, S., Spence, J.S., Hart, J.J., Bartz, E.K., Didehbani, N., Keebler, M.W., Gardner, C.M., Strain, J.F., DeFina, L.F., et al. (2015). Neural mechanisms of brain plasticity with complex cognitive training in healthy seniors. Cereb. Cortex *25*, 396–405.

Charlton, R.A., Schiavone, F., Barrick, T.R., Morris, R.G., and Markus, H.S. (2010). Diffusion tensor imaging detects age related white matter change over a 2 year follow-up which is associated with working memory decline. J. Neurol. Neurosurg. Psychiatry *81*, 13–19.

Cohen, R.M., Small, C., Lalonde, F., Friz, J., and Sunderland, T. (2001). Effect of apolipoprotein E genotype on hippocampal volume loss in aging healthy women. Neurology *57*, 2223–2228.

Cole, J.H., Ritchie, S.J., Bastin, M.E., Valdés Hernández, M.C., Muñoz Maniega, S., Royle, N., Corley, J., Pattie, A., Harris, S.E., Zhang, Q., et al. (2018). Brain age predicts mortality. Mol. Psychiatry *23*, 1385–1392.

Concha, L. (2014). A macroscopic view of microstructure: using diffusion-weighted images to infer damage, repair, and plasticity of white matter. Neuroscience *276*, 14–28.

Conner, T.S. and Mehl, M.R. (2015). Ambulatory assessment: methods for studying everyday life. In: Emerging Trends in the Social and Behavioral Sciences. R.A. Scott and S.M. Kosslyn, eds. (Hoboken, NJ: Wiley). pp. 1–15.

Cowan, N. (1995). Attention and Memory : An Integrated Framework (Oxford, UK: Oxford University Press).

Cox, S.R., Ritchie, S.J., Tucker-Drob, E.M., Liewald, D.C., Hagenaars, S.P., Davies, G., Wardlaw, J.M., Gale, C.R., Bastin, M.E., and Deary, I.J. (2016). Ageing and brain white matter structure in 3513 UK Biobank participants. Nat. Commun. *7*, 13629.

Craik, F.I.M. and McDowd, J.M. (1987). Age differences in recall and recognition. J. Exp. Psychol. Learn. Mem. Cogn. *13*, 474–479.

Curran, P.J. and Bauer, D.J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. Annu. Rev. Psychol. *62*, 583–619.

Dagley, A., LaPoint, M., Huijbers, W., Hedden, T., McLaren, D.G., Chatwal, J.P., Papp, K.V., Amariglio, R.E., Blacker, D., Rentz, D.M., et al. (2017). Harvard aging brain study: dataset and accessibility. Neuroimage *144*, 255–258.

Daneman, M. and Carpenter, P.A. (1980). Individual differences in working memory and reading. J. Verbal Learning Verbal Behav. *19*, 450–466.

Daugherty, A.M., Haacke, E.M., and Raz, N. (2015). Striatal iron content predicts its shrinkage and changes in verbal working memory after 2 years in healthy adults. J. Neurosci. *35*, 6731–6743.

de Lange, A.-M.G., Bråthen, A.C.S., Rohani, D.A., Grydeland, H., Fjell, A.M., and Walhovd, K.B. (2017). The effects of memory training on behavioral and microstructural plasticity in young and older adults. Hum. Brain Mapp. *38*, 5666–5680.

de Lange, A.-M.G., Cecilie, A., Bråthen, S., Rohani, D.A., Fjell, A.M., and Walhovd, K.B. (2018). The temporal dynamics of brain plasticity in aging. Cereb. Cortex *28*, 1857–1865.

de Mooij, S.M.M., Henson, R.N.A., Waldorp, L.J., and Kievit, R.A. (2018). Age differentiation within grey matter, white matter and between memory and white matter in an adult lifespan cohort. J. Neurosci. *38*, 5826–5836.

Deary, I.J. (2001). Human intelligence differences: a recent history. Trends Cogn. Sci. *5*, 127–130.

DeCarli, C., Massaro, J., Harvey, D., Hald, J., Tullberg, M., Au, R., Beiser, A., D'Agostino, R., and Wolf, P.A. (2005). Measures of brain morphology and infarction in the framingham heart study: establishing what is normal. Neurobiol. Aging *26*, 491–510.

Descartes, R. (2013). Meditations on first philosophy. In: Meditations on First Philosophy: With Selections from the Objections and Replies. J. Cottingham, ed. (Cambridge: Cambridge University Press).

Dickerson, B.C., Fenstermacher, E., Salat, D.H., Wolk, D.A., Maguire, R.P., Desikan, R., Pacheco, J., Quinn, B.T., Van der Kouwe, A., Greve, D.N., et al. (2008). Detection of cortical thickness correlates of cognitive performance: reliability across MRI scan sessions, scanners, and field strengths. Neuroimage *39*, 10–18.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. Front. Psychol. *5*, 781.

Dormann, C. and Griffin, M.A. (2015). Optimal time lags in panel studies. Psychol. Methods *20*, 489–505.

Drag, L.L. and Bieliauskas, L.A. (2010). Contemporary review 2009: cognitive aging. J. Geriatr. Psychiatry Neurol. *23*, 75–93.

Draganski, B., Ashburner, J., Hutton, C., Kherif, F., Frackowiak, R.S.J., Helms, G., and Weiskopf, N. (2011). Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). Neuroimage *55*, 1423–1434.

Engvig, A., Fjell, A.M., Westlye, L.T., Moberget, T., Sundseth, Ø., Larsen, V.A., and Walhovd, K.B. (2010). Effects of memory training on cortical thickness in the elderly. Neuroimage *52*, 1667–1676.

Engvig, A., Fjell, A.M., Westlye, L.T., Moberget, T., Sundseth, Ø., Larsen, V.A., and Walhovd, K.B. (2012). Memory training impacts short-term changes in aging white matter: a longitudinal diffusion tensor imaging study. Hum. Brain Mapp. *33*, 2390–2406.

Esteban, O., Birman, D., Schaer, M., Koyejo, O.O., Poldrack, R.A., and Gorgolewski, K.J. (2017). MRIQC: advancing the automatic prediction of image quality in MRI from unseen sites. PLoS One *12*, e0184661

Eyler, L.T., Sherzai, A., Kaup, A.R., and Jeste, D.V. (2011). A review of functional brain imaging correlates of successful cognitive aging. Biol. Psychiatry *70*, 115–122.

Fahrenberg, J. (1979). Das Komplementaritätsprinzip in der psychophysiologischen Forschung und psychosomatischen Medizin. Z. Klin. Psychol. Psychother. *27*, 151–167.

Falk, E.B., Hyde, L.W., Mitchell, C., Faul, J., Gonzalez, R., Heitzeg, M.M., Keating, D.P., Langa, K.M., Martz, M.E., Maslowsky, J.,

et al. (2013). What is a representative brain? Neuroscience meets population science. Proc. Natl Acad. Sci. U.S.A. *110*, 17615–17622.

Ferrer, E. and McArdle, J.J. (2003). Alternative structural models for multivariate longitudinal data analysis. Struct. Equ. Model. A Multidiscip. J. *10*, 493–524.

Ferrer, E., Salthouse, T.A., Stewart, W.F., and Schwartz, B.S. (2004). Modeling age and retest processes in longitudinal studies of cognitive abilities. Psychol. Aging *19*, 243–259.

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., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Holland, D., Dale, A.M., Walhovd, K.B., et al. (2013). Critical ages in the life course of the adult brain: nonlinear subcortical aging. Neurobiol. Aging *34*, 2239–2247.

Fjell, A.M., Westlye, L.T., Grydeland, H., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Dale, A.M., and Walhovd, K.B. (2014). Accelerating cortical thinning: unique to dementia or universal in aging? Cereb. Cortex *24*, 919–934.

Fjell, A.M., Sneve, M.H., Storsve, A.B., Grydeland, H., Yendiki, A., and Walhovd, K.B. (2016). Brain events underlying episodic memory changes in aging: a longitudinal investigation of structural and functional connectivity. Cereb. Cortex *26*, 1272–1286.

Fjell, A.M., Sneve, M.H., Grydeland, H., Storsve, A.B., and Walhovd, K.B. (2017). The disconnected brain and executive function decline in aging. Cereb. Cortex *27*, 2303–2317.

Fleischman, D.A., Wilson, R.S., Gabrieli, J.D.E., Bienias, J.L., and Bennett, D.A. (2004). A longitudinal study of implicit and explicit memory in old persons. Psychol. Aging *19*, 617–625.

Folstein, M.F., Folstein, S.E., and McHugh, P.R. (1975). A practical method for grading the cognitive state of patients for the clinician. J. Psychiatr. Res. *12*, 189–198.

Foster, E.D. and Deardorff, A. (2017). Open Science Framework (OSF). J. Med. Libr. Assoc. *105*, 203–206.

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., and Sauer, H. (2013). BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. PLoS One *8*, e67346.

Gelman, A., Hill, J., and Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. J. Res. Educ. Eff. *5*, 189–211.

Goddings, A.-L., Mills, K.L., Clasen, L.S., Giedd, J.N., Viner, R.M., and Blakemore, S.-J. (2014). The influence of puberty on subcortical brain development. Neuroimage *88*, 242–251.

Goh, J.O., An, Y., and Resnick, S.M. (2012). Differential trajectories of age-related changes in components of executive and memory processes. Psychol. Aging *27*, 707–719.

Gorbach, T., Pudas, S., Lundquist, A., Orädd, G., Josefsson, M., Salami, A., de Luna, X., and Nyberg, L. (2017). Longitudinal association between hippocampus atrophy and episodic-memory decline. Neurobiol. Aging *51*, 167–176.

Gorgolewski, K.J., Auer, T., Calhoun, V.D., Craddock, R.C., Das, S., Duff, E.P., Flandin, G., Ghosh, S.S., Glatard, T., Halchenko, Y.O., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci. Data *3*, 1–9.

Gorgolewski, K.J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M.M., Churchill, N.W., Cohen, A.L., Craddock, R.C., Devenyi, G.A., et al. (2017). BIDS apps: improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. PLoS Comput. Biol. *13*, 1–16.

Grady, C.L. (2012). The cognitive neuroscience of ageing. Nat. Rev. Neurosci. *13*, 491–505.

Gregory, T., Callaghan, A., Nettelbeck, T., and Wilson, C. (2009). Inspection time predicts individual differences in everyday functioning among elderly adults: testing discriminant validity. Australas. J. Ageing *28*, 87–92.

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. A Multidiscip. J. *19*, 268–292.

Guye, S., Röcke, C., Mérillat, S., von Bastian, C.C., and Martin, M. (2016). Adult lifespan. In: Cognitive training: An Overview of Features and Applications. T. Strobach and J. Karbach, eds. (Berlin: Springer). pp. 45–55.

Hakun, J.G., Zhu, Z., Brown, C.A., Johnson, N.F., and Gold, B.T. (2015). Longitudinal alterations to brain function, structure, and cognitive performance in healthy older adults: a fMRI-DTI study. Neuropsychologia *71*, 225–235.

Hamaker, E.L. and Grasman, R.P.P.P. (2015). To center or not to center ? Investigating inertia with a multilevel autoregressive model. Front. Psychol. *5*, 1492.

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.

Healey, M.K., Ngo, K.W.J., and Hasher, L. (2014). Below-baseline suppression of competitors during interference resolution by younger but not older adults. Psychol. Sci. *25*, 145–151.

Hedden, T. and Gabrieli, J.D.E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. Nat. Rev. Neurosci. *5*, 87–96.

Hedman, A.M., van Haren, N.E.M., Schnack, H.G., Kahn, R.S., and Hulshoff Pol, H.E. (2012). Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. Hum. Brain Mapp. *33*, 1987–2002.

Heinen, R., Bouvy, W.H., Mendrik, A.M., Viergever, M.A., Biessels, G.J., and de Bresser, J. (2016). Robustness of automated methods for brain volume measurements across different MRI field strengths. PLoS One *11*, e0165719.

Henry, J.D., MacLeod, M.S., Phillips, L.H., and Crawford, J.R. (2004). A meta-analytic review of prospective memory and aging. Psychol. Aging *19*, 27–39.

Hertzog, C. and Nesselroade, J.R. (2003). Assessing psychological change in adulthood: an overview of methodological issues. Psychol. Aging *18*, 639–657.

Hertzog, C., Kramer, A.F., Wilson, R.S., Lindenberger, U., and Nesselroade, J.R. (2009). Enrichment effects on adult cognitive development. Psychol. Sci. *9*, 1–65.

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*, 150–164.

Hofer, S.M. and Sliwinski, M.J. (2001). Understanding ageing: an evaluation of research designs for assessing the interdependence of ageing-related changes. Gerontology *47*, 341–352.

Hofer, S.M., Flaherty, B.P., and Hoffman, L. (2006). Cross-sectional analysis of time-dependent data: mean-induced association in age-heterogeneous samples and an alternative method based on sequential narrow age-cohort samples. Multivariate Behav. Res. *41*, 165–187.

Hoffmann, L. (2012). Considering alternative metrics of time: does anybody really know what time is? In: Advances in Longitudinal Methods in the Social and Behavioral Sciences. J. Harring and G. Hancock, eds. (Charlotte, NC: Information Age Publishing). pp. 255–287.

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.

Hohman, T.J., Tommet, D., Marks, S., Contreras, J., Jones, R., and Mungas, D. (2017). Evaluating alzheimer's disease biomarkers as mediators of age-related cognitive decline. Neurobiol. Aging *58*, 120–128.

Hultsch, D.F., Hertzog, C., Small, B.J., McDonald-Miszczak, L., and Dixon, R.A. (1992). Short-term longitudinal change in cognitive performance in later life. Psychol. Aging *7*, 571–584.

Ihle, A., Hering, A., Mahy, C.E.V., Bisiacchi, P.S., and Kliegel, M. (2013). Adult age differences, response management, and cue focality in event-based prospective memory: a meta-analysis on the role of task order specificity. Psychol. Aging *28*, 714–720.

Insel, K., Morrow, D., Brewer, B., and Figueredo, A. (2006). Executive function, working memory, and medication adherence among older adults. J Gerontol. – Ser. B Psychol. Sci. Soc. Sci. *61*, 102–107.

Ioannidis, J.P.A., Munafò, M.R., Fusar-Poli, P., Nosek, B.A., and David, S.P. (2014). Publication and other reporting biases in cognitive sciences: detection, prevalence, and prevention. Trends Cogn. Sci. *18*, 235–241.

Jacobucci, R., Grimm, K.J., and McArdle, J.J. (2017). A comparison of methods for uncovering sample heterogeneity: structural equation model trees and finite mixture models. Struct. Equ. Model. *24*, 270–282.

Jäncke, L. (2017). Lehrbuch kognitive Neurowissenschaften (2nd ed.) (Bern: Hogrefe).

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

Johansson, B., Hofer, S.M., Allaire, J.C., Maldonado-Molina, M.M., Piccinin, A.M., Berg, S., Pedersen, N.L., and McClearn, G.E. (2004). Change in cognitive capabilities in the oldest old : the effects of proximity to death in genetically related individuals over a 6-year period. Psychol. Aging *19*, 145–156.

Jones, D.K. (2010). Challenges and limitations of quantifying brain connectivity *in vivo* with diffusion MRI. Imaging Med. *2*, 341–355.

Jovicich, J., Marizzoni, M., Sala-Llonch, R., Bosch, B., Bartrés-Faz, D., Arnold, J., Benninghoff, J., Wiltfang, J., Roccatagliata, L., Nobili, F., et al. (2013). Brain morphometry reproducibility in multi-center 3 T MRI studies: a comparison of cross-sectional and longitudinal segmentations. Neuroimage *83*, 472–484.

Juraska, J.M. and Lowry, N.C. (2012). Neuroanatomical changes associated with cognitive aging. Curr. Top Behav. Neurosci. *10*, 137–162.

Karama, S., Bastin, M.E., Murray, C., Royle, N.A., Penke, L., Muñoz Maniega, S., Gow, A.J., Corley, J., Valdés Hernández Mdel, C., Lewis, J.D., et al. (2014). Childhood cognitive ability accounts for associations between cognitive ability and brain cortical thickness in old age. Mol. Psychiatry *19*, 555–559.

Kaup, A.R., Mirzakhanian, H., Jeste, D.V., and Eyler, L.T. (2011). A review of the brain structure correlates of successful cognitive aging. J. Neuropsychiatry Clin. Neurosci. *23*, 6–15.

Kennelly, S.P., Lawlor, B.A., and Kenny, R.A. (2009). Blood pressure and dementia – a comprehensive review. Ther. Adv. Neurol. Disord. *2*, 241–260.

Kiar, G., Gorgolewski, K.J., Kleissas, D., Roncal, W.G., Litt, B., Wandell, B., Poldrack, R.A., Wiener, M., Vogelstein, R.J., Burns, R., et al. (2017). Science in the cloud (SIC): a use case in MRI connectomics. Gigascience *6*, 1–10.

Kievit, R.A., Brandmaier, A.M., Ziegler, G., van Harmelen, A.-L., de Mooij, S.M.M., Moutoussis, M., Goodyer, I.M., Bullmore, E., Jones, P.B., Fonagy, P., et al. (2018). Developmental cognitive neuroscience using latent change score models: a tutorial and applications. Dev. Cogn. Neurosci. *33*, 99–117.

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.

Kivimäki, M., Luukkonen, R., Batty, G.D., Ferrie, J.E., Pentti, J., Nyberg, S.T., Shipley, M.J., Alfredsson, L., Fransson, E.I., Goldberg, M., et al. (2018). Body mass index and risk of dementia: analysis of individual-level data from 1.3 million individuals. Alzheimer's Dement. *14*, 601–609.

Kliegel, M., Jäger, T., and Phillips, L.H. (2008). Adult age differences in event-based prospective memory: a meta-analysis on the role of focal versus nonfocal cues. Psychol. Aging *23*, 203–208.

Kliegel, M., Ballhausen, N., Hering, A., Ihle, A., Schnitzspahn, K.M., and Zuber, S. (2016). Prospective memory in older adults: where we are now and what is next. Gerontology *62*, 459–466.

Köhncke, Y., Laukka, E.J., Brehmer, Y., Kalpouzos, G., Li, T.-Q., Fratiglioni, L., Bäckman, L., and Lövdén, M. (2016). Three-year changes in leisure activities are associated with concurrent changes in white matter microstructure and perceptual speed in individuals aged 80 years and older. Neurobiol. Aging *41*, 173–186.

Kruschke, J.K. (2010). What to believe: Bayesian methods for data analysis. Trends Cogn. Sci. *14*, 293–300.

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.

Langa, K.M., Larson, E.B., Crimmins, E.M., Faul, J.D., Levine, D.A., Kabeto, M.U., and Weir, D.R. (2017). A comparison of the prevalence of dementia in the United States in 2000 and 2012. J. Am. Med. Assoc. Intern. Med. *177*, 51–58.

Larson, E.B., Yaffe, K., and Langa, K.M. (2013). New insights into the dementia epidemic. N. Engl. J. Med. *369*, 2275–2277.

Lawton, M.P., Moss, M., Hoffman, C., Grant, R., Ten Have, T., and Kleban, M.H. (1999). Health, valuation of life, and the wish to live. Gerontologist *39*, 406–416.

Leong, R.L.F., Lo, J.C., Sim, S.K.Y., Zheng, H., Tandi, J., Zhou, J., and Chee, M.W.L. (2017). Longitudinal brain structure and cognitive changes over 8 years in an East Asian cohort. Neuroimage *147*, 852–860.

Leow, A.D., Yanovsky, I., Parikshak, N., Hua, X., Lee, S., Toga, A.W., Clifford, R.J., Bernstein, M.A., Britson, P.J., Gunter, J.L., et al. (2009). Alzheimer's disease neuroimaging initiative: a 1-year follow up study using tensor-based morphometry correlating degenerative rates, biomarkers and cognition. Neuroimage *45*, 645–655.

Lerner, R.M., Schwartz, S.J., and Phelps, E. (2009). Problematics of time and timing in the longitudinal study of human fevelopment: theoretical and methodological issues. Hum. Dev. *52*, 44–68.

Liem, F., Mérillat, S., Bezzola, L., Hirsiger, S., Philipp, M., Madhyastha, T., and Jäncke, L. (2015). Reliability and statistical power analysis of cortical and subcortical FreeSurfer metrics in a large sample of healthy elderly. Neuroimage *108*, 95–109.

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Kharabian Masouleh, S., Huntenburg, J.M., Lampe, L., Rahim, M., Abraham, A., Craddock, R.C., et al. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. Neuroimage *148*, 179–188.

Light, L.L., Prull, M.W., LaVoie, D.J., and Healy, M.R. (2000). Dual-process theories of memory in old age. In Models of Cognitive Aging. T.J. Perfect and E.A. Maylor, eds. (Oxford, UK: Oxford University Press). pp. 238–300.

Lillard, A.S. and Erisir, A. (2011). Old dogs learning new tricks: neuroplasticity beyond the juvenile period. Dev. Rev. *31*, 207–239.

Lindenberger, U. and Ghisletta, P. (2009). Cognitive and sensory declines in old age: gauging the evidence for a common cause. Psychol. Aging *24*, 1–16.

Lindenberger, U., Singer, T., and Baltes, P.B. (2002). Longitudinal selectivity in aging populations: separating mortality-associated versus experimental components in the Berlin Aging Study (BASE). J Gerontol. – Ser. B Psychol. Sci. Soc. Sci. *57*, 474–482.

Lindenberger, U., von Oertzen, T., Ghisletta, P., and Hertzog, C. (2011). Cross-sectional age variance extraction: what's change got to do with it? Psychol. Aging *26*, 34–47.

Lindquist, M.A. and Mejia, A. (2015). Zen and the art of multiple comparisons. Psychosom. Med. *77*, 114–125.

Liu, H., Wang, L., Geng, Z., Zhu, Q., Song, Z., Chang, R., and Lv, H. (2016). A voxel-based morphometric study of age- and sex-related changes in white matter volume in the normal aging brain. Neuropsychiatr. Dis. Treat. *12*, 453–465.

Lockhart, S.N. and DeCarli, C. (2014). Structural imaging measures of brain aging. Neuropsychol. Rev. *24*, 271–289.

Lövdén, M., Bodammer, N.C., Kühn, S., Kaufmann, J., Schütze, H., Tempelmann, C., Heinze, H.-J., Düzel, E., Schmiedek, F., Lindenberger, U. (2010). Experience-dependent plasticity of white-matter microstructure extends into old age. Neuropsychologia *48*, 3878–3883.

Lövdén, M., Schaefer, S., Noack, H., Bodammer, N.C., Kühn, S., Heinze, H.J., Düzel, E., Bäckman, L., and Lindenberger, U. (2012). Spatial navigation training protects the hippocampus against age-related changes during early and late adulthood. Neurobiol. Aging *33*, 620.e9–620.e22.

Lövdén, M., Köhncke, Y., Laukka, E.J., Kalpouzos, G., Salami, A., Li, T.-Q., Fratiglioni, L., and Bäckman, L. (2014). Changes in perceptual speed and white matter microstructure in the corticospinal tract are associated in very old age. Neuroimage *102*, 520–530.

MacCallum, R.C., Kim, C., Malarkey, W.B., and Kiecolt-Glaser, J.K. (1997). Studying multivariate change using multilevel models and latent curve models. Multivariate Behav. Res. *32*, 215–253.

Madhyastha, T., Mérillat, S., Hirsiger, S., Bezzola, L., Liem, F., Grabowski, T., and Jäncke, L. (2014). Longitudinal reliability of tract-based spatial statistics in diffusion tensor imaging. Hum. Brain Mapp. *35*, 4544–4555.

Madsen, S.K., Gutman, B.A., Joshi, S.H., Toga, A.W., Jack Jr, C.R., Weiner, M.W., and Thompson, P.M. (2013). Mapping dynamic changes in ventricular volume onto baseline cortical surfaces in normal aging, MCI, and Alzheimer's disease. Multimodal Brain Image Anal. *8159*, 84–94.

Mak, E., Su, L., Williams, G.B., Watson, R., Firbank, M., Blamire, A.M., and O'Brien, J.T. (2015a). Longitudinal assessment of global and regional atrophy rates in Alzheimer's disease and dementia with Lewy bodies. NeuroImage Clin. *7*, 456–462.

Mak, E., Su, L., Williams, G.B., Watson, R., Firbank, M.J., Blamire, A.M., and O'Brien, J.T. (2015b). Progressive cortical thinning and subcortical atrophy in dementia with Lewy bodies and Alzheimer's disease. Neurobiol. Aging *36*, 1743–1750.

Mak, E., Su, L., Williams, G.B., Firbank, M.J., Lawson, R.A., Yarnall, A.J., Duncan, G.W., Mollenhauer, B., Owen, A.M., Khoo, T.K., et al. (2017). Longitudinal whole-brain atrophy and ventricular enlargement in nondemented Parkinson's disease. Neurobiol. Aging *55*, 78–90.

Mandolesi, L., Polverino, A., Montuori, S., Foti, F., Ferraioli, G., Sorrentino, P., and Sorrentino, G. (2018). Effects of physical exercise on cognitive functioning and wellbeing: biological and psychological benefits. Front. Psychol. *9*, 1–11.

Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., and Buckner, R.L. (2010). Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J. Cogn. Neurosci. *22*, 2677–2684.

Matthews, F.E., Stephan, B.C.M., Robinson, L., Jagger, C., Barnes, L.E., Arthur, A., Brayne, C., and CFAS Collaboration. (2016). A two decade dementia incidence comparison from the Cognitive Function and Ageing Studies I and II. Nat. Commun. *7*, 11398.

McArdle, J.J. (1988). Dynamic but structural equation modeling of repeated measures data. In: Handbook of Multivariate Experimental Psychology. Perspectives on Individual Differences. J.R. Nesselroade and R.B. Cattle, eds. (Boston, MA: Springer). pp. 561–614.

McArdle, J.J. (2009). Latent variable modeling of differences and changes with longitudinal data. Annu. Rev. Psychol. *60*, 577–605.

McArdle, J.J. and Bell, R.Q. (2000). An introduction to latent growth models for developmental data analysis. In: Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples. T.D. Little, K.U. Schnabel and J. Baumert, eds. (Mahwah, NJ: Lawrence Erlbaum Associates). pp. 69–107, 269–281.

McArdle, J.J. and Epstein, D. (1987). Latent growth curves within developmental structural equation models. Child Dev. *58*, 110–133.

McArdle, J.J. and Hamagami, F. (1992). Modeling incomplete longitudinal and cross-sectional data using latent growth structural models. Exp. Aging Res. *18*, 145–166.

McArdle, J.J. and Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In: Decade of Behavior. New Methods for the Analysis of Change. L.M. Collins and A.G. Sayer, eds. (Washington, DC: APA Press). pp. 139–175.

McLaughlin, S.J., Jette, A.M., and Connell, C.M. (2012). An examination of healthy aging across a conceptual continuum: prevalence estimates, demographic patterns, and validity. J. Gerontol. – Ser. A Biol. Sci. Med. Sci. *67*, 783–789.

Mehta, P.D. and West, S.G. (2000). Putting the individual back into individual growth curves. Psychol. Methods *5*, 23–43.

Merkle, E.C. and Rosseel, Y. (2016). Blavaan: Bayesian structural equation models via parameter expansion. Retrieved from http://arxiv.org/abs/1511.05604.

Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. *19*, 1523–1536.

Mills, K.L. and Tamnes, C.K. (2014). Methods and considerations for longitudinal structural brain imaging analysis across development. Dev. Cogn. Neurosci. *9*, 172–190.

Mills, K.L., Lalonde, F., Clasen, L.S., Giedd, J.N., and Blakemore, S.-J. (2014). Developmental changes in the structure of the social brain in late childhood and adolescence. Soc. Cogn. Affect. Neurosci. *9*, 123–131.

Mills, K.L., Goddings, A.-L., Herting, M.M., Meuwese, R., Blakemore, S.-J., Crone, E.A., Dahl, R.E., Güroğlu, B., Raznahan, A., Sowell, E.R., et al. (2016). Structural brain development between childhood and adulthood: convergence across four longitudinal samples. Neuroimage *141*, 273–281.

Miyake, A. and Friedman, N.P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. Curr. Dir. Psychol. Sci. *21*, 8–14.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., and Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex 'frontal lobe' tasks: a latent variable analysis. Cogn. Psychol. *41*, 49–100.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., and The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med *6*, 1–6, Article no.: e1000097.

Möller, C., Hafkemeijer, A., Pijnenburg, Y.A.L., Rombouts, S.A.R.B., van der Grond, J., Dopper, E., van Swieten, J., Versteeg, A., Steenwijk, M.D., Barkhof, F., 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.

Moon, S.Y., de Souto Barreto, P., Chupin, M., Mangin, J.F., Bouyahia, A., Fillon, L., Andrieu, S., and Vellas, B. (2017). Associations between white matter hyperintensities and cognitive decline over 3 years in non-dementia older adults with memory complaints. J. Neurol. Sci. *379*, 266–270.

Morcom, A.M. and Johnson, W. (2015). Neural reorganization and compensation in aging. J. Cogn. Neurosci. *27*, 1275–1285.

Morey, R.A., Selgrade, E.S., Wagner, H.R., Huettel, S.A., Wang, L., and McCarthy, G. (2010). Scan – rescan reliability of subcortical brain volumes derived from automated segmentation. Hum. Brain Mapp. *31*, 1751–1762.

Mozolic, J.L., Hayasaka, S., and Laurienti, P.J. (2010). A cognitive training intervention increases resting cerebral blood flow in healthy older adults. Front. Hum. Neurosci. *4*, 16.

Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E-J., Ware, J.J., Ioannidis, J.P.A. (2017). A manifesto for reproducible science. Nat. Hum. Behav. *1*, 1–9, Article no.: 0021.

Muthén, B. (2001). Latent variable mixture modeling. In: New Developments and Techniques in Structural Equation Modeling. G.A. Marcoulides and R.E. Schumacker, eds. (NJ: Lawrence Erlbaum Associates). pp. 1–33.

Nilsson, J. and Lövdén, M. (2018). Naming is not explaining: future directions for the 'cognitive reserve' and 'brain maintenance' theories. Alzheimers. Res. Ther. *10*, 34.

Nooner, K.B., Colcombe, S.J., Tobe, R.H., Mennes, M., Benedict, M.M., Moreno, A.L., Panek, L.J., Brown, S., Zavitz, S.T., Li, Q., et al. (2012). The NKI-Rockland sample: a model for accelerating the pace of discovery science in psychiatry. Front. Neurosci. *6*, 152.

Nord, C.L., Valton, V., Wood, J., and Roiser, J.P. (2017). Power-up: a reanalysis of 'power failure' in neuroscience using mixture modelling. J. Neurosci. *13*, 3592–3516.

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., and Bäckman, L. (2012). Memory aging and brain maintenance. Cell Press *16*, 292–305.

Oberauer, K. (2009). Design for a working memory. Psychol. Learn. Motiv. *51*, 45–100.

Palmer, L.J. (2007). UK Biobank: bank on it. Lancet *369*, 1980–1982.

Pannese, E. (2011). Morphological changes in nerve cells during normal aging. Brain Struct. Funct. *216*, 85–89.

Pardoe, H.R., Kucharsky Hiess, R., and Kuzniecky, R. (2016). Motion and morphometry in clinical and nonclinical populations. Neuroimage *135*, 177–185.

Park, D.C. and Festini, S.B. (2016). Theories of memory and aging: a look at the past and a glimpse of the future. J. Gerontol. – Ser. B Psychol. Sci. Soc. Sci. *72*, 82–90.

Park, D.C. and Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. Annu. Rev. Psychol. *60*, 173–196.

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.

Park, D.C., Polk, T.A., Park, R., Minear, M., Savage, A., and Smith, M.R. (2004). Aging reduces neural specialization in ventral visual cortex. Proc. Natl Acad. Sci. U.S.A. *101*, 13091–13095.

Park, S., Ryu, S.-H., Yang, J.-J., Kwon, H., Youn, J.-H., Cho, S.-J., and Lee, J.-Y. (2018). Neural predictors of cognitive improvement by multi-strategic memory training based on metamemory in older adults with subjective memory complaints. Sci. Rep. *8*, 1–11.

Persson, J., Pudas, S., Lind, J., Kauppi, K., Nilsson, L.-G., and Nyberg, L. (2012). Longitudinal structure-function correlates in elderly reveal MTL dysfunction with cognitive decline. Cereb. Cortex *22*, 2297–2304.

Persson, N., Ghisletta, P., Dahle, C.L., Bender, A.R., Yang, Y., Yuan, P., Daugherty, A.M., and Raz, N. (2014). Regional brain shrinkage over 2 years: individual differences and effects of pro-inflammatory genetic polymorphisms. Neuroimage *103*, 334–348.

Persson, N., Ghisletta, P., Dahle, C.L., Bender, A.R., Yang, Y., Yuan, P., Daugherty, A.M., and Raz, N. (2016). Regional brain shrinkage and change in cognitive performance over 2 years: the bidirectional influences of the brain and cognitive reserve factors. Neuroimage *126*, 15–26.

Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., et al. (2010). Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. Neurology *74*, 201–209.

Plude, D.J. and Hoyer, W.J. (1986). Age and the selectivity of visual information processing. Psychol. Aging *1*, 4–10.

Poldrack, R.A., Baker, C.I., Durnez, J., Gorgolewski, K.J., Matthews, P.M., Munafò, M.R., Nichols, T.E., Poline, J.B., Vul, E., Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat. Rev. Neurosci. *18*, 115–126.

Prins, N.D. and Scheltens, P. (2015). White matter hyperintensities, cognitive impairment and dementia: an update. Nat. Rev. Neurol. *11*, 157–165.

Pudas, S., Josefsson, M., Rieckmann, A., and Nyberg, L. (2018). Longitudinal evidence for increased functional response in frontal cortex for older adults with hippocampal atrophy and memory decline. Cereb. Cortex *28*, 936–948.

Rast, P. and Hofer, S.M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. Psychol. Methods *19*, 133–154.

Rast, P., Kennedy, K.M., Rodrigue, K.M., Robinson, P.R.A.W., Gross, A.L., McLaren, D.G., Grabowski, T., Schaie, K.W., and Willis, S.L. (2017). APOEε4 genotype and hypertension modify 8-year cortical thinning: five occasion evidence from the Seattle Longitudinal Study. Cereb. Cortex *28*, 1934–1945.

Raudenbush, S.W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. Annu. Rev. Psychol. *52*, 501–525.

Raudenbush, S.W. and Bryk, A.S. (2002). Hierarchical Linear Models: Applications and Data Analysis Methods (2nd ed.) (Thousand Oaks: SAGE Publications).

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.

Raz, N., Gunning, F.M., Head, D., Dupuis, J.H., McQuain, J., Briggs, S.D., Loken, W.J., Thornton, A.E., and Acker, J.D. (1997). Selective aging of the human cerebral cortex observed *in vivo*: differential vulnerability of the prefrontal gray matter. Cereb. Cortex *7*, 268–282.

Raz, N., Lindenberger, U., Rodrigue, K.M., Kennedy, K.M., Head, D., Williamson, A., Dahle, C., Gerstorf, D., and Acker, J.D. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cereb. Cortex *15*, 1676–1689.

Raz, N., Rodrigue, K.M., Kennedy, K.M., and Acker, J.D. (2007). Vascular health and longitudinal changes in brain and cognition in middle-aged and older adults. Neuropsychology *21*, 149–157.

Raz, N., Lindenberger, U., Ghisletta, P., Rodrigue, K.M., Kennedy, K.M., and Acker, J.D. (2008). Neuroanatomical correlates of fluid intelligence in healthy adults and persons with vascular risk factors. Cereb. Cortex *18*, 718–726.

Resnick, S.M., Pham, D.L., Kraut, M.A., Zonderman, A.B., and Davatzikos, C. (2003). Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. J. Neurosci. *23*, 3295–3301.

Reuter-Lorenz, P.A. and Park, D.C. (2010). Human neuroscience and the aging mind: a new look at old problems. J. Gerontol. – Ser. B Psychol. Sci. Soc. Sci. *65*, 405–415.

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.

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.

Reuter, M., Tisdall, M.D., Qureshi, A., Buckner, R.L., van der Kouwe, A.J.W., and Fischl, B. (2015). Head motion during MRI acquisition reduces gray matter volume and thickness estimates. Neuroimage *107*, 107–115.

Ritchie, S.J., Bastin, M.E., Tucker-Drob, E.M., Muñoz Maniega, S., Engelhardt, L.E., Cox, S.R., Royle, N.A., Gow, A.J., Corley, J., Pattie, A., et al. (2015a). Coupled changes in brain white matter microstructure and fluid intelligence in later life. J. Neurosci. *35*, 8672–8682.

Ritchie, S.J., Dickie, D.A., Cox, S.R., Valdes Hernandez, M.C., Corley, J., Royle, N.A., Pattie, A., Aribisala, B.S., Redmond, P., Muñoz Maniega, S., et al. (2015b). Brain volumetric changes and cognitive ageing during the eighth decade of life. Hum. Brain Mapp. *36*, 4910–4925.

Robitaille, A., Piccinin, A.M., Muniz-Terrera, G., Hoffman, L., Johansson, B., Deeg, D.J.H., Aartsen, M.J., Comijs, H.C., and Hofer, S.M. (2013). Longitudinal mediation of processing speed on age-related change in memory and fluid intelligence. Psychol. Aging *28*, 887–901.

Rodriguez-Laso, A., McLaughlin, S.J., Urdaneta, E., and Yanguas, J. (2018). Defining and estimating healthy aging in spain: a cross-sectional study. Gerontologist *58*, 388–398.

Rohrer, J.M. (2018). Thinking clearly about correlations and causation: graphical causal models for observational data. Adv. Methods Pract. Psychol. Sci. *1*, 27–42.

Rönnlund, M., Nyberg, L., Bäckman, L., and Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: cross-sectional and longitudinal data from a population-based study. Psychol. Aging *20*, 3–18.

Rosen, A.F.G., Roalf, D.R., Ruparel, K., Blake, J., Seelaus, K., Villa, L.P., Ciric, R., Cook, P.A., Davatzikos, C., Elliott, M.A., et al. (2018). Quantitative assessment of structural image quality. Neuroimage *169*, 407–418.

Sachdev, P., Wen, W., Chen, X., and Brodaty, H. (2007). Progression of white matter hyperintensities in elderly individuals over 3 years. Neurology *68*, 214–222.

Sala-Llonch, R., Junqué, C., Arenaza-Urquijo, E.M., Vidal-Piñeiro, D., Valls-Pedret, C., Palacios, E.M., Domènech, S., Salvà, A., Bargalló, N., and Bartrés-Faz, D. (2014). Changes in whole-brain functional networks and memory performance in aging. Neurobiol. Aging *35*, 2193–2202.

Sala-Llonch, R., Idland, A.V., Borza, T., Watne, L.O., Wyller, T.B., Brækhus, A., Zetterberg, H., Blennow, K., Walhovd, K.B., and Fjell, A.M. (2017). Inflammation, amyloid, and atrophy in the aging brain: relationships with longitudinal changes in cognition. J. Alzheimer's Dis. *58*, 829–840.

Salat, D.H., Kaye, J.A., and Janowsky, J.S. (2002). Greater orbital prefrontal volume selectively predicts worse working memory performance in older adults. Cereb. Cortex *12*, 494–505.

Salat, D.H., Tuch, D.S., Hevelone, N.D., Fischl, B., Corkin, S., Rosas, H.D., and Dale, A.M. (2005). Age-related changes in prefrontal white matter measured by diffusion tensor imaging. Ann. NY Acad. Sci. *1064*, 37–49.

Salthouse, T.A. (1996). The processing-speed theory of adult age differences in cognition. Psychol. Rev. *103*, 403–428.

Salthouse, T.A. (2003). Memory aging from 18 to 80. Alzheimer's Dis. Assoc. Disord. *17*, 162–167.

Salthouse, T.A. (2010). Selective review of cognitive aging. J. Int. Neuropsychol. Soc. *16*, 754–760.

Salthouse, T.A. (2011). Neuroanatomical substrates of age-related cognitive decline. Psychol. Bull. *137*, 753–784.

Salthouse, T.A. (2012). Consequences of age-related cognitive declines. Annu. Rev. Psychol. *63*, 201–226.

Salthouse, T.A. (2014). Why are there different age relations in cross-sectional and longitudinal comparisons of cognitive functioning? Curr. Dir. Psychol. Sci. *23*, 252–256.

Salthouse, T.A. and Pink, J.E. (2008). Why is working memory related to fluid intelligence? Psychon. Bull. Rev. *15*, 364–371.

Salthouse, T.A., Schroeder, D.H., and Ferrer, E. (2004). Estimating retest effects in longitudinal assessments of cognitive functioning in adults between 18 and 60 years of age. Dev. Psychol. *40*, 813–822.

Satz, P., Cole, M.A., Hardy, D.J., and Rassovsky, Y. (2011). Brain and cognitive reserve: mediator(s) and construct validity, a critique. J. Clin. Exp. Neuropsychol. *33*, 121–130.

Savalia, N.K., Agres, P.F., Chan, M.Y., Feczko, E.J., Kennedy, K.M., and Wig, G.S. (2017). Motion-related artifacts in structural brain images revealed with independent estimates of in-scanner head motion. Hum. Brain Mapp. *38*, 472–492.

Schacter, D.L. (1987). Implicit memory: history and current status. J. Exp. Psychol. Learn. Mem. Cogn. *13*, 501–518.

Schaie, K.W. (1996). Intellectual Development in Adulthood: The Seattle Longitudinal Study (New York: Cambridge University Press).

Schaie, K.W. (2005). What can we learn from longitudinal studies of adult development? Res. Hum. Dev. *2*, 133–158.

Schaie, K.W. and Willis, S.L. (2010). The Seattle Longitudinal Study of adult cognitive development. ISSBD Bull. *57*, 24–29.

Schmidt, R., Fazekas, F., Kapeller, P., Schmidt, H., and Hartung, H.-P. (1999). MRI white matter hyperintensities: 3-year follow-up of the Austrian Stroke Prevention Study. Neurology *53*, 132–139.

Schmidt, R., Ropele, S., Enzinger, C., Petrovic, K., Smith, S., Schmidt, H., Matthews, P.M., and Fazekas, F. (2005). White matter lesion progression, brain atrophy, and cognitive decline: the Austrian stroke prevention study. Ann. Neurol. *58*, 610–616.

Schmiedek, F., Hildebrandt, A., Lövdén, M., Lindenberger, U., and Wilhelm, O. (2009). Complex span versus updating tasks of working memory: the gap is not that deep. J. Exp. Psychol. Learn. Mem. Cogn. *35*, 1089–1096.

Schnitzspahn, K.M., Ihle, A., Henry, J.D., Rendell, P.G., and Kliegel, M. (2011). The age-prospective memory-paradox: an exploration of possible mechanisms. Int. Psychogeriatrics *23*, 583–52.

Seresinhe, C.I., Preis, T., and Moat, H.S. (2015). Quantifying the impact of scenic environments on health. Sci. Rep. *5*, 16899.

Serra, L., Musicco, M., Cercignani, M., Torso, M., Spanò, B., Mastropasqua, C., Giulietti, G., Marra, C., Bruno, G., Koch, G., et al. (2015). Cognitive reserve and the risk for Alzheimer's disease: a longitudinal study. Neurobiol. Aging *36*, 592–600.

Sexton, C.E., Walhovd, K.B., Storsve, A.B., Tamnes, C.K., Westlye, L.T., Johansen-Berg, H., and Fjell, A.M. (2014). Accelerated changes in white matter microstructure during aging: a longitudinal diffusion tensor imaging study. J. Neurosci. *34*, 15425–15436.

Shafto, M.A., Tyler, L.K., Dixon, M., Taylor, J.R., Rowe, J.B., Cusack, R., Calder, A.J., Marslen-Wilson, W.D., Duncan, J., Dalgleish, T., et al. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. BMC Neurol. *14*, 1–25.

Shehzad, Z., Giavasis, S., Li, Q., Benhajali, Y., Yan, C., Yang, Z., Milham, M., Bellec, P., and Craddock, C. (2015). The preprocessed connectomes project quality assessment protocol – a resource for measuring the quality of MRI data. Front. Neurosci. Conference Abstract: Neuroinformatics (Cairns, Cairns, Australia).

Sherif, T., Rioux, P., Rousseau, M.-E., Kassis, N., Beck, N., Adalat, R., Das, S., Glatard, T., Evans, A.C. (2014). CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. Front. Neuroinform. *8*, 54.

Silbert, L.C., Nelson, C., Howieson, D.B., and Kaye, J.A. (2008). Impact of white matter hyperintensity volume progression on rate of cognitive and motor decline. Neurology *71*, 108–113.

Silberzahn, R., Uhlmann, E.L., Martin, D.P., Anselmi, P., Aust, F., Awtrey, E., Bahník, Š., Bai, F., Bannard, C., Bonnier, E., et al. (2018). Many analysts, one dataset: making transparent how variations in analytical choices affect results. Adv. Methods Pract. Psychol. Sci. *1*, 337–356.

Singer, J.D. and Willett, J.B. (2003). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence (Oxford: Oxford University Press).

Sliwinski, M.J., Hoffman, L., and Hofer, S.M. (2010). Evaluating convergence of within-person change and between-person age differences in age-heterogeneous longitudinal studies. Res. Hum. Dev. Dev. *7*, 45–60.

Snijders, T.A.B. and Bosker, R.J. (1999). Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (London: SAGE Publications).

Snyder, H.R., Miyake, A., and Hankin, B.L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. Front. Psychol. *6*, 328.

Song, S.-K., Sun, S.-W., Ramsbottom, M.J., Chang, C., Russell, J., and Cross, A.H. (2002). Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. Neuroimage *17*, 1429–1436.

Song, Z., Farrell, M.E., Chen, X., and Park, D.C. (2018). Longitudinal accrual of neocortical amyloid burden Is associated with microstructural changes of the fornix in cognitively normal adults. Neurobiol. Aging *68*, 114–122.

Sowell, E.R., Thompson, P.M., and Toga, A.W. (2004). Mapping changes in the human cortex throughout the span of life. Neuroscience *10*, 372–392.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. J. Int. Neuropsychol. Soc. *8*, 448–460.

Stern, Y. (2009). Cognitive reserve. Neuropsychologia *47*, 2015–2028.

Stern, Y., Arenaza-Urquijo, E.M., Bartrés-Faz, D., Belleville, S., Cantilon, M., Chetelat, G., Ewers, M., Franzmeier, N., Kempermann, G., Kremen, W.S., et al. (2018). Whitepaper: defining and investigating cognitive reserve, brain reserve, and brain maintenance. Alzheimer's Dement. 1–7.

Storsve, A.B., Fjell, A.M., Tamnes, C.K., Westlye, L.T., Overbye, K., Aasland, H.W., and Walhovd, K.B. (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.

Storsve, A.B., Fjell, A.M., Yendiki, A., and Walhovd, K.B. (2016). Longitudinal changes in white matter tract integrity across the adult lifespan and its relation to cortical thinning. PLoS One *11*, e0156770.

Sullivan, E.V., Rohlfing, T., and Pfefferbaum, A. (2010). Quantitative fiber tracking of lateral and interhemispheric white matter

systems in normal aging: relations to timed performance. Neurobiol. Aging *31*, 464–481.

Suo, C., Singh, M.F., Gates, N., Wen, W., Sachdev, P., Brodaty, H., Saigal, N., Wilson, G.C., Meiklejohn, J., Singh, N., et al. (2016). Therapeutically relevant structural and functional mechanisms triggered by physical and cognitive exercise. Mol. Psychiatry *21*, 1633–1642.

Tamnes, C.K., Walhovd, K.B., Dale, A.M., Østby, Y., Grydeland, H., Richardson, G., Westlye, L.T., Roddey, J.C., Hagler Jr, D.J., Due-Tønnessen, P., et al. (2013). Brain development and aging: overlapping and unique patterns of change. Neuroimage *68*, 63–74.

Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Cam-CAN, and Henson, R.N. (2017). The cambridge centre for ageing and neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. Neuroimage *144*, 262–269.

Teipel, S.J., Meindl, T., Wagner, M., Stieltjes, B., Reuter, S., Hauenstein, K.-H., Filippi, M., Ernemann, U., Reiser, M.F., Hampel, 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.

ten Brinke, L.F., Davis, J.C., Barha, C.K., and Liu-Ambrose, T. (2017). Effects of computerized cognitive training on neuroimaging outcomes in older adults: a systematic review. BMC Geriatr. *17*, 1–20.

Treitz, F.H., Heyder, K., and Daum, I. (2007). Differential course of executive control changes during normal aging. Aging Neuropsychol. Cogn. *14*, 370–393.

Tucker-Drob, E.M. and Salthouse, T.A. (2008). Adult age trends in the relations among cognitive abilities. Psychol. Aging *23*, 453–460.

Tulving, E. (1972). Episodic and semantic memory. In: Organization of Memory. E. Tulving and W. Donaldson, eds. (Oxford, England: Academic Press). pp. 381–403.

United Nations. (2017). World population prospects: The 2017 revision, key findings and advance tables. Retrieved from https://esa.un.org/unpd/wpp/publications/Files/WPP2017_KeyFindings.pdf.

Valizadeh, S.A., Hänggi, J., Mérillat, S., and Jäncke, L. (2017). Age prediction on the basis of brain anatomical measures. Hum. Brain Mapp. *38*, 997–1008.

Valkanova, V., Eguia Rodriguez, R., and Ebmeier, K.P. (2014). Mind over matter – what do we know about neuroplasticity in adults? Int. Psychogeriatrics *26*, 891–909.

Van der Elst, W., Molenberghs, G., Van Boxtel, M.P.J., and Jolles, J. (2013). Establishing normative data for repeated cognitive assessment: a comparison of different statistical methods. Behav. Res. Methods *45*, 1073–1086.

Van Gerven, P.W.M. and Guerreiro, M.J.S. (2016). Selective attention and sensory modality in aging: curses and blessings. Front. Hum. Neurosci. *10*, 147.

Van Petten, C., Plante, E., Davidson, P.S.R., Kuo, T.Y., Bajuscak, L., and Glisky, E.L. (2004). Memory and executive function in older adults: relationships with temporal and prefrontal gray matter volumes and white matter hyperintensities. Neuropsychologia *42*, 1313–1335.

Verbeke, G. and Davidian, M. (2009). Joint models for longitudinal data: Introduction and overview. In: Longitudinal Data Analysis. G. Fitzmaurice, M. Davidian, G. Verbeke, and G. Molenberghs, eds. (Boca Raton, FL: Chapman and Hall/CRC).

Verbeke, G. and Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data (New York: Springer). pp. 319–326.

Verhaeghen, P. (2011). Aging and executive control: reports of a demise greatly exaggerated. Curr. Dir. Psychol. Sci. *20*, 174–180.

Verhaeghen, P. and Cerella, J. (2002). Aging, executive control, and attention: a review of meta-analyses. Neurosci. Biobehav. Rev. *26*, 849–857.

Verhaeghen, P., Steitz, D.W., Sliwinski, M.J., and Cerella, J. (2003). Aging and dual-task performance: a meta-analysis. Psychol. Aging *18*, 443–460.

Verhaeghen, P., Martin, M., and Sędek, G. (2012). Reconnecting cognition in the lab and cognition in real life: the role of compensatory social and motivational factors in explaining how cognition ages in the wild. Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn. *19*, 1–12.

Vidal-Piñeiro, D., Sneve, M.H., Nyberg, L.H., Mowinckel, A.M., Sederevicius, D., Walhovd, K.B., and Fjell, A.M. (2018). Maintained frontal activity underlies high memory function over 8 years in aging. Cereb. Cortex, bhy177. https://doi.org/10.1093/cercor/bhy177

Vijayakumar, N., Mills, K.L., Alexander-Bloch, A., Tamnes, C.K., and Whittle, S. (2017). Structural brain development: a review of methodological approaches and best practices. Dev. Cogn. Neurosci. *33*, 129–148.

Voelkle, M.C. and Oud, J.H.L. (2013). Continuous time modelling with individually varying time intervals for oscillating and non-oscillating processes. Br. J. Math. Stat. Psychol. *66*, 103–126.

Voelkle, M.C., Oud, J.H.L., Davidov, E., and Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. Psychol. Methods *17*, 176–192.

Vogelstein, J.T., Mensh, B., Häusser, M., Spruston, N., Evans, A.C., Kording, K., Amunts, K., Ebell, C., Muller, J., Telefont, M., et al. (2016). To the cloud! A grassroots proposal to accelerate brain science discovery. Neuron *92*, 622–627.

von Oertzen, T. (2010). Power equivalence in structural equation modelling. Br. J. Math. Stat. Psychol. *63*, 257–272.

von Oertzen, T. and Brandmaier, A.M. (2013). Optimal study design with identical power: an application of power equivalence to latent growth curve models. Psychol. Aging *28*, 414–428.

von Oertzen, T., Brandmaier, A.M., and Tsang, S. (2015). Structural equation modeling with Ωnyx. Struct. Equ. Model. Multidiscip. J. *22*, 148–161.

Walhovd, K.B., Westlye, L.T., Amlien, I., Espeseth, T., Reinvang, I., Raz, N., Agartz, I., Salat, D.H., Greve, D.N., Fischl, B., et al. (2011). Consistent neuroanatomical age-related volume differences across multiple samples. Neurobiol. Aging *32*, 916–932.

Walhovd, K.B., Johansen-Berg, H., and Káradóttir, R.T. (2014). Unraveling the secrets of white matter – bridging the gap between cellular, animal and human imaging studies. Neuroscience *276*, 2–13.

Walhovd, K.B., Krogsrud, S.K., Amlien, I.K., Bartsch, H., Bjørnerud, A., Due-Tønnessen, P., Grydeland, H., Hagler Jr, D.J., Håberg, A.K., Kremen, W.S., et al. (2016). Neurodevelopmental origins of lifespan changes in brain and cognition. Proc. Natl Acad. Sci. U.S.A. *113*, 9357–9362.

Wang, L. and Maxwell, S.E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. Psychol. Methods *20*, 63–83.

Wardlaw, J.M., Valdés Hernández, M.C., and Muñoz-Maniega, S. (2015). What are white matter hyperintensities made of? Relevance to vascular cognitive impairment. J. Am. Heart Assoc. *4*, 1–19.

Wasylyshyn, C., Verhaeghen, P., and Sliwinski, M.J. (2011). Aging and task switching: a meta-analysis. Psychol. Aging *26*, 15–20.

Weiskopf, N., Mohammadi, S., Lutti, A., and Callaghan, M.F. (2015). Advances in MRI-based computational neuroanatomy. Curr. Opin. Neurol. *28*, 313–322.

Wenger, E., Schaefer, S., Noack, H., Kühn, S., Mårtensson, J., Heinze, H.-J., Düzel, E., Bäckman, L., Lindenberger, U., Lövdén, M. (2012). Cortical thickness changes following spatial navigation training in adulthood and aging. Neuroimage *59*, 3389–3397.

Wenger, E., Brozzoli, C., Lindenberger, U., and Lövdén, M. (2017). Expansion and renormalization of human brain structure during skill acquisition. Trends Cogn. Sci. *21*, 930–939.

Westlye, L.T., Walhovd, K.B., Dale, A.M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., Grydeland, H., Tamnes, C.K., Ostby, Y., and Fjell, A.M. (2010). Life-span changes of the human brain white matter: diffusion tensor imaging (DTI) and volumetry. Cereb. Cortex *20*, 2055–2068.

Whitwell, J.L., Crum, W.R., Watt, H.C., and Fox, N.C. (2001). Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. Am. J. Neuroradiol. *22*, 1483–1489.

Willett, J.B. (1989). Some results on reliability for the longitudinal measurement of change: implications for the design of studies of individual growth. Educ. Psychol. Meas. *49*, 587–602.

Willis, S.W., Rast, P., McLaren, D., Grabowski, T., Kennedy, K., Rodridgue, K., Gross, A., Robinson, P., and Schaie, K.W. (2013). Comparison of findings of longitudinal change in cortical thickness between cross-sectional and longitudinal processing streams in freesurfer. Meeting for the Society for Neuroscience. San Diego, CA.

World Health Organization. (1946). Preamble to the constitution of the World Health Organization as adopted by the international health conference, New York, 19–22 June, 1946. New York. Retrieved from http://whqlibdoc.who.int/hist/official_records/constitution.pdf.

World Health Organization. (2015). World Report on Ageing and Health (World Health Organization). Retrieved from http://www.who.int/ageing/events/world-report-2015-launch/en/.

World Health Organization. (2016). World Health Statistics 2016: Monitoring Health for the SDGs, Sustainable Development Goals. Retrieved from http://www.who.int/gho/publications/world_health_statistics/2016/en/.

Yarkoni, T. and Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. Perspect. Psychol. Sci. *12*, 1100–1122.

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.

Yendiki, A., Reuter, M., Wilkens, P., Rosas, H.D., and Fischl, B. (2016). Joint reconstruction of white-matter pathways from longitudinal diffusion MRI data with anatomical priors. Neuroimage *127*, 277–286.

Yuan, P., Voelkle, M.C., and Raz, N. (2018). Fluid intelligence and gross structural properties of the cerebral cortex in middle-aged and older adults: a multi-occasion longitudinal study. Neuroimage *172*, 21–30.

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.

Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., and Alexander, D.C. (2012). NODDI: practical *in vivo* neurite orientation dispersion and density imaging of the human brain. Neuroimage *61*, 1000–1016.

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.

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.

Web references

Lifebrain (n.d). Retrieved May 16, 2019, from http://lifebrain.uio.no/.

GitHub (n.d). Retrieved May 16, 2019, from http://github.com/.

Open Science Framework (n.d). Retrieved May 16, 2019, from http://osf.io/.

OpenNeuro (n.d). Retrieved May 16, 2019, from http://openneuro.org/.

Open Neuroimaging Laboratory (n.d). Retrieved May 16, 2019, from http://openneu.ro/.