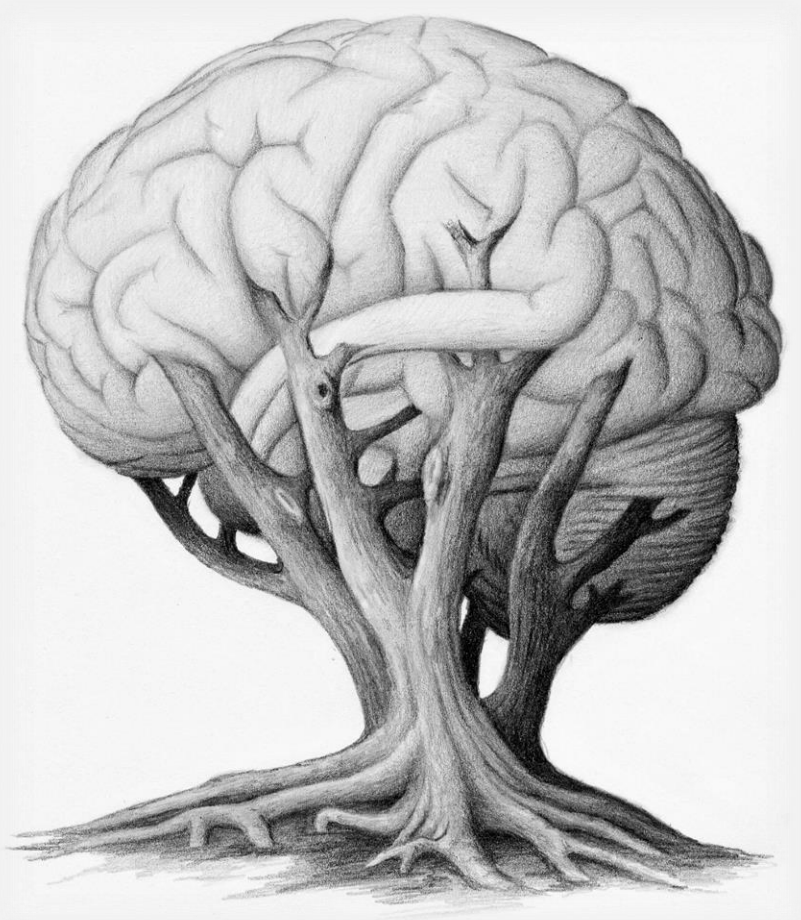
Cognitive performance across the lifespan and domains



Suzanne Swagerman

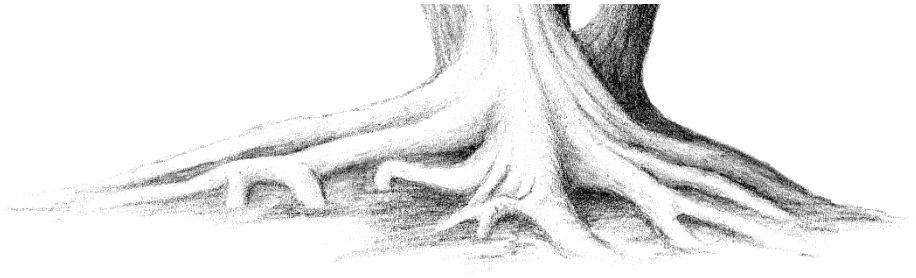
Cognitive performance across the lifespan and domains

Suzanne Swagerman

Table of contents

Chapter 1

General introduction



Cognitive functioning refers to a person's ability to process thoughts and use existing knowledge to generate new knowledge. It encompasses the set of all mental abilities and conscious and unconscious processes related to for example attention, memory and working memory, judgment and evaluation, reasoning and "computation", problem solving and decision making, speech and comprehension of language. In most healthy individuals the brain is capable of learning new knowledge in each of these areas. This capacity is most notable in infanthood and early childhood, the periods of time when most humans are best able to absorb and use new information. Children learn new words, concepts, and ways to express themselves on a weekly or even daily basis. The capacity to learn slows down as the individual ages, but overall cognitive function should not diminish on a large scale in healthy adult individuals. In later adulthood, however, effects of cognitive aging will increasingly come in play, although not equally strong in all domains. Functioning in some cognitive domains decreases predictably, such as speed of information processing and working memory, whereas functioning in other domains can be maintained or even improved with aging, such as vocabulary, knowledge and wisdom (deeper understanding to apply knowledge, Blazer, Yaffe, & Liverman, 2015).

Cognitive functioning is associated with multiple components of mental health and well-being: cognitive dysfunctions (in attention, working memory, executive functioning and memory) are often part of psychiatric disorders (e.g., schizophrenia and depression), or a key component in developmental disorders like dyslexia (reading problems), dyscalculia (arithmetic problems), attention deficit hyperactivity disorder (ADHD) or autism (characterized by social and communication problems, and repetitive and inflexible behaviors). Optimal development and maintenance of cognitive abilities is therefore of great importance to all members of the population: not only to excel in academics or work, but also to reduce problems in everyday life. In older adults, staying sharp of mind and retaining a good memory is a major concern as they impact on the ability to carry out daily activities and retaining autonomy and quality of life (Blazer et al., 2015). Cognitive aging presents an important societal challenge as our current society is faced with increasing numbers of elderly people, with an expected increase in the percentage of people of over 65 years in Europe from 14% in 2010 to 25% in 2050 (World Health Organization, 2015).

Within psychology, the concept of cognitive functioning is closely related to abstract concepts such as mind and intelligence, and global indices of cognitive functioning are often indexed by general intelligence ('IQ') or educational achievement. However, many more separate cognitive functions, like attention, working memory, reasoning or emotion processing, can be assessed separately by a variety of neurocognitive tests. Whether operationalized as intelligence or

as the performance on neurocognitive tests of more specific cognitive skills, large individual differences are found across the entire life span. In view of the importance of cognitive functioning for mental health and well-being in everyday life, understanding the determinants and modifiers of these differences remains a major research mission.

Causes of individual differences

Multiple factors cause differences between individuals in their level of cognitive functioning. Two of these, age and sex, are fixed effects that cannot be changed by any intervention. They should always be taken into account in the analysis of individual differences as they can exert substantial effects. This is most evident for age. During childhood and adolescence many cognitive functions are still developing, whereas cognitive performance gradually declines during older age (Salthouse, 2009).

Sex differences for some cognitive functions are apparent already during childhood (Gur et al., 2012) and further increase during adolescence. Sex differences may be related to the fact that males and females differ in hormone levels as well as brain structure, but sociocultural factors have also been suggested to play an important role (Halpern, Benbow, Geary, Hyde, & Gernsbacher, 2007). The most commonly suggested sex differences are a female advantage for verbal skills and a male advantage for spatial skills. However, sex differences may not be limited to these domains but also be present in, for example, memory functioning (Andreano & Cahill, 2009). Men and women appear to differ in their sensitivity to effects of cognitive aging, with cognition in women relatively more spared (Maylor et al., 2007). This sex by age interaction is complicated by detrimental effects of menopause. However, overall, meta-analyses suggest that for the majority of traits sex differences are small or trivial (Hyde, 2014).

Even when accounting for age and sex by stratification or covariate analysis, vast individual differences in cognitive abilities remain and these differences are seen across the large arsenal of neurocognitive tests available. Performance on these neurocognitive tests relies on the activation of specific brain areas and networks, and this activation also differs among people. Functional magnetic resonance imaging (fMRI) studies have shown that intelligence is associated with neural activation patterns and brain connectivity (Bassett et al., 2009; Cole, Yarkoni, Repovs, Anticevic, & Braver, 2012; Koenis et al., 2015; Langer et al., 2012; Park & Friston, 2013; Ramsden et al., 2011). Further, more efficient brain networks are associated with higher intelligence scores (Schmithorst & Holland, 2007; Song et al., 2008; van den Heuvel, Stam, Kahn, & Hulshoff Pol, 2009).

Neurobiological markers are not only associated with cognitive functioning, but with other components of mental health as well, as abnormal structure and function have been shown for the majority of psychiatric disorders (Etkin, Gyurak, & O'Hara, 2013).

Currently, many studies focus on trajectories of abnormal brain development, in addition to structural alterations at a specific time point (Giedd et al., 2015; Gu et al., 2015; Rapoport & Gogtay, 2008). This stresses the importance of healthy brain structure and function for normal cognitive and mental functioning. The brain is still under intense development in children and adolescents. Of special relevance is an understanding of brain development during adolescence, as the brain undergoes extensive reorganization both structurally and functionally during this period of life, when there is significant cognitive, emotional and social development but also the highest incidence of onset of psychiatric disorders (Lenroot & Giedd, 2006; Paus, Keshavan, & Giedd, 2008). Knowing which factors influence variation in brain structure and function throughout the lifespan provides insight into the pathways guiding normal and abnormal brain development, and ultimately into mechanisms underlying neuropsychiatric disorders.

Genes and environment as causes of individual differences

The etiology of variation in cognitive and neurobiological functioning is for a significant part explained by genetic differences between individuals. Studies of general intelligence (IQ), brain volume and brain function indicate that these traits are under relatively strong genetic influence, although the size of heritability estimates (the proportion of total trait variance explained by genetic factors) may depend on age. In this sense, age not only influences the level of cognitive function of an individual, but also modifies the importance of genetic factors.

The heritability of IQ is well established and increases from childhood to adulthood (Haworth et al., 2010), but the heritability of specific cognitive skills is less clear, partly because fewer studies have focused on the assessment of specific skills, and partly because a broader range of instruments has been used across studies. For brain volumes and brain function, a similar situation exists. The heritability of global brain volumes (e.g., total brain, total grey matter and total white matter volume) is high (Peper, Brouwer, Boomsma, Kahn, & Hulshoff Pol, 2007; Thompson et al., 2001) and well established, whereas fewer studies have focused on more specific indices of brain structure and function (Blokland, de Zubicaray, McMahon, & Wright, 2012). For brain function, as assessed by fMRI or ERP studies in twins and families, heritability tends to be somewhat lower (Jansen, Mous, White, Posthuma, & Polderman, 2015).

Interestingly, several studies have shown that the association between intelligence and brain structure and function is, at least to a large extent, due to shared genetic factors (Brouwer et al., 2014; Koenis et al., 2015; Posthuma et al., 2002).

In genetic epidemiological studies, that make use of the classical twin design, the environmental factors are operationalized as latent factors in e.g., a structural equation model and there is no need for their measurement. The term environmental factors in such models contains all influences on a trait that are modifiable. In this thesis, these factors are labelled under 'environment' at least to the extent that these factors are not themselves caused by genetic factors (Vinkhuyzen, van der Sluis, de Geus, Boomsma, & Posthuma, 2010). When considering cognitive performance, two concrete examples of modifiable environmental influences may be lifestyle factors and current physical health. Modifiable lifestyle factors that are of specific interest include diet and lack of exercise, and physical health includes risk factors for cardiovascular disease, of which chronic hypertension has long been postulated to influence cognitive and brain functioning. For example, people lead increasingly sedentary lifestyles as there is less physical exertion necessary for means of transportation, during day jobs and leisure time activities. It has been suggested that increasing physical activity levels might prevent dementia, and help maintain good brain function in the elderly (Hooghiemstra et al., 2012). In addition, the prevalence of hypertension is increasing, in part because of the increased prevalence of obesity, whereas successful antihypertensive treatment is available. If cognitive functioning is affected by blood pressure, a clear opportunity would present itself. However, the currently hypothesized relationships between exercise and cognitive function and between blood pressure and cognitive function are not supported by uniform and abundant empirical data (Novak & Hajjar, 2010; Singh, Uijtdewilligen, Twisk, van Mechelen, & Chinapaw, 2012).

Measurement of individual differences in cognitive functioning

Individual differences in specific cognitive functions have thus far been studied less frequently and less comprehensively than studies of general intelligence. Different aspects of cognitive performance tend to be positively correlated, but such correlations are not very high. Therefore, measurement of cognitive functions would ideally be performed using instruments that assess the entire cognitive spectrum and distinguish effects of accuracy and speed.

The Brain and Behavior Laboratory of the University of Pennsylvania has developed a test battery that aims to provide exactly this opportunity. For the past few decades, they have been developing and optimizing the web-based Computerized Neurocognitive Battery (CNB), that enables a fast and easy, but yet comprehensive and reliable assessment of the entire range of cognitive functions (Gur et al., 2010). First, the test battery is computerized and web-based, which has several advantages. Test scores will be less influenced by effects due to the researcher collecting the test data, as test instructions are highly standardized and test scores are not sensitive to errors in scoring and calculating. Secondly, there is less of a paper trail compared to traditional pen and paper tests, which is in particular an advantage for studies involving large numbers of participants. Further, the easy and quick assessment creates possibilities for including large numbers of participants in a study.

This makes the CNB a suitable instrument for all studies requiring large sample sizes, for example genome wide association studies. Importantly, test scores on the computerized version compare well to traditional test instruments measuring the same cognitive constructs, and tests have shown to be sensitive to cognitive dysfunctions seen in for example schizophrenia (Gur et al., 2001a; 2001b). In addition, whereas traditional test scores are often based on accuracy, and sometimes use a time limit or response time as the outcome variable, all tests of the CNB (with exception of the motor test) provide an accuracy score and median response time (of all correct responses). Finally, tests were designed to activate specific brain areas: these neuroscience-based tests thus reflect distinct mental processes.

The research in this thesis describes the validation of the Dutch CNB and the analysis of data collected with the CNB in a large Dutch sample, which includes a subgroup of children who are part of BrainScale (van Soelen et al., 2012a): a longitudinal project that follows twin pairs and their siblings from age 9 into adolescence and assesses brain structure, hormone levels and cognition. In this thesis, I seek answers to questions such as: are individual differences in cognitive function and brain development mainly due to genetic factors, and how do effects of genes and environment differ for different cognitive functions? And to what extent do lifestyle and health related factors such as exercise and blood pressure influence cognitive performance?

To address such questions, studies should be carried out in samples that are representative of the general population. Twins are born in every country worldwide, and in many countries, including the Netherlands, and their numbers are increasing (Glasner, van Beijsterveld, Willemsen, & Boomsma, 2013). In the Netherlands, the majority of all twins are dizygotic, meaning that

they originate from two individually fertilized egg cells. This makes them genetically as similar as other brothers and sisters: they share on average 50% of their segregating genes. Other twins are monozygotic, originating from a single fertilized egg cell that, for unknown reasons, splits within the first days after gestation. This results in two individuals who are genetically identical; they share the same DNA sequence. As far as we know, differences at the sequence level are very rare in identical twin pairs (van Dongen, Slagboom, Draisma, Martin, & Boomsma, 2012). Both monozygotic and dizygotic twins share a part of their environment, including prenatal effects, and grow up in the same household and neighborhood, and possibly attend the same school. Of course, both are also exposed to environmental factors that are unique to the individual (for example friends or activities they don't share with their co-twin) and a large proportion of twins will attend separate schools.

The classical twin design uses the difference in genetic similarity between mono- and dizygotic (MZ and DZ) twins to estimate the proportion of total variance in a trait that can be attributed to genetic factors, to shared environmental factors, and to unique environmental factors (Plomin, Defries, Knopik, & Neiderhiser, 2013). In this thesis, some chapters will be based on data from MZ and DZ twins, other chapters include additional family members of twins that took part in the data collection. These additional family members allow for extra hypotheses to be tested, most importantly the hypothesis that cultural transmission may play a role in explaining resemblances of family members.

Outline of this thesis

This thesis addresses neurocognitive test performance across the entire spectrum, across sex and across all ages, and explores how individual differences can be explained by genetic and environmental factors, lifestyle factors in particular.

First, in **chapters 2 and 3** an overview of the main research projects that form the basis of this thesis are described, with respect to the sample of participants and the data collection.

Chapter 4 turns to reading, an important developmental ability that often shows familial risk, where parents with dyslexia have a high change of offspring with reading problems. This chapter explores whether family resemblance for reading (dis)ability might be due to transmission of a genetic liability or due to family environment, including cultural transmission from parents to offspring. In this study, the participants consist of parents and their offspring (twins and siblings).

This design makes it is possible to study cultural transmission from parents to offspring, that is, factors not included in genetic transmission. Further, this design enables a correction of heritability estimates for assortative mating, the phenomenon where partners resemble each other on a given trait. As a result of (strong) assortative mating, siblings (and DZ twins) will be more than 50% genetically alike, which will influence the estimates of the variance components for a trait: estimates of the shared environment will be overestimated and heritability will be underestimated.

The Dutch translation of the CNB first required careful validation, which will be presented in **chapter 5**. In this chapter, I report on reliability indices and effects of age, sex and education on test performance of the CNB. In addition, the possibility of using the CNB as a proxy for traditional intelligence batteries is explored. Next, linear and non-linear effects of aging are presented, and finally heritability of all tests is presented based on analyses in twins, as well as the entire pedigree.

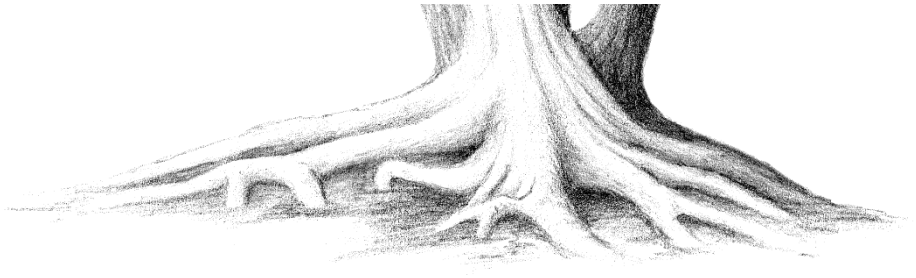
In the next part of the thesis the CNB is used to explore the importance of regular exercise behavior and high blood pressure for individual differences in cognitive functioning across the 17 different domains of the CNB. In **chapter 6** the effect of voluntary exercise behavior on cognitive performance is assessed, while controlling for potential important confounders (sex and age). In **chapter 7** the association between blood pressure and cognitive performance is assessed, again controlling for sex and age.

Neurocognitive testing involved a large sample of children whose brain development had been followed from age 9 onwards. **Chapter 8** examines the heritability and development of subcortical brain volumes during childhood. In a longitudinal twin study, the extent to which subcortical brain volumes are influenced by genetic factors at ages 9 and 12 is explored. This design enables the possibility to test whether new genes are expressed at age 12 and whether there is evidence for genotype by sex interaction. The results are discussed in the broader context of other studies (mainly in adults) on heritability of subcortical structures.

The thesis concludes with a summary and discussion and includes a series of Appendices that detail the data collection and the procedures used to approach and recruit the twin families in these projects, whom I very much want to thank and acknowledge. Without their participation this project would not have been possible.

Chapter 2

Sample description and procedure in studies on the Computerized Neurocognitive Battery in the Dutch population



The Computerized Neurocognitive Battery (CNB), developed by The Brain and Behavior Laboratory of the University of Pennsylvania, was translated into Dutch by the Department of Biological Psychology, Vrije Universiteit Amsterdam. A large data collection project with the Dutch CNB took place between 2010 and 2013. Initially, the CNB was tested in a group of 30 students (pilot phase) and next in a group of over 1100 participants from the Netherlands Twin Register (NTR).

Characteristics of participating families in the CNB study

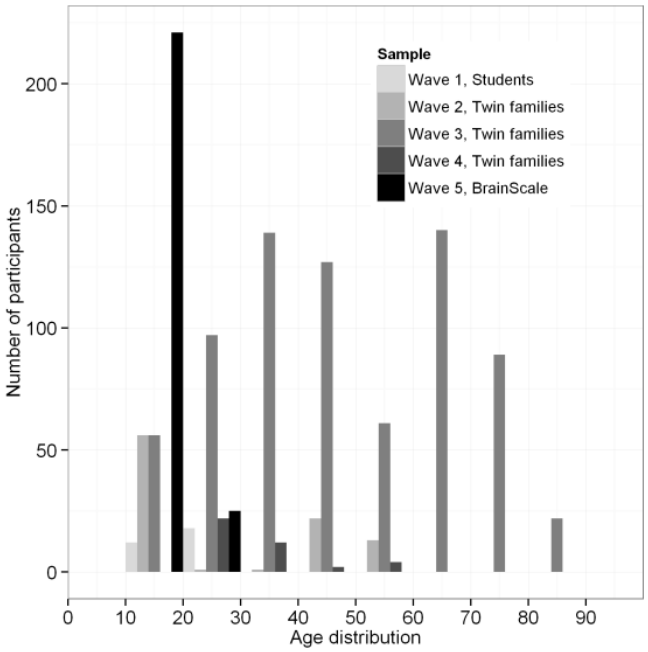
The complete sample with data from the neuropsychological testing project was tested in 5 different waves. During the pilot phase (wave 1), 30 undergraduate students participated in data collection in the lab. Students signed up themselves, and received study credits for participation.

Families from the NTR in the second and third wave were recruited based on living area (provinces near Amsterdam) and age. In the second wave, twin pairs around age 14 and 15 were selected, and their parents and siblings were allowed to participate as well during these home visits (in total 93 participants from 26 families). The third wave recruited elderly participants of the NTR, preferably with available genotype data in the database, and again all family members were invited to participate (in total 731 participants from 276 families).

A fourth wave consisted of a group of 20 twin pairs from an MRI study in twin pairs discordant for obsessive compulsive symptoms (den Braber et al., 2013b) who took part in cognitive testing while they were in the Amsterdam Medical Center for the appointment.

The final group consisted of participants of the BrainScale study (wave 5). The group of twins and siblings in the BrainScale project were acquired throughout the Netherlands in a combined research project between the NTR and the Brain Center Rudolf Magnus from the University Medical Center Utrecht. These participants were selected from the NTR in 2005 when the first data collection started, and were currently participating in the third wave of data collection in this project. In total this sample consisted of 176 twins and 70 siblings (in total 139 females, 107 males, mean age 17.45, $SD = 1.32$). A detailed description of the sample and data collection during the third assessment can be found in Chapter 3.

Figure 1. Number of participants per age cohort in the different waves of data collection.



In total, neuropsychological data of 1140 participants were collected. The final sample was comprised of 668 female and 472 male participants between the ages of 10 and 86 (mean age = 37.73, *SD* = 20.86). A graphical indication of the age distribution is given in Figure 1. The twin-family part of the sample consisted of 1110 participants from 431 NTR families. The majority of this sample was part of a twin pair (618). The rest of the sample were parents of twins (126 fathers, 160 mothers), siblings of twins (144), spouses of twins (43) and siblings (10), and children of twins (6) and siblings (3).

Procedure

These studies were approved by the Medical Ethics Review Committee of the Vrije Universiteit Medical Center Amsterdam and the Central Committee on Research Involving Human Subjects (wave 5), and research procedures were performed in accordance with the Declaration of Helsinki. Examples of letters, brochures and documents can be found in the Appendices (1 to 5).

Invitation

All participants were sent an invitation letter including a brochure (Appendix 1 and 2) with detailed information about the study and procedures. Besides general information about the study purpose and procedure, it stated that participants would receive gift vouchers, and were compensated for travel expenses. Further, a summary of their results on the computerized tests would be mailed to them afterwards as a token of appreciation (Appendix 4).

Following this letter, participants were contacted by phone to inquire whether they had received the letter and were willing to participate. A telephone protocol was used that specified for each moment of contact: the date and time, what was discussed, whether each participant of a family was willing to participate, when to call back (if necessary), the reason for not participating (if applicable), the confirmed date of the appointment, and the date of sending the confirmation letter including study materials. Participants were provided the option to choose whether they preferred a home visit or an appointment at the Vrije Universiteit Amsterdam (VU) laboratory, except for participants of the MRI study which was always taking place in the hospital. When families agreed to participate, each individual was sent a confirmation of the appointment and consent forms (Appendix 3). Consent forms were signed by parents as well as children when the child was younger than 18.

Procedures

The data collection took place at the participants' home (536), the VU laboratory (318), the Amsterdam Medical Center (40), and the University Medical Center Utrecht (246). Depending on the wave of data collection, experimental procedures varied slightly. Measurements and instruments are described below.

The experimental procedure would always start with an explanation. Next, basic information about medication use and education was gathered. After participation, adults were asked to fill in the latest survey of the Adult NTR at home and all NTR participants were asked to collect buccal epithelium for DNA isolation.

In the first wave (the university students) and the second wave (young twins and their parents and siblings) the procedure started with the reading test. Then participants completed the Computerized Neurocognitive Battery (CNB). During the first and second designated breakpoint of the CNB, blood pressure was measured. During the third breakpoint a standardized interview was administered which included questions about education or occupation, sleep, smoking, drinking, exercise, time spent walking and biking, menstruation and general health (Appendix 5). The procedure took on average 2 hours per person.

After the first two waves, a few measurements were added to the procedure. First, cardiac autonomous nervous system (ANS) activity was recorded by a non-invasive device. Electrodes of the VU University Ambulatory Monitoring System were placed on the back and chest (VU-AMS, de Geus, Willemsen, Klaver, & Van Doornen, 1995; Willemsen, de Geus, Klaver, Van Doornen, & Carroll, 1996). This recording provides measurements of for example heart rate (interbeat interval and variability), heart rate variability, T-wave amplitude and pre-ejection period. During a 5-minute measurement, participants watched a calming movie of a beach during which a baseline recording of autonomous nervous system activity was made. Also an additional measurement of blood pressure was taken during this resting baseline. Measurements of length and weight were taken, and hip and waist circumference were measured.

Participants in wave 4 (opposite-sex twins) and wave 5 (BrainScale twins and siblings) also had an MRI scan of the brain. When their co-twin was in the scanner, participants in wave 4 would follow the same CNB testing protocol as participants from the first and second wave (thus without ANS recording). The procedure of wave 5 will be described in detail in Chapter 3.

Instruments and measurements

Behavioral data

Questionnaire

Participants from the NTR were asked to fill in the most recent survey ('survey 8'). This questionnaire contains questions on emotional and behavior problems (ASR), well-being, lifestyle, exercise behavior, sedentary behavior, and family functioning (Willemsen et al., 2013).

Interview

In addition to the CNB, participants were asked about, or filled out a questionnaire on lifestyle (e.g., drinking, smoking, exercise behavior).

Education

Prior to the CNB administration participants were asked about their own educational background, as well as that of their parents. Level of education was defined as the sum of years involved in elementary, secondary and higher education if the educational curriculum (per year) was completed.

Medication

In the confirmation letter, participants were asked to show, or bring with them to the appointment, packages of medication they (recently) used. The brand and substance name, frequency and reason for using the medication were registered.

Physical examination

Length and weight; hip and waist circumference

Measurements of length and weight were obtained by measurement at the day of testing (wave 3, 4 and 5). Before measurement, participants were asked to take off their shoes. In wave 3 and 4, hip and waist circumference were measured with a tape-measure.

Blood pressure and heart rate

Blood pressure and heart rate were measured in a sitting position with an Omron automatic blood pressure monitoring device. The cuff was attached to the non-dominant arm. Depending on the wave of data collection, measurements were taken once (wave 5), twice (wave 1, 2 and 4), or three times including a baseline measurement (wave 3).

Neuropsychological assessment

Computerized Neurocognitive Battery (CNB)

The CNB (Computerized Neurocognitive Battery) is a Dutch translation of the current web-based CNB (Gur et al., 2012). It includes a total of 17 tests, resulting in measures of performance accuracy (the percentage or number of correct responses) and response time in five global cognitive functions. These functions, their corresponding test and which cognitive domain they specifically measure are given in Table 1. For more detailed descriptions of these tests we refer to Gur et al., (2010, 2012) and the Supplementary materials of Chapter 5.

Prior to the administration of each of the CNB's tests, instructions were read out loud to the participant by the administrator, after which participants were provided with practice trials (memory tests and the Conditional Exclusion Test excluded). Practice trials had to be completed successfully before the actual trials started. The administrators kept track of whether the participant's test scores were valid or not, for example based on the participant's (lack of) motivation, the presence of distracters, or computer issues. On top of this, automated test score validation occurred upon upload to the Pennsylvania web servers that host the CNB. At the Pennsylvania web servers automated scores were generated for a number of variables describing the performance on the various subtests in great

detail. The main outcome variables extracted from these scores, reflecting overall accuracy and speed of test performance, are listed in Table 2, including the mean score (and SD) for the participants from wave 1-4.

Reading ability

The participants were instructed to read out loud, within one minute, as many words as possible from a card with 116 words. The list was adapted from the "Three Minutes Reading Task", which is frequently used in the Dutch educational system (Cito, 1995).

Table 1. Overview of global cognitive functions, corresponding tests and the cognitive domain they measure

Cognitive function	Test name	Cognitive domain measured
Executive-control	Continuous Performance Test	attention
	Letter N-Back Test	working memory
	Conditional Exclusion Test	abstraction & mental flexibility
Episodic memory	Face Memory Test	face memory
	Word Memory Test	verbal memory
	Visual Object Learning Test	spatial memory
Complex cognition	Matrix Reasoning Test	nonverbal reasoning
	Verbal Reasoning Test	language reasoning
	Line Orientation Test	spatial ability
Social cognition	Emotion Identification Test	emotion identification
	Emotion Differentiation Test	emotion differentiation
	Age Differentiation Test	age differentiation
Sensorimotor speed	Motor Praxis Test	sensorimotor speed
	Finger Tapping Test	motor speed

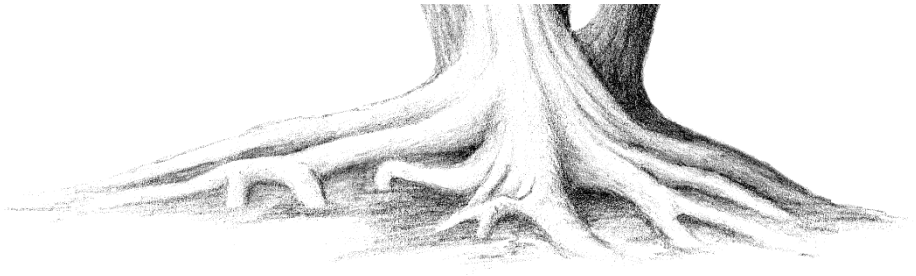
Table 2. Overview of main output variables (wave 1-4)

Task or measure used	Main output phenotype	N	Mean ± SD
Cognition			
Reading ability (1 minute)	Total correct words	892	94.13 ± 14.72
CNB, Cognitive domain			
Attention	True positive responses (#)	886	1.78 ± 0.80
	Median RT (ms)	886	2923.23 ± 1486.68
Abstraction and mental flexibility	Correct categories (#)	884	54.82 ± 5.64
	Median RT (ms)	884	486.55 ± 50.92
Working memory	True positive responses (#)	877	18.66 ± 1.97
	Median RT (ms)	876	542.73 ± 121.65
Face Memory	Total correct (#)	882	31.45 ± 3.54
	Median RT (ms)	882	2013.87 ± 555.49
Face Memory -delayed	Total correct (#)	880	31.97 ± 3.57
	Median RT (ms)	880	1869.45 ± 503.48
Verbal memory	Total correct (#)	884	36.16 ± 2.93
	Median RT (ms)	884	1611.63 ± 390.10
Verbal memory - delayed	Total correct (#)	883	34.80 ± 3.36
	Median RT (ms)	883	1587.70 ± 396.61
Spatial memory	Total correct (#)	875	15.84 ± 2.32
	Median RT (ms)	875	2048.73 ± 569.66
Spatial memory - delayed	Total correct (#)	872	15.25 ± 2.36
	Median RT (ms)	872	1885.24 ± 537.35
Nonverbal reasoning	Total correct (#)	887	13.06 ± 5.21
	Median RT (ms)	886	10861.37 ±7341.78
Language reasoning	Percentage correct (%)	878	69.89 ± 21.07
	Median RT (ms)	877	8399.40 ± 3339.09
Spatial ability	Total correct (#)	877	12.66 ± 3.76
	Median RT (ms)	877	10707.51 ± 4101.30
Emotion Identification	Total correct (#)	891	31.71 ± 3.52
	Median RT (ms)	891	2353.91 ± 725.51
Emotion Differentiation	Total correct (#)	889	27.71 ± 3.58
	Median RT (ms)	889	3862.67 ± 1445.49
Age Differentiation	Total correct (#)	879	26.57 ± 4.03
	Median RT (ms)	879	3424.03 ± 1579.68
Sensorimotor speed	Total correct (#)	888	19.94 ± 0.41
	Median RT (ms)	888	822.75 ± ± 235.47
Motor speed	Total taps in 1 minute (#)	882	109.66 ± 15.75
Physical examination			
Height	Centimeters	769	172.91 ± 9.82
Weight	Kilogram	769	76.06 ± 14.99
Diastolic blood pressure	mmHG	890	76.46 ± 11.28
Systolic blood pressure	mmHG	890	129.00 ± 18.10
Heart rate	Beats per minute	890	67.30 ± 10.35
Hip circumference	Centimeters	770	100.20 ± 0.62
Waist circumference	Centimeters	770	83.72 ± 14.03

Note: Median RT refers to the median response time per individual for all correct responses on a test. Mean RT refers to the mean of RT medians across individuals

Chapter 3

Sample description and data collection in the BrainScale study: an Adolescent Longitudinal Twin Study into the Genetic Etiology of Individual Differences



The group of twins and their siblings in the BrainScale project forms a sample that is followed longitudinally since age 9. The BrainScale project is a cooperation between the NTR and the UMC Utrecht, studying the influences on brain and cognition throughout healthy development. Participants were invited for the first assessment in 2004 when the data collection started. This was around the 9th birthday of the twins, and over a 1.5 year period 330 children from 112 families participated. Between 2007 and 2009 the second assessment took place, this time around the 12th birthday of the twins. In total, 261 children were willing to participate again. Details of the data collection at ages 9 and 12 years are described in the dissertations of M. van Leeuwen (2008), J. S. Peper (2008) and I. L. C. van Soelen (2011). An overview of the project was also published in 2012 (van Soelen et al., 2012a). As part of this thesis project, twins and siblings were invited for the third time. This chapter provides the details of the third wave of data collection that took place between 2012 and 2014.

Participants and return rate

All families that have participated at the first assessment (112 families with 224 twins and 96 siblings) were invited to participate again, with the exception of 3 families who had indicated that they did not want be approached for follow-up. In total, 246 participants from 89 families agreed to participate (77% return rate) in this third wave. Reasons for not participating were: too busy with school and/or work (21), one or more family members did not want to participate (43), living abroad (5), or unable to reach by mail and phone (3). Of all individuals that originally participated but did not return at the third assessment, 37 were only part of the first wave. On the other hand, 31 individuals were not part of the second wave but decided to participate again (see Table 1 for the number of participants included at the neuropsychological assessment and MRI scans at three assessments). In total, a large number of participants have participated in all three assessments: 154 participants completed the MRI procedure and 212 participants finished cognitive testing on all time points.

Participants in the third assessment consisted of 176 twins (95 female, 81 male) and 70 siblings (44 female, 26 male) from 89 families. Twin pairs were almost equally distributed for zygosity: monozygotic male (16), dizygotic male (17), monozygotic female (22), dizygotic female (18) and dizygotic opposite sex (16). Twin pairs were incomplete in two families. Twins were 16 or 17 years old at the day of testing (mean age 16.85, SD = 0.36). The mean age of the siblings was 19.26 (SD = 1.30).

Table 1. Number of participants included in the neuropsychological tests and MRI scan at the first, second and third assessments.

Assessment	Neuropsychological testing	MRI scan
1+2+3	212	154
1+2	47	37
1+3	31	73
2+3	2	7
Only 1	37	37
Only 2	-	1
Only 3	1	7
Total	330	316

Procedure

This study was approved by the Central Committee on Research involving Human subjects of the Netherlands (CCMO), and research procedures were performed in accordance with the Declaration of Helsinki.
Examples of letters, brochures and documents can be found in the appendices (6 – 10).

Invitation

All participants and their parents were sent an invitation letter including a brochure (Appendix 6-8) with more specific information about the study and procedure. Besides general information about the study purpose and procedure, it stated that participants would receive gift vouchers, and compensation for travel expenses. Participants were asked to come to the University Medical Center Utrecht (UMCU) for a six-hour visit. Further, as a token of appreciation, a summary of their results on the computerized tests (Appendix 4) and a printed image of their brain from the MRI scan would be mailed to them afterwards.

Following the invitation letter, participants were contacted by phone to inquire whether they had received the letter and were willing to participate. A telephone protocol was used that specified for each moment of contact: the date and time, what was discussed, whether each participant of a family was willing to participate, when to call back (if necessary), the reason for not participating (if applicable), the confirmed date of the appointment, and the date of sending the confirmation letter including study materials.

The twins and siblings in the BrainScale study could participate together or make separate appointments if this was preferred, for example because of different school or work schedules. In addition, participants could choose to start in the morning or the afternoon, and could take part in the study on weekdays as well as during the weekend.

Confirmation

When families agreed to participate, each individual was sent a confirmation of the appointment and consent forms (Appendix 9 and 10). This confirmation letter further included the additional materials (for collection of cheek swabs and/or urine and saliva) and documents (instructions for buccal and/or hormone collection, questionnaires, MRI checklist, directions to the hospital).

Experimental procedures of the BrainScale study

The data collection took place in the UMCU, similar to the second assessment. At the first assessment, the cognitive test protocol took place at the Vrije Universiteit Amsterdam and the MRI scan and physical examinations were made at UMCU. For a family of three children, the test day lasted approximately six hours (including lunch, test protocol is described in Table 2). Depending on the availability of the MRI scanner and the preference of the participants, the protocol for a family of three could be as follows: 1) morning neuropsychological tests, afternoon MRI scans (9:00-15:00); or 2) afternoon MRI scans, then neuropsychological tests (13:00-19:00); or 3) afternoon neuropsychological tests then MRI scans (14:00-20:00). If participants came alone, or with two persons, the protocol lasted 4 to 5 hours.

Instruments and measurements

Changes and continuity in data collection

Compared to the data collection on the first two assessments, most conditions and instruments have stayed the same on the third assessment, to the extent that this was possible. MRI scanners and protocols did not change, physical assessments were the same and participants collected urine and saliva at home in the same way on two consecutive days. However, the cognitive testing was adapted, so that psychometric IQ scores were now assessed by the test for adults and the collection of neurocognitive tests was replaced by an extensive computerized cognitive test battery. Details of the data collection of the third assessment will be described below. The main outcome variables of cognition, health and lifestyle variables collected at the third assessment are listed in Table 3, including the mean and SD separately for the twins and siblings.

Neuropsychological assessment

The Computerized Neurocognitive Battery

The test version and conditions of the CNB were similar to the settings as are described in Chapter 2. Within approximately 1.15 hour, performance scores - both accuracy and speed - on 17 tests were acquired.

Table 2. Test protocol for the third assessment.

Welcome at the UMC Utrecht	
Collection of questionnaires, saliva and urine samples, cheek swabs	15 minutes
Explain procedure and sign consent forms	15 minutes
Neuropsychological test protocol	
WAIS-III Intelligence test	45 minutes
Break	15 minutes
Computerized Neurocognitive Battery	75 minutes
Lunch break	45 minutes
Afternoon program	
MRI scan	50 minutes
Corsi task	5 minutes
Iowa Gambling task	10 minutes
One minute reading test	2 minutes
Physical examination (length, weight, blood pressure, Tanner stage)	10 minutes

Note: the order of the afternoon program was different for each participant because of MRI scanner availability, and could vary depending on the length of the neuropsychological test assessment.

IQ

A selection of subtests of the Wechsler Adult Intelligence Scale – Third version (WAIS-III, Wechsler, 1997) was administered (in order of administration: Vocabulary, Block Design, Similarities and Matrix Reasoning). Raw test scores were standardized based on the age of the participant. Then a correction for the number of excluded subtests was performed to be able to calculate IQ scores: the sum of the standardized score of Vocabulary and Similarities was multiplied by 2.5 (verbal IQ), the sum of standardized scores of Block Design and Matrix Reasoning was multiplied by 3 (performance IQ), the sum of verbal and performance IQ (total IQ).

Corsi block task

The Corsi block task measures short-term spatial memory. On a computer screen, nine blocks were presented, that were scattered across the screen. These blocks would light up for one second, and the participants were instructed to click, using the computer mouse, the previously lighted blocks in exactly the same order. Immediately after the trial, participants received feedback on whether their response was correct or not. Starting with 2 trials with a length of 2 blocks, trials would increase in length by one block if the participant was successful on at least one of these trials, with a maximum of all 9 blocks. If both trials of a series were unsuccessful the test was terminated. The performance measure on this task was the total number of trials the participant completed successfully.

Iowa gambling task

The Iowa gambling task measures decision making and self-regulation in a delayed reward task, where the participant has to make decisions that may be more advantageous on either the short-term or the long-term (Bechara, Damasio, Damasio, & Anderson, 1994). Performance on this test is related to damage to the frontal cortex, addiction, and risk-taking behavior in adolescents (Brevers, Bechara, Cleeremans, & Noel, 2013; Crone & van der Molen, 2004). Four decks of cards were presented on the computer screen, each deck with a different reward schedule. Participants had to click on a deck in order to receive either an award (most often) or a penalty. The magnitude of the reward and the penalty depended on the deck. The deck that has high reward on the short-term will also include high penalty, whereas the deck with low reward on the short-term includes less penalty, making this more advantageous on the long-term. The task started with a fictional $2000 and the participants were instructed to choose decks while trying to earn as much money as they could. The task finished after 100 cards. The outcome was the total amount of money at the end of the task.

Reading ability

The participants were instructed to read out loud, within one minute, as many words as possible from a card with 116 words. The list was adapted from the "Three Minutes Reading Task", which is frequently used in the Dutch educational system (Cito, 1995).

Behavioral data

Questionnaires

Twins, siblings and their parents were asked to fill in standardized questionnaires. Participants filled in the Dutch Health and Behavior Questionnaire (DHBQ), which includes questions on emotional and behavior problems (Youth Self-Report, Achenbach, 1991), well-being, lifestyle, exercise behavior, sedentary behavior, and family functioning (van Beijsterveldt et al., 2013). Parents were asked to fill in the Adult Self-Report (ASR, Achenbach & Rescorla, 2003) about themselves (80 fathers, 88 mothers) and the Child Behavior Checklist (CBCL, Achenbach & Rescorla, 2001) about their children. We received 241 surveys from the participants and 423 from their parents. The DHBQ contains questions about the specific type of education.

Smoking behavior

At the start of the testing day, participants were individually (in private) asked about their current or previous smoking behavior. When they mentioned they

had smoked or were still smoking, the researcher asked for the exact age of initiation (as close as possible), frequency of smoking and the age of quitting.

Education

Participants were asked about their own educational background, as well as that of their parents. Level of education was defined as the sum of years involved in elementary, secondary and higher education if the educational curriculum (per year) was completed.

Medication

Participants were asked to bring with them any packages of medication they used at the moment of testing, or had very recently used, but in this young group medication use was rare and complete information was not always provided by the participants. Concerning use of contraceptive pills, girls were asked to mention the brand of the pill on the questionnaire about urine collection.

Physical examination, hormone and buccal sample collection

Length and weight

Participants were asked to take off their shoes before body height and weight were measured.

Blood pressure and heart rate

Blood pressure and heart rate were measured in a sitting position with an Omron automatic blood pressure monitoring device. The cuff was attached to the non-dominant arm. To measure blood pressure and heart rate, participants were asked to remain relaxed and still, and refrain from talking and laughing.

Tanner stage

Tanner stages were determined with a self-report questionnaire, on the basis of secondary sexual characteristics using the five stages of development devised by Marshall & Tanner (1969; 1970, see for data on this sample Koenis et al., 2013). After explanation by the researcher, the researcher left the room and participants were asked to fill in their developmental status on black and white photographs of the different pubertal stages. Stage 1 represents no pubertal development and full maturation is represented at stage 5. Girls were asked about breast development and pubic hair growth; boys were asked about genital development and pubic hair growth. In boys, genital stage was divided in penis and testes development. Testes volume was reported on a 4-item scale (compared size with ovals: 1) 1-3 ml; 2) 4-6 ml; 3) 7-11ml; 4) 12-25 ml) and boys were also asked to rate testes volume with an orchidometer.

Table 3. Overview cognition, health and lifestyle variables collected at the 3th assessment. Means and SD are given separately for twins and siblings.

Task or measure used	Main output phenotype	N total (twins/sibling	Mean ± SD twins	Mean ± SD siblings	N total 1st	N total 2nd
Cognition						
Age	Years	176 / 70	17.0 ± .4	19.3 ± 1.3		
Intelligence (WAIS)	Total IQ	176 / 70	10.4 ± 13.1	107.5 ± 14.1	224 / 102	178 / 81
	Verbal IQ	176 / 70	105.3 ± 13.6	108.2 ± 15.6	224 / 102	178 / 81
	Performance IQ	176 / 70	102.1 ± 12.0	108.0 ± 13.7	224 / 102	178 / 81
Reading ability (1 minute)	Total correct words	176 / 69	91.2 ± 14.8	95.8 ± 14.5	209 / 85	167 / 76
Spatial memory (Corsi)	Total correct items	176 / 70	9.2 ± 1.6	9.2 ± 1.7	221 / 101	173 / 79
Decision making (Iowa gambling)	Total gain across items	176 / 70	1678.1 ± 619.7	1858.6 ± 748.1	-	-
CNB, Cognitive domain						
Attention	True positive responses(#)	176 / 70	54.4 ± 4.9	55.9 ± 4.5	-	-
	Median RT (ms)	176 / 70	495.7 ± 43.2	486.4 ± 38.2	-	-
Abstraction / mental flexibility	Correct categories (#)	176 / 70	2.1 ± .6	2.1 ± .6	-	-
	Median RT (ms)	176 / 70	2359.2 ± 882.5	2471.7 ± 757.3	-	-
Working memory	True positive responses(#)	175 / 70	19.0 ± 1.7	19.1 ± 1.1	-	-
	Median RT (ms)	175 / 70	519.0 ± 107.7	524.4 ± 84.0	-	-
Verbal memory	Total correct (#)	176 / 70	36.7 ± 2.5	36.7 ± 2.6	-	-
	Median RT (ms)	176 / 70	1422.0 ± 199.5	1376.7 ± 198.0	-	-
Verbal memory - delayed	Total correct (#)	176 / 70	36.0 ± 2.9	35.8 ± 2.9	-	-
	Median RT (ms)	176 / 70	1402.5 ± 232.1	1355.0 ± 217.2	-	-
Face Memory	Total correct (#)	176 / 70	31.2 ± 3.5	31.8 ± 3.3	-	-
	Median RT (ms)	176 / 70	1959.0 ± 49.2	1945.4 ± 491.7	-	-
Face Memory -delayed	Total correct (#)	176 / 70	32.7 ± 3.5	32.4 ± 3.4	-	-
	Median RT (ms)	176 / 70	173.4 ± 425.1	1691.4 ± 325.7	-	-
Spatial memory	Total correct (#)	176 / 70	16.4 ± 2.2	16.5 ± 2.1	-	-
	Median RT (ms)	176 / 70	1696.9 ± 401.9	1707.1 ± 366.1	-	-
Spatial memory - delayed	Total correct (#)	176 / 70	15.8 ± 2.3	16.0 ± 2.7	-	-
	Median RT (ms)	176 / 70	1538.7 ± 353.5	1562.5 ± 297.5	-	-
Nonverbal reasoning	Total correct (#)	176 / 70	16.4 ± 4.0	17.7 ± 3.9	-	-
	Median RT (ms)	176 / 70	10195.2±5135.5	11229.4 ± 5844.9	-	-
Language reasoning	Percentage correct (%)	176 / 70	65.4 ± 18.7	68.6 ± 19.5	-	-
	Median RT (ms)	176 / 70	8739.5 ± 3325.0	8513.8 ± 3271.0	-	-

Table 3 – continued.

		Main output phenotype	N total (twins/siblings)	Mean ± SD twins	Mean ± SD siblings	N total 1st	N total 2nd
Spatial ability		Total correct (#)	176 / 70	13.6 ± 3.5	14.3 ± 3.7	-	-
		Median RT (ms)	176 / 70	9624.9 ± 2828.7	10177.5 ±2443.2	-	-
Emotion Identification		Total correct (#)	176 / 70	33.6 ± 2.9	33.8 ± 2.8	-	-
		Median RT (ms)	176 / 70	1962.0 ± 368.7	202.6 ±	-	-
Emotion Differentiation		Total correct (#)	176 / 70	28.7 ± 3.2	29.5 ± 2.7	-	-
		Median RT (ms)	176 / 70	3164.2 ± 882.1	3266.9 ± 814.6	-	-
Age Differentiation		Total correct (#)	176 / 70	27.5 ± 3.3	28.1 ± 3.0	-	-
		Median RT (ms)	176 / 70	2546.5 ± 842.1	2626.9 ± 813.4	-	-
Sensorimotor speed		Total correct (#)	176 / 70	20.0 ± 0.0	20.0 ± 0.0	-	-
		Median RT (ms)	176 / 70	472.5 ± 1044.0	68.6 ± 102.6	-	-
Motor speed		Total taps in 1 minute	176 / 69	38.7 ± 78.3	59.6 ± 7.0	-	-
Physical examination							
Height		Centimeters	176 / 70	173.6 ± 8.2	175.1 ± 9.5	218 / 99	174 / 78
Weight		Kilogram	176 / 70	64.1 ± 9.5	7.1 ± 1.6	218 / 99	174 / 78
Diastolic blood pressure		mmHG	176 / 70	71.0 ± 9.4	72.7 ± 9.6	-	-
Systolic blood pressure		mmHG	176 / 70	129.1 ± 13.4	133.6 ± 14.4	-	-
Heart rate		Beats per minute	176 / 70	64.7 ± 11.6	65.1 ± 12.0	-	-
Smoking		Yes, no, stopped (%)	133 / 54	23, 71, 6	26, 70, 4	-	-
Puberty							
Tanner (stage 1:6)	Boys	Penis development	77 / 26	1, 2, 13, 42, 19	0, 0, 3, 11, 12	108 / 44	83 / 31
		Pubic hair	78 / 26	1, 0, 1, 1, 46, 13	0, 0, 0, 3, 15, 8	107 / 44	84 / 31
		Testis size	78 / 26	0, 1, 34, 34	0, 2, 9, 15	-	-
	Girls	Breast development	96 / 44	0, 0, 0, 40, 56	0, 0, 0, 6, 38	109 / 54	86 / 43
		Pubic hair	96 / 44	4, 4, 1, 13,55, 19	0, 0, 1, 5, 20, 18	108 / 53	80 / 44
Questionnaires							
Child behavior checklist (CBCL)		Report by parents	174 / 69			158 / 92	198 / 81
Dutch health behavior questionnaire (DHBQ)		Report by participant	173 / 68			-	175 / 81
Adult self-rating (ASR)		Report by parents	80 fathers, 88 mothers			-	-

Assessment of hormone levels

Similar to the previous occasions, participants were asked to collect saliva and morning urine for the assessment of reproductive hormonal levels (LH, FSH and estrogen from urine), and testosterone from saliva (Koenis et al., 2013). Because of their hormonal cycle, girls were asked to collect urine and saliva for assessment of hormone levels at a specific time point during their menstrual cycle, namely in the early follicular phase when hormone levels are relatively low. To minimize effects of contraceptive pills, they were asked to collect morning urine and saliva at the 6th and 7th day of their menstrual cycle. They could send the samples by mail to the Vrije Universiteit Amsterdam (VU). Boys were asked to collect samples in the two days prior to the test day and bring the samples to the UMCU. They could also send the samples by mail.

Participants were asked to fill in the time and date of the collection. Girls had a more extensive questionnaire that included questions about their menstrual cycle and contraceptive pill use.

Samples were stored in the refrigerator as soon as possible. After pipetting a small volume of urine for hormone level assessment, all samples were stored at -20°C at the VU. Saliva samples were stored at - 20°C as soon as possible.

Urine samples were collected for 166 twins and 62 siblings. Saliva samples were collected for 166 twins and 63 siblings.

Buccal epithelium

Participants were asked to collect buccal swabs in the morning and evening, on 2 days. They were instructed not to eat, drink or brush their teeth prior to collection. Collection of buccal epithelium was done by rubbing cotton buds along the inside of the mouth. The swabs were then placed in a tube with buffer. Participants could bring the tubes with them to the UMCU at the test day, or send the tubes by postal mail. DNA was isolated at the *Avera Institute for Human Genetics,* Sioux Falls. All samples were tested on single nucleotide polymorphic (SNP) markers to establish zygosity (van Beijsterveldt et al., 2013).

MRI scan protocol

At the third assessment of the BrainScale study, the same scan parameters were used as in the previous two test assessments (see Table 4) where participants were scanned at a Philips Achieva scanner at 1.5T at all measurements (Brouwer et al., 2012; Peper et al., 2008; van Soelen et al., 2012b). At the start of the test day, the scan procedure was explained to the participants. They were allowed to watch a movie or listen to music during the structural scans. During the last 10 minutes the resting state functional MRI scan was made, for which participants

were asked to close their eyes and try to think of nothing specific. Afterwards, participants were asked if they had remained awake during the scan and if they managed to think of nothing specific. The total scan protocol took about 45-50 minutes per child.

Dental braces

Presence of dental wires (top / bottom / both) was asked at the start of the test day since this may distort the MRI image.

Table 4. Scan protocol and MR acquisition details

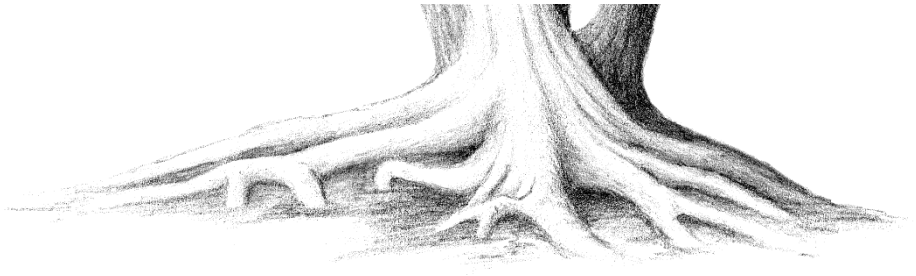
	MR acquisition details	Duration
1.	Scout scan, sagittal T1 weighted; TR = 13 ms; TE = 4.6 ms	1 min.
2.	Dual Echo – Turbo Spin Echo (DE-TSE) clinical scan, transversal T2 weighted; TR=2200 ms, TE= 9ms; 19 slices of 6 mm; slicegap 1 mm; flipangle 90°; FOV: 230 mm / 90%	2 min.
3.	Three Dimensional - Fast Field Echo (3D-FFE) T1 weighted scan; coronal; 256 x 212 acquisition matrix; 256 × 256 reconstruction matrix; 160–180 contiguous slices of 1.2 mm; TR = 30 ms; TE = 4.6 ms; flipangle 30°; FOV: 256 mm / 65%	7 min.
4.	Diffusion Tensor Imaging (DTI) scan using SENSE coil; 15-64 Directions; b-factor 1000; 60 slices of 2.5 mm; slicegap 0.0 mm; 96 x 96 acquisition matrix; reconstruction matrix 128 x 128; flipangle 90°; FOV: 240 mm; TE = 60–88 ms; no cardiac gating.	11 min.
5.	Magnetization Transfer Imaging (MTI) scan; 60 transverse slices of 2.5 mm; slicegap 0.0 mm; 128 x 96 acquisition matrix; reconstruction matrix 128 x 128; flip angle 8°; FOV: 240 mm / 78%; TR = 37.5 ms; TE = 3.73 ms.	7 min.
6.	Dual Echo - Turbo Spin Echo using SENSE, transversal T2 weighted; parallel imaging, sensefactor 2; TR/TE1/TE2 6000/18/80 ms; 120 slices of 2 mm; 256 x 195 acquisition matrix; reconstruction matrix 256 x 256; slicegap 0 mm; flipangle 90° ; FOV: 240 mm / 79 %	7 min.
7.	Resting State Scan using SENSE coil; parallel imaging, sensefactor 1.8; 3D T2* weighted field echo EPI (FEEPI) scan; Timeseries 800-1200 scans, single scan duration 0.5-0.7 sec; sagittal scan orientation; acquisition matrix: 64 x 33; reconstruction matrix: 64 x 64; flip angle 9°; 36 slices; FOV: 256 mm; 4 mm isotropic voxels; TR=21.1 ms; TE= 31.10 (shifted echo).	10 min.

Note: at first assessment, T2 (scan 6) and rs-fMRI (scan 7) were not included; all other scan parameters were the same.

Chapter 4

Genetic transmission of reading ability

This chapter is published as: Swagerman, S. C., van Bergen, E., Dolan, C. V., de Geus, E. J. C., Koenis, M. M. G., Hulshoff Pol, H. E., & Boomsma, D. I. Genetic transmission of reading ability. *Brain & Language*. Accepted for publication.



Introduction

Dyslexia, usually conceptualized as the lower tail of the word reading-ability distribution, tends to run in families. Children of dyslexic parents, as well as siblings of dyslexic children, have a higher change of developing dyslexia themselves (Snowling, Gallagher, & Frith, 2003; Torppa, Lyytinen, Erskine, Eklund, & Lyytinen, 2010; van Bergen, van der Leij, & de Jong, 2014; Vogler, Defries, & Decker, 1985). Their heightened risk is utilized in studies seeking neuro-anatomical, neuro-functional, cognitive, and environmental precursors of dyslexia. For instance, it has been found that children with familial risk have altered structural brain networks in language areas (Hosseini et al., 2013) and impaired auditory processing (Lyytinen et al., 2005; van der Leij et al., 2013). Despite the ubiquitous use of this familial-risk design in reading and language research, what remains to be resolved is the nature of the transmitted risk (van Bergen et al., 2014). A mainly genetic cause for reading ability and disability implies that parents with reading problems pass on less advantageous genes, whereas a mainly environmental explanation would mean that these parents create a less advantageous home-literacy environment. Which of these two is the main driver has consequences for the interpretation of dyslexia precursors seen in at-risk children.

Evidence for the genetic explanation comes from twin and family studies, which indicate that genetic factors explain a large part of individual differences in children's word-level reading ability (henceforth called 'reading ability'). Reading ability (or decoding) is typically assessed by asking participants to read a list of words, and measuring accuracy or a combination of accuracy and speed (called fluency). The heritability of dyslexia and reading ability is high (60-70%) from a young age onwards (Hawke, Wadsworth, & Defries, 2006; Kovas et al., 2013). The heritability might be higher for timed compared to untimed tasks (Petrill et al., 2012). The current study was conducted in a large Dutch twin-family sample. The Dutch orthography (writing system) is less complex compared to English (Seymour et al., 2003). Hence, accuracy is close to ceiling and reading ability in Dutch is typically measured using fluency tasks (Patel, Snowling, & de Jong, 2004). This might be related to the even higher heritability found for reading ability in Dutch children (around 80%, van Leeuwen, van den Berg, Peper, Hulshoff Pol, & Boomsma, 2009). However, Samuelsson et al., (2008) did not find differences in heritability between orthographies. Alternatively, the high heritability found in the Netherlands may be due to the egalitarian educational system, which reduces environmental variance. Besides children, our study also includes adults. In adults, the heritability of reading has hardly been studied.

One study in adult men found somewhat lower though still robust heritability estimates (45%, Kremen et al., 2005).

Evidence for environmental influences comes from twin studies, which sometimes find a significant influence of the environment that is shared between twins (Olson, Keenan, Byrne, & Samuelsson, 2014; Taylor, Roehrig, Hensler, Connor, & Schatschneider, 2010). This could be due to environmental transmission from parent to child, or due to other, indirect, effects having to do with sharing a household. Several studies indicate which shared household factors correlate with reading ability. Aspects identified thus far include the number of books in a household, how much parents read, and socio-economic status (Evans, Kelley, Sikora, & Treiman, 2010; Leseman & de Jong, 1998; Manolitsis, Georgiou, & Parrila, 2011). However, correlates that are observed in the home environment do not necessarily represent an environmental cause, since such factors may be influenced by the genotype of the parents who provide the home environment (Kendler & Baker, 2007). As parents both transmit their genes and provide the child with the home environment, this may induce a gene-environment correlation, that is, the home environment that the child experiences is related to his or her genotype. If a parental characteristic (e.g., reading ability) still influences an offspring's characteristic after controlling for common genes that influence both generations, then this influence acts through the environment, referred to as cultural transmission.

Thus far, only a few studies explored the association between children's and parents' reading ability. A Dutch and a Finnish familial risk study showed a moderate correlation between parents' and children's reading fluency (Torppa, Eklund, van Bergen, & Lyytinen, 2011; van Bergen, de Jong, Plakas, Maassen, & van der Leij, 2012). A recent Dutch family study (based on an unselected sample) reported a parent-offspring correlation for reading fluency of 0.35 (van Bergen, Bishop, van Zuijen, & de Jong, 2015). Two English studies tried to disentangle genetic and environmental influences within the family. A study that included parent and (adoptive) child data (Kirkpatrick, Legrand, Iacono, & Mcgue, 2011) provides a genetically sensitive design. The study employed a broad construct of literacy (Wide Range Achievement Test), but did not explicitly test the nature of familial transmission. However, the pattern of correlations did not point to cultural transmission. Another adoption study (Wadsworth, Corley, Hewitt, Plomin, & Defries, 2002) showed that reading accuracy of parents and their biological offspring correlated around 0.2, whereas the association among parents and adopted children was absent. As adoptive children can only resemble parents because of cultural transmission, this study suggests that cultural transmission of reading ability is lacking.

We aim to further investigate this possibility in an extended twin design, that combines the strength of the classical twin study with the option to study cultural transmission, when twins and their parents have been phenotyped on the same measures. In our study, we used a fluency task in a different orthography, thereby extending empirical research on genetic and cultural transmission of reading in a different culture.

Returning to van Bergen et al., (2015) and Wadsworth et al., (2002), they report spouse correlations of 0.16 and 0.26 respectively, indicating non-random, or assortative, mating. We are unaware of other studies reporting assortative mating for reading ability, but its presence is important for several reasons: it may bias heritability estimates downwards if not taken into account in a classical twin design (i.e., data from mono- and dizygotic twins), while simultaneously suggesting a larger influence of shared environment (Cavalli-Sforza & Bodmer, 1971; Eaves, Fulker, & Heath, 1989). Assortative mating may also signify that offspring of dyslexic parents are particularly vulnerable, as they may inherit genetic and environmental risk factors from both parents.

Here, we aimed to explore the association between parents' and offspring's reading skills further: in a sample of Dutch twins, their siblings and their parents, we estimated resemblance of family members on a commonly used word-reading task. We test if offspring resemble their parents, if there is assortative mating between parents, if there is resemblance among offspring and if this resemblance is larger for monozygotic twin pairs than for dizygotic pairs and non-twin siblings. This is the first general-population study that explores the family resemblance of reading ability in a genetically-sensitive design.

Methods

Participants

Participants were recruited from the Netherlands Twin Register (NTR, Boomsma et al., 2006; van Beijsterveldt et al., 2013; Willemsen et al., 2013). Reading scores were collected in two samples. The first sample, which we will refer to as the twin-sibling sample (n = 310 NTR participants), consists of twin pairs with their older sibling from a longitudinal study on the development of brain and cognition (BrainScale, van Soelen et al., 2012a). Measurements took place around the twins' 9th, 12th and 17th birthday. If available, reading data of the first measurement were used (n = 294), otherwise from the third measurement (n = 16). This sample consisted of 47 monozygotic (22 male, 25 female) and 70 dizygotic twin pairs (21 male, 21 female, 18 opposite sex). Data for 41 brothers and 53 sisters aged between 9 and 21 years (M = 12.62, SD = 2.61) were simultaneously collected.

The second sample is a parent-offspring sample, consisting of 894 NTR participants from a population-based study on cognition and psychophysiology (Swagerman et al., 2015a). For this study, we included 436 twins (34 male and 72 female MZ twin pairs, 19 male and 40 female DZ twin pairs, and 50 opposite sex pairs), 33 brothers (mean age 35.91, *SD* = 16.10), 38 sisters (mean age 35.67, *SD* = 18.79), 125 fathers (mean age 63.95, *SD* = 10.20), and 158 mothers (mean age 61.34, *SD* = 10.83).

In total, data were available for 1100 participants from 431 families, of which 386 had at least two family members. On average, the mean age of this sample was 34.40 (*SD* = 22.92). These participants are representative of the general population: on average, adults had engaged in 14 years of education (range 6-20 years).

Materials

Reading test

Participants were given a list of Dutch words and were asked to correctly read out loud as many words as possible within one minute. Each participant was assessed on one of two highly similar tests, which we will refer to as one-minute-test 1 (OMT1) and one-minute-test 2 (OMT2).

OMT1. The OMT1 consists of 120 multisyllabic words, increasing in difficulty from two to four syllables (list 3C, Verhoeven, 1995). The manual reports a reliability of 0.86-0.92 in 9- to 12-year-olds (Moelands, Kamphuis, & Verhoeven, 2008).

OMT2. The OMT2 consists of 116 words of increasing difficulty (list B of Brus & Voeten, 1999). The first 10 words are monosyllabic. Thereafter they increase from two to four syllables. The reliability is 0.76-0.96 in 9- to 13-year-olds (van den Bos, Lutje Spelberg, Scheepstra, & de Vries, 1994).

The OMT1 was used on the first measurement of the twin-sibling sample (209 twins and 85 siblings), and the OMT2 was used on the third measurement of the twin-sibling sample and in the parent-offspring sample (443 twins, 80 siblings, all fathers and mothers). Both OMT1 and OMT2 were assessed in an independent sample of 122 9-year olds (end Grade 3; unpublished data of Peter F. de Jong). In this sample the tasks correlated 0.90. This correlation falls in the range of test reliabilities and corroborates that the OMT1 and OMT2 measure the same construct.

Procedure

Participants were first approached by mail, followed by a telephone call asking about willingness to cooperate and possible exclusion criteria.

Data collection in the twin-sibling sample took place in the University Medical Center Utrecht and Vrije Universiteit (VU) Amsterdam, and participants in the parent-offspring sample could choose if they preferred a home assessment or a lab assessment (at the VU or the Amsterdam Medical Center). Prior to starting the test protocol, procedures were explained to the participants, who signed informed consents. For children up to 16 years parents had to give consent as well. Procedures for both studies were approved by the Medical Ethics Review Committee of the Vrije Universiteit Medical Center Amsterdam and the Central Committee on Research Involving Human Subjects.

Statistical analyses

As an indicator of sample representativeness, we compared raw reading scores of 9-year olds (n = 210) to available national norms. For adults no norm scores are available. Raw test scores showed a good approximation to a normal distribution with a tail towards the lower scores, with 50% of the participants scoring between 90 and 120. Approximately 10% of all participants was able to read all the words of the list within a minute, but since they often made 1 or 2 mistakes, 5% of the participants had a final score of 116 words within a minute. First, using SPSS (IBM Corp., 2011), test scores were corrected for linear and quadratic effects of age in participants under age 18 (Figure 1), because reading ability increases throughout the first years of education. The standardized residuals (z-scores) were saved for further analyses. Test scores of participants over age 18 were standardized to z-scores without correcting for age. Figure 1 illustrates how the age correction and standardization of scores results in a mean score around 0 across samples, tests, and age.

Figure 1. Scatterplot of raw (panel A) and standardized (panel B) reading scores against age. Panel A shows the linear relationship in the OMT1 where reading score increases with age (including 95% confidence interval).

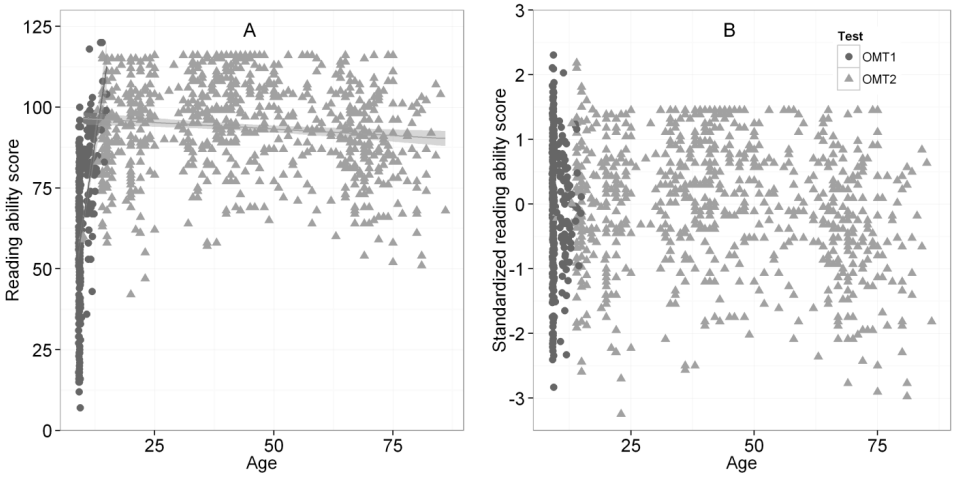
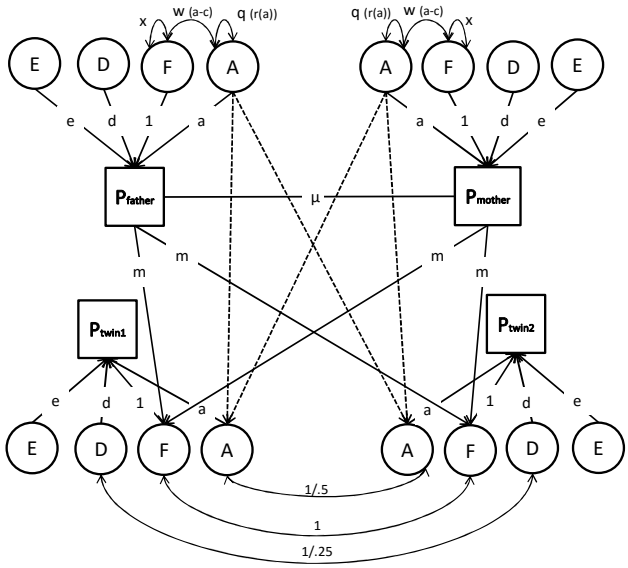


Figure 2. Path diagram of the extended twin family design.



P = phenotypic variance of father, mother or twin (siblings are not represented in this figure). A = additive genetic effects, E = unique environment including measurement error, D = non-additive genetic effects, F = family environment, μ = assortative mating co-path, m = cultural transmission, q = additive genetic covariance, x = family environment covariance, w = covariance between A and F (variances of other latent factors are 1). The dotted line is fixed to 0.5.

All standardized scores were analyzed using structural equation modeling in OpenMx (Boker et al., 2011). Figure 2 represents a path diagram of the extended twin family design (ETFD, Heath, Kendler, Eaves, & Markell, 1985; Keller et al., 2009). In the ETFD, the covariance structure between family members can be used to estimate the relative contribution of additive genetic (A), non-additive genetic (or dominant, D), family environment (F, passed on via cultural transmission from parents to offspring), sibling environment (S, shared by twins and siblings only) and unique environmental factors including measurement error (E).

Including data of parents reduces parameter bias if there is assortative mating present because assortative mating induces correlated genetic effects between parents and offspring which inflate correlations amongst offspring (Eaves et al., 1989; Fulker, 1982; Heath et al., 1985). Furthermore, data of siblings has been shown to increase statistical power to detect non-additive genetic influence (Posthuma & Boomsma, 2000). Parents transmit half of their segregating genes to offspring (dotted line in Figure 2 between A parent and A twin).

Monozygotic twins (MZ) share (nearly) 100% of their genetic material, whereas dizygotic twins (DZ) and non-twin siblings share on average 50% of the additive and 25% of the dominant genetic variance. This is represented as correlations of 1.0 between MZ twins and 0.5 (A) and 0.25 (D) between DZ twins and siblings. The family environment (F) correlates 1.0 between twins / siblings. However, a model estimating A, E, D, F and S simultaneously is not identified: therefore either D, F or S should be fixed to 0. The ETFD provides the possibility to model assortative mating between spouses (μ in Figure 2), which, if present, would result in an increased resemblance between all twins and siblings. Lastly, cultural transmission is indicated in Figure 2 by path m.

We fitted a model with A, E, F and D parameters, including cultural transmission to be able to test our primary hypothesis regarding cultural versus genetic transmission. Given that the DZ twin correlations are less than half the MZ correlation and the correlation between parents and offspring is low, dominance genetic effects (D) were modeled instead of shared sibling environment (S = 0). Since there may be sex differences in reading ability (Rutter et al., 2004), sex was included as a covariate. Parameter estimation was by raw-data maximum likelihood as implemented in OpenMx. The fit of nested submodels was compared by likelihood-ratio tests, based on the difference in minus twice the log likelihood (-2LL) between two models. The difference has a chi-square (χ2) distribution with the degrees of freedom (df) equaling the difference in df between the two models. If constraining parameters in a nested model did not result in a significantly worse fit (α= 0.05) this more parsimonious model was deemed the best fitting model. In submodels, the different means for family members, the significance of the covariate sex, and the significance of assortative mating and cultural transmission were tested by setting this parameter to 0.

Results

The average score of 9-year olds on the OMT1 was 56.95 (*SD* = 19.57), which falls within the 40-60th percentile of the national norm scores (Jongen & Krom, 2010). On the OMT2 the average score for participants of 18 years and older was 94.42 (*SD* = 14.67). For participants under 18, there was a significant effect of age (β = 3.1, $p < 0.001$) and age squared (β = -2.5, $p < 0.001$). There was no significant effect of sex on the mean (β = 0.12, $\Delta\chi^2(1) = 1.75$, $p = 0.19$).The MZ twin correlation was 0.62 (confidence interval (CI): 0.51-0.73) whereas the dizygotic correlation and twin-sibling correlation was 0.26 (CI: 0.17-0.32). The MZ correlation thus was larger than twice the DZ or sibling correlation. The parent-offspring correlation was 0.18 (CI: 0.03-0.31) and there was a spouse correlation of 0.38 (CI: 0.22-0.53).

Such a pattern of correlation among family members is consistent with a model that attributes resemblance to additive genetic factors, these are the factors that contribute to resemblance among all biological relatives, and to non-additive genetic factors. Non-additive genetic factors, or genetic dominance, contributes to resemblance among siblings, but not to the resemblance of parents and offspring (Falconer & Mackay, 1996). Maximum likelihood estimates for the additive genetic factors were 28% (CI: 0-43%) and for dominant genetic factors 36% (CI: 18-65%), resulting in a broad-sense heritability estimate of 64%. The remainder of the variance is attributed to unique environmental factors and measurement error (35%, CI: 29-44%). The path loading of the co-path between spouses (μ in Figure 2) was 0.38, indicating significant assortative mating: dropping this parameter from the model resulted in a significantly worse fit ($\Delta\chi^2(1) = 13.61$, $p < 0.001$). Dominance could not be omitted from the model without a significant change of fit ($\Delta\chi^2(1) = 19.4$, $p < 0.001$). There was no evidence for cultural transmission (m = 0.006, $\Delta\chi^2(1) = 2.9$, $p = 0.09$). Therefore the ADE model without cultural transmission was deemed the best fitting model.

Discussion

In this study we aimed to test if the family resemblance which has been reported for reading ability and disability is caused by genetic or cultural transmission. To our knowledge, we were the first to explore this using a sample including twins, their parents and siblings. Secondly, we aimed to test if assortative mating is present. We found that individual differences in reading ability were mainly caused by genetic factors, both additive and non-additive. Environmental factors that are shared between parents and children did not contribute to familial resemblance and no evidence was found for cultural transmission from parents to their offspring. In the remainder we will start with limitations, followed by discussion of modelling findings and their scientific and clinical implications.

This study has some limitations. First of all, although the sample size is considerable (> 1000 individuals), on the family level this study is smaller. Therefore we may be under powered to detect small effects. Secondly, the assumption of the ETFD is that etiological sources of variance are the same for parents and their offspring. That is, that the same genes play an equally large role for all family members, even if they belong to different generations. However, this is not necessarily the case: the influence of genetic factors may increase with age, as is shown for psychometric IQ (Haworth et al., 2010). In our sample, the twin group includes children, adolescents as well as adults and elderly (20% is over 40 years of age).

Therefore, should it be the case that heritability increases with age, our estimate would represent an average over the lifespan and will therefore be somewhat higher than if it were based on younger twins alone. In addition, we found a somewhat larger component of E on reading ability compared to other genetic studies which found estimates around 10% (Harlaar, Spinath, Dale, & Plomin, 2005; Samuelsson et al., 2007). This may reflect larger measurement error, or reflect genuine environmental influences that are not shared among family members. Adding a second measure of reading ability and working with a common-factor score may have reduced measurement error and allowed for the possibility to distinguish between these alternatives.

Notwithstanding these limitations, this is the first study to analyze data on reading ability with parents and their twin-offspring. This design is better suited to provide a comprehensive understanding of why family members resemble each other. From the model-fitting analysis it can be concluded that familial resemblance is caused by genetic factors: the broad sense heritability (variance due to additive + non-additive genetic factors) is 64%. We do not know of other studies that have found evidence for non-additive (or dominant) genetic influence on reading (dis)ability (e.g., Kirkpatrick et al., 2011).

Reading ability of spouses appeared to be correlated (assortative mating, 0.38), which is in line with findings from Wadsworth et al., (2002: 0.26) and other studies of traits that correlate with reading, like intelligence (Vinkhuyzen, van der Sluis, Maes, & Posthuma, 2012: 0.37), but lower than found by van Bergen et al., (2015: 0.16). As noted in the introduction, assortative mating may render children of a parent with dyslexia extra vulnerable, as their other parent may also exhibit below-average reading skills. Indeed, children of a dyslexic parent who go on to develop dyslexia themselves are more likely to have a second parent with reading difficulties (Gilger, Hanebuth, Smith, & Pennington, 1996; van Bergen, de Jong, Maassen, & van der Leij, 2014). Another implication of the finding of assortative mating is that future studies should take this into account, as it may bias the heritability estimates. Some twin studies report evidence for shared-environmental influences (reading disability e.g., Friend, Defries, & Olson, 2008; Harlaar et al., 2005; reading ability e.g., Petrill et al., 2007; Taylor & Schatschneider, 2010). However, these influences may have been overestimated in the presence of assortative mating. Regarding parent-offspring resemblance, the estimate of the parent-offspring correlation (0.18) is of similar magnitude to correlations with biological children reported by Wadsworth et al., (2002: 0.16-0.26) but lower than reported on another Dutch sample (van Bergen et al., 2015, 0.32-0.38).One consequence of genetic non-additivity (genetic dominance) is that parent-offspring resemblance is lower than sib-sib

resemblance. Whereas siblings share part (25%) of the variance due to genetic dominance, parents and offspring do not (Falconer & Mackay, 1996).

In conclusion, after taking into account the genetic liability that is passed on from parent to child and assortative mating, there is no additional effect of parental reading ability to offspring reading ability. This absence of cultural transmission is in line with the findings from Wadsworth et al., (2002), despite differences in reading measure, language, culture, and study design. For psychopathology in contrast, cultural transmission has sometimes been demonstrated (Maes, Silberg, Neale, & Eaves, 2007; McAdams et al., 2014). Therefore, for psychopathology, intervention aimed at the parents would also benefit mental health of their children. In the case of children with reading disability, we would advise that interventions should focus on the child's, and not the parents' reading skills. However, this does not mean that parental characteristics other than reading ability are not passed on through cultural transmission. An example of this might be the school parents choose for their child: school choice may not be related to parents' reading ability (but e.g., based on religious affiliation), but school choice may impact on children's reading ability (Taylor et al., 2010). School choice would then be an environmental influence which is passed on from parent to child.

As mentioned in the introduction, familial risk studies seek neurobehavioral precursors of dyslexia. The current study speaks to whether familial risk is in fact genetic or environmental in nature. The types of analyses that were employed in this paper depend on population-based data and would not be possible in dyslexic families: there would be a restriction of range within parental reading scores (i.e., they all score in the lower tail of the distribution) and without substantial variance, computing covariance would be futile. Our results suggest that the precursors for reading disability observed in familial risk studies are caused by genetic, not environmental, liability from parents. That is, having family risk does not reflect experiencing a less favorable literacy environment, but receiving less favorable genetic variants.

Chapter 5

The Computerized Neurocognitive Battery: validation, aging effects, and heritability across cognitive domains

This chapter is published as: Swagerman, S. C., de Geus, E. J. C., Kan, K. J., van Bergen, E., Nieuwboer, H. A., Koenis, M. M. G., Hulshoff Pol, H. E., Gur, R. E., Gur, R. C., & Boomsma, D. I. (2016). The Computerized Neurocognitive Battery: validation, aging effects, and heritability across cognitive domains. *Neuropsychology, 30,* 53-64.

Introduction

Cognitive performance varies greatly among individuals. Possible sources of individual variation are gender, age, and genetic and environmental factors. Studies on cognitive functioning increasingly aim to find the biological basis of cognition in brain substrates or genetic variants. These neurobiological and genetic association studies on individual differences in cognition require reliable and well-defined phenotypes obtained in large numbers of participants. Such studies would benefit greatly from the availability of cognitive tests that are optimally suited to explore mechanistic neurobiological and neurodevelopmental models in large samples. Understanding how cognitive functions develop across the lifespan and how they are influenced by environmental and genetic factors is critical for elucidating healthy and pathological brain function.

As cognitive functions may be differentially sensitive to sources of variation, both basic functions, such as processing speed or attention, and more complex functions, like reasoning or emotion processing, require consideration. Notably, neurocognitive tests based on functional neuroimaging are designed to activate specific brain systems, while traditional neuropsychological and intelligence tests may activate multiple brain systems simultaneously, making the latter less suitable in neurobiological studies (Gur, Erwin, & Gur, 1992).

In order to address the need for an efficient and comprehensive neurocognitive battery, the Brain Behavior Laboratory of the University of Pennsylvania has developed the web-based Computerized Neurocognitive Battery (CNB, Gur et al., 2001b; Gur et al., 2010; Gur et al., 2012). This battery is the result of an iterative validation process during which tests and test items were selected. Tests aim to target specific brain regions, which was validated in functional brain imaging studies (Roalf et al., 2014). Since its introduction, the CNB has undergone minor revisions including shortening of tests and adding new ones. The current version of the CNB (Gur et al., 2012) yields quantitative performance (accuracy and speed) measures in five neurobehavioral functions: executive-control, memory, complex cognition, social cognition, and sensorimotor speed. More specifically, within these five neurobehavioral functions the battery assesses performance across 14 cognitive domains, which are described in Table 1 and described in Supplementary material S1.

The need for an efficient and reliable neurocognitive battery extends beyond the English speaking countries for large-scale genetic, developmental and aging studies. For this reason we translated test instructions and test items from English into Dutch. International collaborative studies would benefit from the assurance that cognitive batteries can be deployed universally: cognitive

performance and effects such as sex and age should be comparable across countries.

The objectives of this paper are first to estimate validity and reliability of the battery's Dutch translation, secondly to investigate effects of age across cognitive domains, and third to estimate how these cognitive abilities are influenced by environmental and genetic factors. With regard to the validation part of our study, we aim to confirm reliability, validity and feasibility in home and laboratory settings of the CNB in a large population-based sample of 1140 participants (10 – 86 years). Here we present indices of reliability based on internal consistency (Cronbach's alpha) and on inter-correlations among the test scores. To confirm validity, we compare mean scores and effects of sex and age in the Dutch to the U.S. population. In addition, we correlate CNB scores to measures of a person's own and parental level of education. We also consider whether the CNB can provide scores comparable to intelligence scores as derived from traditional intelligence tests. If so, this would provide further convergent validity, because, although individual CNB test scores will be difficult to compare to traditional IQ scores, across batteries the sources of between test covariance can be expected to be the same (Johnson, te Nijenhuis, & Bouchard, 2008), genetic sources in particular (Plomin & Kovas, 2005).

Once we have established that the CNB provides reliable and valid measures of cognition, we can explore the etiology of individual differences in these cognitive phenotypes. These extend beyond sex- and linear age effects: therefore our second aim is to estimate nonlinear effects of age across the lifespan. Many cognitive functions improve as children mature, but with different trajectories for different functions: a well-known example is the late development of executive functions compared to memory (Gur et al., 2012). However, later in life cognitive abilities start to decrease again, especially in the domains of processing speed, memory and executive functioning, although there is currently little agreement on the time of onset of this decline (Deary et al., 2009; Salthouse, 2009; Schaie, 2005). Cognitive aging is most often studied in a small age range (i.e. only elderly), usually including only one or a few cognitive functions. Here we will explore the patterns of development across cognitive domains and covering the lifespan.

Our third and final aim regards environmental and genetic effects on the cognitive tests. Initial studies on a subset of the tests show heritability estimates between 10 and 70% (Calkins et al., 2010; Greenwood et al., 2007; Gur et al., 2007) in the U.S. population. These estimates are based on selected samples of schizophrenia patients.

We will extend these findings by estimating heritability for all accuracy and speed scores in an unselected sample, which facilitates generalization to the general population. We will also estimate heritability of the common variance among the CNB test scores. Since indicators of common variance among psychometric IQ tests. i.e., general factors of intelligence, are the most heritable among the indicators of intelligence, with an estimated heritability coefficient of 50 to 80% (Jensen, 1998; Plomin, 2012), we expect a high heritability. If so in our analyses, this would further confirm validity.

Heritability was estimated using two approaches, both based on the resemblance in cognitive performance among family members as a function of their genetic relatedness. Half of our sample consisted of twins; the other half of parents, siblings, and children of twins and siblings. The first approach is based on information from the mono- and dizygotic twin pairs, who are of the same age by definition, and estimate the extent to which their resemblance is due to shared genes, or common environmental influences shared by offspring growing up in the same family. In the second approach we extend the analyses to data from the entire pedigree, i.e. all family members, where cross-generation resemblance is analyzed simultaneously with the resemblance in twin pairs. These pedigree-based analyses provide information on genetic stability across generations.

Method

Participants

Participants were mainly recruited by the Netherlands Twin Register, which is a population-based register that recruits twins and other multiples, their parents, siblings, spouses, and offspring (NTR, Boomsma et al., 2006; van Beijsterveldt et al., 2013; Willemsen et al., 2013). In total there were 1140 participants, mainly (n = 1110) from 431 families who were recruited from all regions in the Netherlands. The other 30 subjects were university students. Most participants (621) were part of a twin pair or triplet. Twin pairs were monozygotic (54 male, 100 female pairs) or dizygotic (42 male, 60 female, 71 opposite-sex pairs). The rest of the sample consisted of siblings (150), parents of twins (279), partners of twins and siblings (51), and offspring of twins and siblings (9). The age range was from 10 to 86 (M = 37.73, SD = 20.86). The figure in Supplementary material S2 depicts the age distribution of these 472 males (41.4%) and 668 females. On average, participants had 12.92 years of education (SD = 3.29). The average number of years of education in their parents was 12.34 (similar to the average in the Dutch population, UNESCO Institute for Statistics, 2013).

Procedure

Studies and procedures were approved by the Medical Ethics Review Committee of the Vrije Universiteit Medical Center Amsterdam and the Central Committee on Research Involving Human Subjects. Participants were approached by mail. When they (and possibly other family members) were willing to participate, a structured telephone call followed. This phone call had the purpose of informing participants and of asking about exclusion criteria. Exclusion criteria were epilepsy or paralysis, and physical problems that would influence test performance (like a broken arm or severe vision problems).

Testing took place at the Vrije Universiteit laboratory (n = 358), at the participants' home (n = 536), or in the laboratory of the University Medical Center Utrecht (n = 246). In all settings, test conditions were controlled to prevent disturbance or distractions. Prior to the start of the testing, the administrator fully explained the procedure, after which written informed consent was obtained. Participants of 12 years of age and older signed themselves. For children up to 16 years parents needed to sign as well. Following the CNB protocol from the Brain Behavior Laboratory, participants were asked to complete a reading test (Swagerman et al., 2015b). For none of the participants did the reading test indicate that they were unable to complete the CNB. Participants received a gift voucher and compensation for their traveling costs. All participants received feedback on their performance, in the form of a graph in which their score was ranked with participants of the same age.

Standardized procedures were followed for both the home and laboratory test location. The participant sat at a desk, with the test administrator behind him or her. Macbooks were used for administration with identical mouse and screen settings. All participants were instructed to use only the mouse and spacebar for responses (laptop mousepad was disabled).

Prior to the start of each test, the administrator read the test instructions out loud to the participant, after which the participant was provided with practice trials (except for the memory tests and the Conditional Exclusion Test). The practice trials had to be completed successfully in order to start the test. During the cognitive assessment, the experimenters kept track of whether test scores were valid, based on the participant's apparent motivation or interruption of the test session. Automated test score validation occurred upon upload to the Pennsylvania web servers that host the CNB (Gur et al., 2012). Completion of the battery lasted on average 1.5 hours (ranging between approximately 50 minutes and three hours), including optional breaks at three designated points.

A subsample of adolescent participants (n = 246, 14 – 22 years old), took part in the BrainScale study on development of brain and cognition (van Soelen et al., 2012a). These participants completed a shortened version of the Wechsler Intelligence Scale for Adults (WAIS, Wechsler, 1997) on the same day as they were assessed on the CNB.

Measures

In addition to the CNB, participants were asked about, or filled out a questionnaire on lifestyle (drinking, smoking, exercise behavior) and medication use.

Cognitive battery

The Dutch translation of the current CNB includes a total of 17 tests, yielding measures of performance (accuracy and speed) in 14 cognitive domains (Table 1; Supplementary Material S1). All test instructions and test items were translated from English into Dutch, and back translated by a professional translator. In addition, the frequency of the words in the A and B versions of the Word Memory Test and Verbal Reasoning Test were compared, to ensure that both versions were of equal difficulty.

Accuracy was defined as the percentage or number of correct responses on a test. Measures of *speed* were derived from the median response time in milliseconds of correct responses, and were multiplied by -1. Hence for both accuracy and speed, higher scores denote better performance. The Finger Tapping Test (TAP) did not provide accuracy scores: the score reflected the number of taps one can produce within 10 seconds over 6 attempts. TAP score thus constitutes a speed score, where a relatively high score denotes relatively fast motor speed.

Psychometric IQ

The shortened WAIS included two verbal and two performance tests, which were, in order of assessment, Vocabulary (verbal), Block Design (nonverbal), Similarities (verbal), and Matrix Reasoning (nonverbal). Using normative tables per age group, raw test scores were transformed into standardized scale scores (Wechsler, 1997). Then a correction for the number of excluded subtests was applied (2 out of 6 verbal and 2 out of 5 nonverbal tests) in order to obtain total (TIQ), verbal (VIQ), and performance IQ (PIQ).

Years of education

Participants were asked how many years of education they and their parents had completed. Repeating a school year did not count as an extra year. In case the same type of education was repeated at a higher level (e.g., economics degree

at college level followed by university level), only the number of years at the highest level was counted. Parental education was defined as the mean number of years of paternal and maternal education, or of one of them if the other was unknown.

Statistical Analyses

Validity and reliability

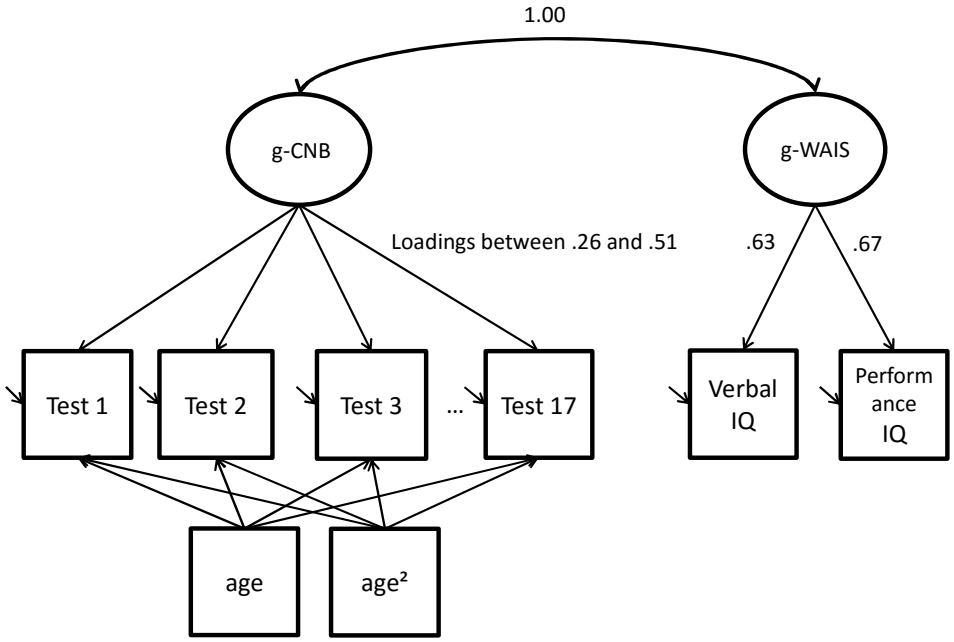
Excluding test scores of children under 13 (n = 4) and scores that were judged invalid (0.8%), we calculated in SPSS 21.0 (IBM Corp., 2011) for each test the average accuracy score, average speed score, average duration, and the Cronbachs's alpha coefficient of internal consistency (not possible for the Conditional Exclusion Test). Further, correlations among accuracy scores, correlations among speed scores, and per test the correlation between accuracy score and speed score were calculated (all while correcting for effects of sex and age). Accuracy and speed scores were skewed. In addition, the data had to be considered as clustered since the study involved family members. Statistical analyses (other than the genetic analyses) thus required correction of the standard errors of the parameters. This was accomplished by analyzing the data in the statistical program R (version 3.1.1, R Core Team, 2014) using packages lavaan (Rosseel, 2012) and lavaan.survey (Oberski, 2014), by including family number as cluster variable (each student received a unique family number), and by opting for robust sandwich estimation. This procedure allowed for the analysis of clustered, non-normally distributed, but continuous outcome variables.

Following Gur et al. (2010) we obtained sex differences on all cognitive measures, and correlations between performance scores and education as well as parental education. Because own educational level is meaningful only after the typical age that maximal academic training can be achieved, we restricted these analyses to a subsample over age 30 (n = 632, M = 14.21, SD = 3.38).

In the literature, the variance that is common to IQ subtest scores is usually described by the latent variable referred to as general intelligence or simply 'g' (Jensen, 1998; Spearman, 1904). A strong correlation between the common variance in CNB test performance and general intelligence (as derived from traditional batteries) would imply that once performance measures on the CNB are aggregated, a CNB sum score would be similar to a traditional Total IQ score. TIQ can be considered to constitute the most accurate proxy of general intelligence (after the g-factor score). The WAIS VIQ, PIQ and TIQ scores that were available in the subsample therefore provided the opportunity to test this using the following approach.

We selected all CNB accuracy scores, those on the Motor Praxis test excluded, since WAIS scores are based on accuracy scores rather than speed scores, and concern cognitive abilities and knowledge rather than motor skills. Next, we forced a confirmatory oblique 2-factor model on the (WAIS and CNB accuracy) data, in which the CNB scores loaded on one latent factor (labeled '*g*-CNB' in Figure 1) and the WAIS VIQ and PIQ scores on the other (labeled '*g*-WAIS'). As WAIS IQ scores are already age corrected, we added linear and (mean-centered) quadratic age terms as predictors of the CNB scores to make them comparable to the WAIS. The correlation between the two latent factors was considered to indicate the strength to which the common variance among the CNB accuracy test scores relates to general intelligence as assessed by the WAIS. A high correlation would indicate that the CNB can provide a valid and reliable estimation of general intelligence. To be able to confirm this, we obtained factor scores on the *g*-CNB and correlated these with WAIS TIQ scores. This correlation was interpreted as a measure of both reliability and cross-validity.

Figure 1. Oblique two-factor model of overlap in variance of CNB tests and WAIS Verbal and Performance IQ scales.



Circles represent the two latent variables that describe common variance among CNB tests (labeled g-CNB) and common variance among WAIS subtests (labeled g-WAIS). Squares represent the observed CNB test scores and WAIS Verbal and Performance IQ scores. Double-headed arrows between two variables represent correlations and single-headed arrows between two variables represent standardized regression effects (factor loadings included). Any other single-headed arrows represent residuals

Analyses of aging effects

Relations between test performance scores and age were analyzed according to a model in which the scores on a particular test were regressed on age (across the age range in the data: 13-86 years old) and on (mean centered) age squared.

Heritability analyses

To estimate heritability, data of monozygotic (MZ) twins who are (nearly) genetically identical and dizygotic (DZ) twin pairs who share on average half of their segregating genes were analyzed first (Boomsma, Busjahn, & Peltonen, 2002). Because MZ and DZ twins differ in their genetic similarity, genetic effects are suggested for a trait if the MZ correlation is higher than the DZ correlation. Effects of common environment shared by twins are suggested to also contribute to twin resemblance when the DZ correlation is larger than half the MZ correlation. Modeling of twin data was performed in OpenMx (Boker et al., 2011) by raw-data maximum likelihood. All speed scores were log-transformed prior to analysis to reduce skewness (to the right towards slow response times) and heteroscedasticity (more variance with older age). First, in a saturated model, means, variances and twin correlations were estimated for monozygotic (MZ) and dizygotic (DZ) twin pairs. Next, parameters representing the influence of additive genetic factors (A), common environment shared by twins (C) and unique environment (E, including measurement error) were estimated (Plomin et al., 2013). The model included sex, age and (mean centered) age^2 as moderators of the mean scores.

Secondly, heritability was estimated in Mendel (Lange et al., 2013; Lange, Westlake, & Spence, 1976), analyzing the entire pedigree structure including twins. The approach implemented in Mendel takes the entire pedigree information to estimate variance components and allows for the inclusion of all relatives. The effect of common environment (C) was estimated for twins and their non-twin siblings growing up in the same household up to age 22 (mean age when children move out of their parents' house, Statistics Netherlands, 2014). Heritability analyses were performed for the 15 accuracy and 17 speed outcomes. As 98% of all participants had perfect accuracy on the Motor Praxis Test, for the sensorimotor domain only speed was examined. In addition, heritability was estimated for both the factor score on the *g*-CNB and WAIS TIQ scores.

Results

Validity

Internal consistencies and intercorrelations

Table 1 includes general information about the cognitive tests and domains, mean duration, mean accuracy and speed score, and Cronbach's alpha coefficient. These coefficients of internal consistency were high for speed (median = 0.92) and moderate to high for accuracy (median = 0.62). Table 2 summarizes the intercorrelations among the performance scores. When intercorrelations were estimated without correcting for sex and age, results are similar but generally a little stronger. As expected, correlations among accuracy scores were all positive (although the magnitudes ranged considerably, mostly small to moderate). Intercorrelations among speed scores were for the majority positive with magnitudes ranging from small to large. Correlations between accuracy score and speed of each test varied considerably, ranging from negative and large (-0.73, nonverbal reasoning) to positive and moderate (0.26, verbal memory) with a median of 0.07. Some tests were thus characterized by a tendency of better accuracy being accompanied by faster response time, while others were characterized by a tradeoff, where better accuracy was accompanied by slower response time (the nonverbal reasoning test in particular).

Sex differences

Figure 2 depicts the mean sex differences on the performance measures. We found that females tended to score more accurate on all social cognition tests as well as the face and word memory tests (negative effects in Table 3) whereas males showed higher scores in the language reasoning, spatial ability and spatial memory (delayed) tests (positive effects in Table 3). Regarding speed, males were faster on the motor speed and spatial ability tests, and females on the verbal memory (delayed), emotion identification and age differentiation tests.

Education and parental education

Figure 3 provides the correlations between cognitive performance and education. The correlations between years of education and accuracy were all positive, ranging from small (0.16, age differentiation) to moderate (0.49, language reasoning). Those with speed ranged from moderately negative -0.17 (nonverbal reasoning) to moderately positive 0.39 (verbal memory). Their medians were moderate and positive (accuracy 0.29; speed 0.20).

Correlations between mean parental education and cognitive accuracy were also positive and also ranged from small (0.05, sensorimotor speed) to moderate (0.28 nonverbal reasoning). Correlations with parental education and speed

ranged from negative and small (-0.02, nonverbal reasoning) to positive and moderate 0.31 (sensorimotor speed). Both medians were positive but small (accuracy 0.14; speed 0.04).

Relation to psychometric intelligence

The mean IQ scores in the subsample of 246 participants who completed the shortened WAIS were comparable to the population average of 100 (*SD* = 15): VIQ 102.44 (*SD* = 13.76), PIQ 106.15 (*SD* = 14.25), TIQ 103.80 (*SD* = 12.74). The tests that correlated highest with IQ were Word memory and Verbal- and Matrix Reasoning (see Supplementary material S3).Fitting the oblique two-factor model (Figure 1) showed that the latent g-CNB factor and the common g-WAIS factor had to be considered to represent the same construct, because the estimated correlation between the two factors equaled 1.0, denoting a perfect relation. That overall performance on the CNB compares well to cognitive performance as assessed by the traditional WAIS IQ test battery was confirmed by the high correlation between the factor scores on the g-CNB and WAIS Total IQ, which was 0.82. In conclusion, the results imply that, corrected for age effects, overall performance on the CNB compares well to general intelligence as assessed by a psychometric intelligence test battery. This would suggest that one does not need an intelligence battery in addition to the CNB in order to obtain estimates of general intelligence (next to performance measures of specific neurocognitive functioning). In the interest of possible future assessment of intelligence via the CNB, Supplementary material S3 includes a description of how to calculate IQ scores based on CNB test scores.

Analyses of aging effects

The correlations between cognitive performance and age (see Supplementary Figure S4 for illustration) ranged in magnitudes from positively small (0.15, language reasoning) to negatively moderate (-0.35, emotion identification) for accuracy. Associations with speed were all negative, and ranged from small (-0.03, language reasoning) to moderate (-0.53, spatial memory delayed). The contributions of linear and quadratic age effects are detailed in Table 3. Examples of the curvilinear age dependencies are visualized in Figure 4 (see S5 for all CNB tests). In general, the results clearly indicate that test performance tends to decline as a nonlinear function of age, but also that the pattern of decline differs across the cognitive domains. Often, cognitive performance peaked during childhood or adolescence after which performance gradually declined with a steeper slope after this peak: this was seen for many of the speed measures, and accuracy on nonverbal reasoning, attention and most memory tests. However, for other domains, like language reasoning (accuracy), performance increased into middle adulthood and was followed by limited decline.

Table 1. Cognitive domains and test names, and mean administration duration (Time, in minutes), number of participants who completed the test (N), and the test's mean score (M, and SD), Cronbach's alpha coefficients (α) of accuracy score (percentage or number of correct responses) and speed (median response time, in ms).

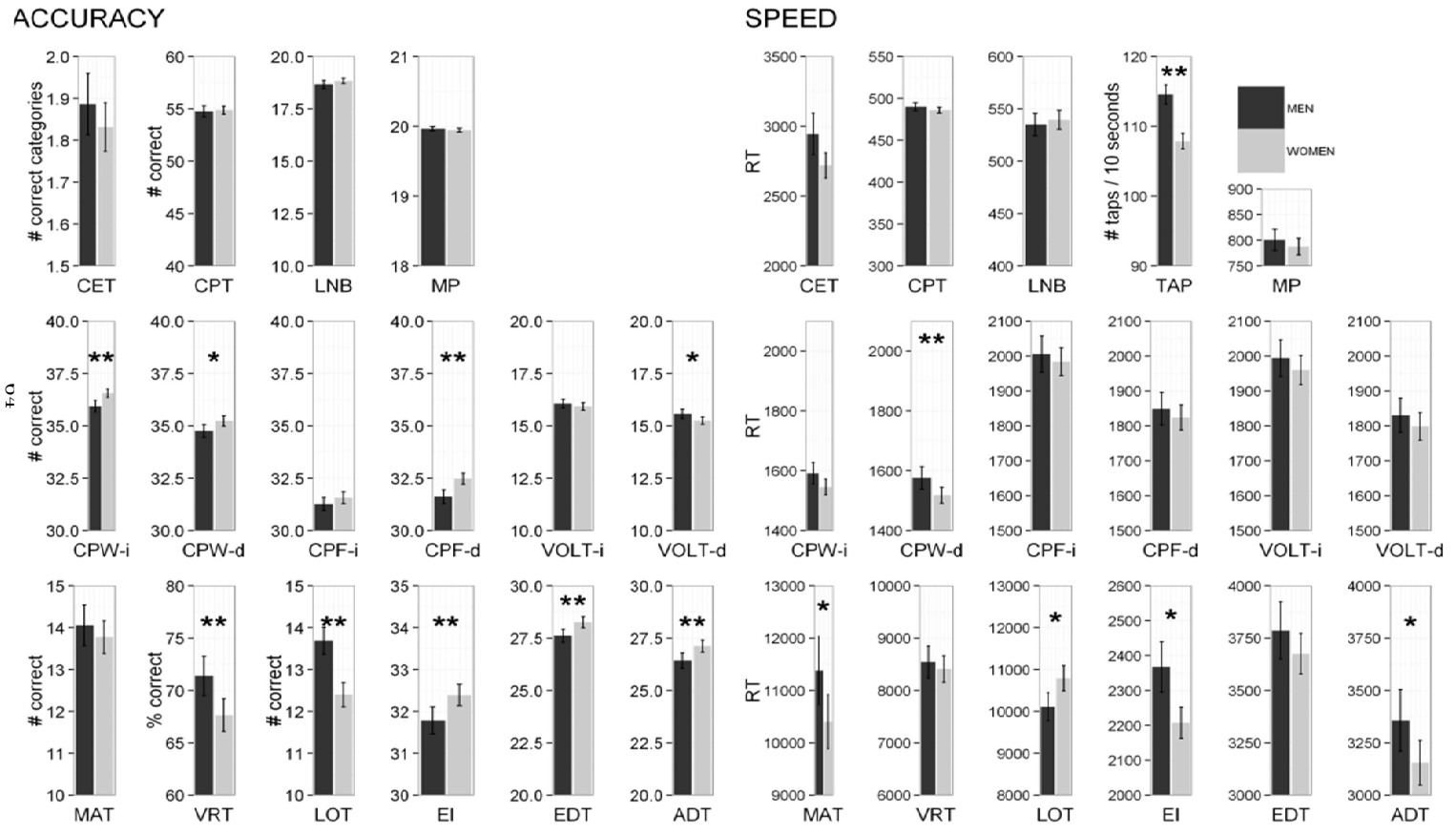
					Accuracy			Speed		
Executive Control	Test name	Label	Time	N	M	SD	α	M	SD	α
Abstraction / flexibility	Penn Conditional Exclusion Test [a]	CET	4.9	1125	1.9	0.8	[d]	2813.3	1392.6	[d]
Attention	Penn Continuous Performance Test [a]	CPT	5.3	1125	54.8	5.4	.86	487.7	49.1	.82
Working memory	Letter-N-Back Test [a]	LNB	9.2	1114	18.8	1.8	.77	537.7	118.0	.80
Memory										
Verbal Memory	Penn Word Memory Test [b]									
	- immediate	CPW-i	3.1	1125	36.3	2.8	.62	1564.5	368.2	.92
	- delayed	CPW-d	1.1	1124	35.0	3.3	.64	1541.7	376.6	.91
Face Memory	Penn Facial Memory Test									
	- immediate	CPF-i	3.9	1123	31.4	3.5	.56	1992.7	544.2	.92
	- delayed	CPF-d	1.5	1121	32.1	3.5	.57	1834.2	489.7	.89
Spatial Memory	Visual Object Learning Test [a]									
	- immediate	VOLT-i	2.7	1117	16.0	2.3	.48	1973.8	554.6	.87
	- delayed	VOLT-d	0.5	1115	15.4	2.4	.48	1811.5	519.7	.86
Complex Cognition										
Nonverbal reasoning	Penn Matrix Reasoning Test	MAT	7.8	1129	13.9	5.2	.90	10806.0	6959.8	.88
Language reasoning	Penn Verbal Reasoning Test [a,b]	VRT	1.8	1123	69.2	20.6	.53	8465.8	3332.5	.74
Spatial ability	Variable Penn Line Orientation Test [a]	LOT	5.5	1119	12.9	3.7	.79	10506.8	3861.8	.97
Social Cognition										
Emotion Identification	Penn Emotion Identification Test	EI	2.3	1132	32.1	3.5	.62	2273.4	685.7	.92
Emotion differentiation	Measured Emotion Differentiation	EDT	3.4	1131	28.0	3.5	.69	3721.0	1369.1	.94
Age Differentiation	Age Differentiation Test	ADT	3.0	1122	26.8	3.9	.74	3238.4	1493.5	.94
Sensorimotor										
Sensorimotor speed	Motor Praxis Test	MP	1.8	1130	20.0	0.4	.93	793.2	221.3	.95
Motor speed	Computerized Finger-Tapping Test[a]	TAP	3.5	[c]	[c]	[c]	[c]	110.6	15.1	.96

[a] short test version. [b] different items for children. [c] no accuracy score available for TAP. [d] not amenable for calculating

Table 2. Intercorrelations between accuracy (upper triangle) and speed (lower triangle) and cross correlations between accuracy and speed (on diagonal, underscored). Correlations are corrected for age and sex.

	Executive Control			Memory						Complex Cognition			Social Cognition			Sensorimotor	
	CET	CPT	LNB	CPWi	CPWd	CPFi	CPFd	VOLTi	VOLTd	MAT	VRT	LOT	EI	EDT	ADT	MP	TAP
CET	.12	.18	.12	.16	.15	.10	.08	.12	.13	.23	.17	.18	.18	.12	.10	-	-
CPT	.17	.21	.20	.16	.22	.21	.16	.11	.14	.27	.24	.21	.16	.19	.16	-	-
LNB	.14	.50	.18	.16	.20	.15	.16	.10	.11	.25	.25	.12	.13	.16	.11	-	-
CPWi	.24	.35	.20	.26	.58	.23	.21	.22	.18	.25	.22	.14	.15	.20	.14	-	-
CPWd	.25	.34	.20	.78	.16	.28	.26	.17	.20	.27	.25	.20	.12	.24	.12	-	-
CPFi	.24	.25	.12	.52	.53	.14	.62	.21	.22	.22	.22	.18	.26	.20	.18	-	-
CPFd	.22	.27	.13	.52	.62	.74	.07	.20	.21	.23	.22	.19	.25	.21	.21	-	-
VOLTi	.28	.21	.06	.51	.57	.55	.55	-.03	.52	.26	.22	.16	.12	.15	.12	-	-
VOLTd	.21	.21	.08	.43	.53	.48	.54	.65	-.06	.26	.21	.21	.10	.19	.15	-	-
MAT	.25	.01	.06	-.01	.05	.14	.16	.15	.19	-.73	.42	.33	.20	.33	.24	-	-
VRT	.24	.17	.15	.18	.21	.26	.22	.25	.27	.31	-.03	.24	.19	.27	.19	-	-
LOT	.27	.21	.11	.35	.38	.39	.41	.44	.45	.28	.33	-.15	.12	.31	.27	-	-
EI	.22	.28	.18	.49	.43	.48	.47	.36	.28	.10	.24	.33	.14	.24	.22	-	-
EDT	.35	.23	.16	.36	.38	.48	.47	.44	.34	.28	.35	.45	.47	-.07	.46	-	-
ADT	.27	.17	.11	.26	.32	.46	.48	.46	.38	.30	.29	.45	.36	.62	-.03	-	-
MP	.15	.27	.18	.50	.40	.25	.27	.24	.17	-.01	.06	.24	.43	.21	.13	-	-
TAP	.07	.24	.14	.32	.28	.17	.18	.13	.14	-.10	.01	.21	.18	.15	.06	.30	-

Figure 2. Mean scores (and their 95% confidence intervals) of cognitive test scores in men (darkgrey) and women (lightgrey).



See Table 1 for abbreviations of cognitive tests. No accuracy score available for TAP. * = $p < 0.05$, ** = $p < 0.01$.

Figure 3. Correlations (and their 95% confidence intervals) of the cognitive tests with participants' own level of education and their parents' level of education. Correlations with accuracy scores are given in black and with speed scores in grey.
See Table 1 for abbreviations of cognitive tests. Note: No accuracy score available for TAP.

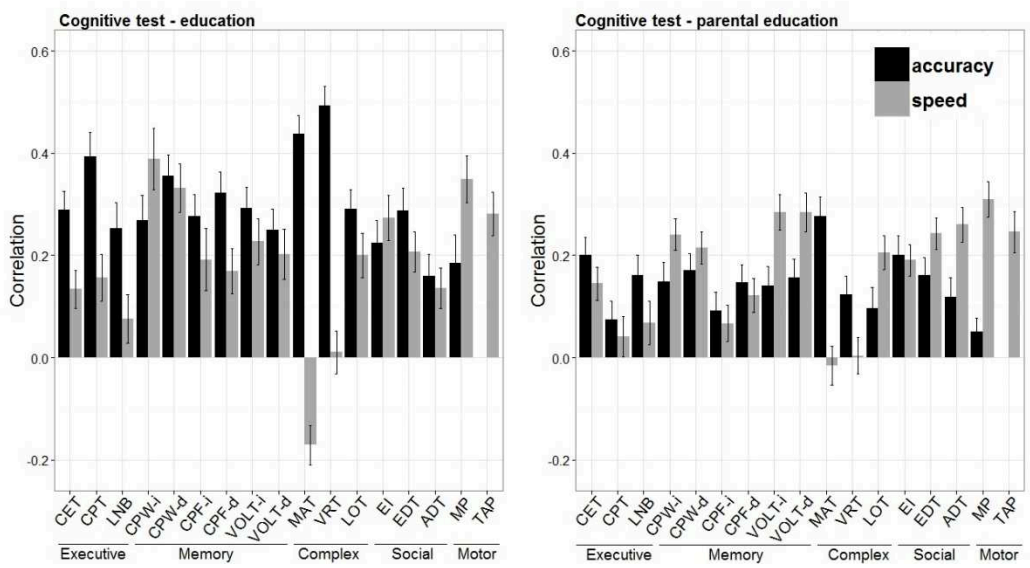


Figure 4. The curvilinear relation between cognitive test scores and age, including 95% confidence intervals. Females are given in black (●), males in grey (▲). Panel A: Language reasoning accuracy. Panel B: Nonverbal reasoning accuracy. Panel C: Sensorimotor speed. Note that cognitive decline is more pronounced in Panels B and C than A.

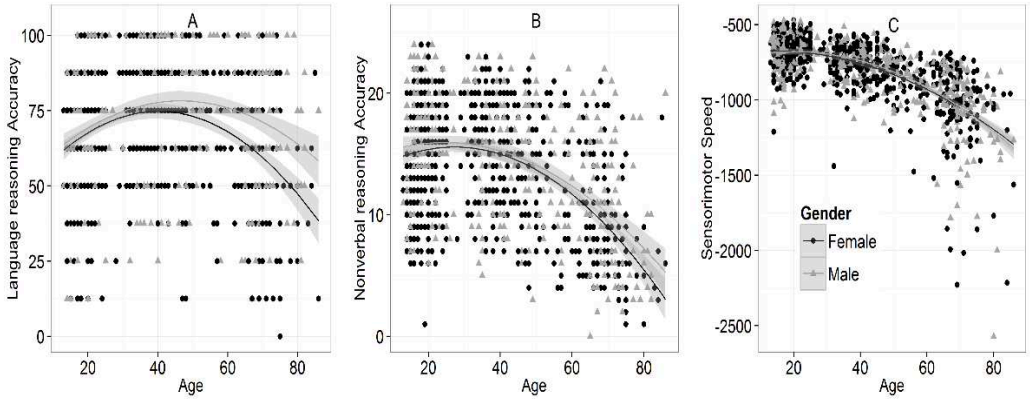


Table 3. Standardized effect size (or correlation) of univariate modeling of effects of age, sex, education (in years, in participants over 30 years of age) and mean parental education (in years) on accuracy and speed.

	Age (in years)		Age² (in years)		Sex (females 0, males 1)		Education (in years)		Parental education (in years)	
Cognitive domain	Accuracy	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy	Speed	Accuracy	Speed
Executive Control										
Abstraction flexibility	-.32**	-.40**	-.10**	-.15**	.04	-.08	.29**	.13**	.20**	.15**
Attention	.10*	-.10**	-.44**	-.32**	-.01	-.04	.39**	.16**	.07*	.04
Working memory	-.17**	-.26**	-.24**	-.02	-.05	.02	.25**	.08	.16**	.07
Memory										
Verbal Memory	-.12**	-.38**	-.10*	-.35**	-.11**	-.06	.27**	.39**	.15**	.24**
delayed	-.21**	-.42**	-.21**	-.28**	-.07*	-.08**	.36**	.33**	.17**	.22**
Face Memory	.01	-.12**	-.32**	-.26**	-.04	-.02	.28**	.19**	.09*	.07
delayed	-.13**	-.26**	-.27**	-.21**	-.12**	-.03	.32**	.17**	.15**	.12**
Spatial Memory	-.24**	-.48**	-.08*	-.15**	.03	-.03	.29**	.23**	.14**	.28**
delayed	-.21**	-.53**	-.06	-.13**	.07*	-.03	.25**	.20**	.16**	.28**
Complex Cognition										
Nonverbal reasoning	-.30**	-.21**	-.26**	.11	.03	-.07*	.44**	-.17**	.28**	-.02
Language reasoning	.15**	-.03	-.29**	-.04	.09**	-.02	.49**	.01	.12**	-.01
Spatial ability	-.15**	-.36**	-.15**	-.30**	.17**	.09**	.29**	.20**	.10**	.21**
Social Cognition										
Emotion Identification	-.35**	-.44**	-.16**	-.28**	-.09**	-.12**	.22**	.27**	.20**	.19**
Emotion Differentiation	-.15**	-.47**	-.21**	-.19**	-.09**	-.04	.29**	.21**	.16**	.24**
Age Differentiation	-.11**	-.51**	-.21**	-.13**	-.09**	-.07*	.16**	.14**	.12**	.26**
Sensorimotor										
Sensorimotor speed	-.10**	-.52**	-.23**	-.24**	.03	-.03	.19**	.35**	.05	.31**
Motor speed		-.21**		-.31**		.22**		.28**		.25**

* = significant at α = .05, ** = significant at α = .01.

Table 4. Variance components explained by additive genetic effects (heritability) based on twins, and based on all family members, including 95% confidence intervals (CI).

		Twins		All family members	
Cognitive domain		Heritability	CI	Heritability	CI
Executive Control					
Abstraction & flexibility	Accuracy	12	0-26	13	03-23
	Speed	41	0-53	38	28-48
Attention	Accuracy	42	19-56	38	26-49
	Speed	30	0-51	40	32-48
Working memory	Accuracy	23	0-41	22	07-37
	Speed	15	0-47	31	20-41
Memory					
Face Memory	Accuracy	31	0-49	34	22-46
	Speed	43	21-56	36	25-47
delayed	Accuracy	35	0-48	31	22-41
	Speed	49	17-60	43	32-54
Verbal Memory	Accuracy	27	1-40	26	13-39
	Speed	41	12-53	44	34-53
delayed	Accuracy	16	0-32	18	03-33
	Speed	36	0-49	36	26-45
Spatial Memory	Accuracy	30	0-44	31	21-40
	Speed	33	0-46	33	24-42
delayed	Accuracy	31	1-44	30	20-39
	Speed	35	1-48	33	23-43
Complex Cognition					
Nonverbal reasoning	Accuracy	52	15-63	40	30-51
	Speed	46	1-57	34	23-45
Language reasoning	Accuracy	29	0-42	37	26-48
	Speed	31	0-49	31	22-39
Spatial ability	Accuracy	46	25-57	49	42-56
	Speed	34	0-51	30	20-41
Social Cognition					
Emotion Identification	Accuracy	27	0-48	14	00-29
	Speed	30	0-50	37	27-46
Emotion Differentiation	Accuracy	4	0-34	17	05-30
	Speed	23	0-50	35	25-45
Age Differentiation	Accuracy	0	0-37	22	12-33
	Speed	40	11-52	32	21-42
Sensorimotor					
Sensorimotor Speed	Speed	19	0-53	45	35-55
Motor speed	Speed	38	14-51	31	19-43
General intelligence					
'g-CNB' (latent factor of CNB accuracy)		70	52-77	68	61-75
Total IQ (WAIS)		75	61-84	-	-

Heritability analyses

Overall, twin correlations (Supplementary Table S6) of monozygotic twin pairs were larger than of dizygotic twin pairs, suggesting effects of genetic influences on individual differences in test performance. Genetic modelling of twin data (Table 4) showed moderate heritability for the majority of the tests. For accuracy, heritability ranged from 0 (ADT) to 52% (nonverbal reasoning, median of 31%). For speed measures, heritability ranged from 15 (working memory) to 49% (face memory delayed, median of 33%). For nearly all cognitive domains, influences of the common environment (C) were absent or small (between 0 and 24%), mostly seen in the social cognition domain.

Heritability estimates based on all available pedigree information were highly similar: between 13 and 49% of the total variance in speed and accuracy could be attributed to genetic factors. These results imply that expression of genes that influence cognitive performance are stable over generations. Individual differences in the factor scores on the latent variable g-CNB were 70% heritable, without any evidence for C, whether based on twin data or on all available family data. This was close to the heritability of Total IQ on the WAIS: 75%.

Discussion

The aim of this paper was threefold: the first was to establish reliability and validity of the Dutch translation of the Computerized Neurocognitive Battery (CNB). The second was to explore how cognitive domains, as measured by the CNB, develop across the lifespan. The third was to estimate how these cognitive abilities are influenced by environmental and genetic factors. We conclude, based on a non-selected sample consisting of family members, that the CNB is a reliable and valid instrument in the Dutch population, with comparable scores to the U.S. studies. As part of the validation objective in our analyses, we report high Cronbach's alpha's across all tests. These indices of internal consistency are slightly lower than those reported by Gur et al., (2010), but this is likely due to the use of shortened tests. Intercorrelations among cognitive tests were of small to moderate magnitude, but of similar magnitude in the Netherlands and the U.S. without correcting for effects of age and sex. The Dutch and U.S. samples further show similar mean accuracy scores. The Dutch sample demonstrated somewhat longer response times than the U.S. sample, which probably reflects the fact that the age range of the Dutch sample was broader and included more elderly (see also below).

Another part of the validation of the CNB concerned exploration of the role of two well-known covariates of cognitive performance: sex and age. Compared to the results from the U.S. sample, we found effects that were overall similar, although small differences can be noticed. For example, in the Dutch study males and females performed about equally well on tests measuring attention and working memory, whereas Gur et al., (2010, Figure 3) report lower attention scores for males and higher working memory for females. However, generalizing across all CNB tests, standardized effect sizes were distributed around zero, which suggests the absence of an overall sex effect. This fits with findings from the literature on intelligence: whenever sex differences are found (also in the Dutch population, e.g., van der Sluis et al., 2008), they are usually test specific and small, and the consensus is that there is no evidence for any sex difference in overall cognitive performance (Hyde, 2014).

Regarding age effects, the broader age range of the Dutch sample is a likely explanation of the finding that correlations with age tended to be stronger in this sample compared to the U.S. sample (Gur et al., 2010). Yet, the overall picture was the same: older age is associated with slower as well as less accurate performance, although across cognitive domains the associations with age vary considerably in strength. CNB results are well in line with previous findings from research into cognitive aging (Salthouse, 2009). These findings have shown that the relation between age and cognitive performance is quadratic: (young) adults often outperform children and adolescents as well as older adults and elderly. Further, they indicate that the shape and rate of cognitive decline tend to differ across domains, and cognitive decline is particularly strong for measures of cognitive speed. In the current sample, cognitive decline in accuracy performance was relatively strong in the domain of attention and nonverbal reasoning. In contrast, decline in verbal reasoning was relatively spared, as the onset was late and the decline progressed at a fairly slow pace.

These observations also fit with the differences in growth curves as derived from traditional psychometric tests. Crystalized cognitive abilities (typically measured by verbal knowledge IQ tests) continue to increase with age, whereas fluid abilities (typically measured by nonverbal cognitive processing tests) show a peak in adulthood followed by decline (Baltes, 1987; Christensen, 2001). It should be noted that our analyses are cross-sectional. This has the disadvantage that they cannot control for cohort effects like the Flynn effect. On the other hand, cross-sectional studies have the advantage that they are not influenced by retest-effects on test scores (Hofer & Sliwinski, 2001; Salthouse, 2009).

Returning to the validation part of our study, convergent validity was indicated by the association of individual test scores with general indices of educational attainment (here operationalized as years of own education and years of parental education), similar to the U.S. population. Positive correlations between cognitive performance and own and parental educational attainment were apparent, although the strengths varied considerably across measures and domains. This held for accuracy measures as well as speed measures. This reiterates the general finding that cognitive performance and educational attainment are associated (Deary & Johnson, 2010), but not equally strong for all measures (Ardila, Ostrosky-Solis, Rosselli, & Gomez, 2000).

We further demonstrated convergent validity of the CNB by the strong relation between the common variance across CNB tests and general intelligence as assessed by the WAIS using a latent factor approach. It should be noted, however, that overall scores on the CNB can never fully predict the total IQ score of the WAIS because observed scores will always be affected by measurement error. Nevertheless the high correlation between the CNB factor scores and WAIS TIQ (0.82) suggests that global measures of CNB performance can be used as a good proxy of the universally used total WAIS IQ.

The CNB is a valuable instrument not only for research, but also for clinical purposes. Clinical neuropsychological examinations regularly include intelligence and cognitive testing, because cognitive dysfunction is often a characteristic of psychiatric disorders (Millan et al., 2012). A well-known example is attention-deficit / hyperactivity disorder, but impairments in attention, memory or planning are also frequently seen in patients with schizophrenia or mood- and anxiety disorders (Castaneda, Tuuio-Henriksson, Marttunen, Suvisaari, & Lonnqvist, 2008; Heinrichs & Zakzanis, 1998; Marvel & Paradiso, 2004). Traditional neuropsychological tests are often designed to obtain a diagnosis on whether cognitive functioning is abnormal. The CNB has a similar clinical utility, since it provides quantitative measures of functioning, and yields a patients' profile of strengths and weaknesses. It may in addition shorten the clinical cognitive assessment, as obtaining global measures from the CNB makes the use of an additional psychometric intelligence test unnecessary. This reduces administration time as well as the burden for patients or participants.

Finally, the heritability analyses showed moderate estimates with wide ranges for both accuracy (1-52%) and speed (14-50%) and are in line with the studies in the U.S. samples (Calkins et al., 2010; Greenwood et al., 2007; Gur et al., 2007). In addition, estimates based on twin data closely resembled those based on family data, demonstrating that heritability estimates do not

necessarily have to be based on twin data, even though twins form a perfectly controlled design because of equal environmental factors like age and prenatal environment. Furthermore, family pedigree analyses enable the study of cross-generation resemblance. From our analyses on cognitive performance, it can be concluded that family members resemble each other mostly because of shared genetic factors, and only to a small extent due to shared environment. The relatively large component of unshared environmental factors is in agreement with other studies on specific neurocognitive traits like attention or working memory (Kremen et al., 2007; Polderman et al., 2007). Similar to heritability estimates of general intelligence (Haworth et al., 2010), the variance common to subtests showed a high heritability of 70%. This is higher than the heritability coefficients of the variance in single CNB test scores, which is in agreement with the common finding that (intelligence) subtests demonstrate lower heritability coefficients than factors of general intelligence (Kan, Wicherts, Dolan, & van der Maas, 2013). Heritability of test scores (compared to *g*) may firstly be reduced due to measurement error. Secondly, genetic effects that influence specific cognitive performance tend to accumulate as a function of the tests' specificity, with aggregated measures showing the highest heritability. As genetic effects on specific cognitive abilities become blurred in general outcome measures like '*g*', we advise future studies to focus on the specific cognitive functions, rather than general cognitive performance measures. In sum, our findings are in line with results from both research into specific neurocognitive functioning and general intelligence, providing vast evidence for the validity of the CNB.

S1. Description of CNB tests.

Tests of the CNB can be divided into five main neurobehavioral functions, each including a selection of a total of 14 cognitive domains.

The 17 tests are assessed in the following order: Motor praxis task, Emotion identification test, Continuous performance test, Face memory test, Word memory test, Letter N-back test, Face memory test - delayed, Word memory test - delayed, Conditional exclusion test, Emotion differentiation test, Finger tapping test, Matrix reasoning test, Visual object learning test, Verbal reasoning test, Age differentiation test, Line orientation test and Visual object learning test – delayed.

1. Executive-control

Conditional exclusion Test (CET)

The Conditional exclusion test measures abstraction and mental flexibility. Participants are instructed to select one out of four objects which they think does not belong. Participants are not informed about which sorting principle (line thickness, shape and size) to follow. The sorting principle changes after six consecutive correct answers. The participant receives feedback after each answer "correct" or "not correct", which may guide their next decision. There is a maximum of 48 trials, without a time limit.

There is no practice session. Accuracy is calculated as follows: number of categories achieved + 1 (to avoid a floor effect if no categories were solved) multiplied by the proportion of correct responses. Speed is the median response time of the correct responses.

Continuous performance test (CPT)

The Continuous performance test measures attention and vigilance. A 7-segment display of red vertical and horizontal lines appear in a frame (resembling a digital clock). Whenever these form a number (or letter on the second half of the test) the participant must press the spacebar as soon as possible. Both conditions are practiced before the actual test starts. Each condition consists of 30 real stimuli and 60 distractors. Stimuli are shown for 300 milliseconds, followed by a blank page for 700 milliseconds, giving 1 second to respond before the next stimulus is shown.

Accuracy is based on the number of true positives, and speed on the median response time for these true positives.

Letter N-back test (LNB)

The Letter N-back test measures working memory. This test consists of 3 conditions: 0-back, 1-back and 2-back (two sessions of each), all of which have a practice session which has to be completed successfully before the actual test begins. During the 0-back, participants are instructed to press the spacebar when the letter that appears on the screen is an "X". In the 1-back participants must press the spacebar whenever the same letter appears on the screen two times in a row. During the 2-back, the participants are supposed to press the spacebar whenever the letter on the screen is the same as the letter before the previous letter. They are instructed to do so as fast as possible, but the next trial is shown after 2.5 seconds.

Accuracy score is based on the number of true positive responses, speed is based on the median reaction time of the true positives.

2. Episodic memory

Face memory test (CPF)

The Face memory test is a measure of face memory. First, participants are shown 20 faces that they will be asked to identify later. Then -the immediate recall- participants are shown a series of 40 faces: the 20 faces they were asked to memorize mixed with 20 novel faces. During the delayed recall (15 – 45 minutes after the immediate recall), participants are again shown 40 faces: the 20 faces they were asked to memorize mixed with 20 novel faces which are different from the distracters shown during the immediate recall.

On both the immediate and delayed recall, participants are instructed to indicate for each face whether they think they have seen the face before by clicking on one of four buttons; "definitely yes", "probably yes", "probably no", and "definitely no". There is no time limit. Facial stimuli are black and white photographs of neutral expressions, balanced for gender and age.

There is no practice session. Accuracy score is based on the number of correct responses (true positives and true negatives), speed is based on the median reaction time of these correct answers.

Word memory test (CPF)

The Word memory test is a measure of verbal memory. First, participants are shown 20 words that they will be asked to identify later. Then -the immediate recall- participants are shown a series of 40 words: the 20 words they were asked to memorize mixed with 20 novel words. During the delayed recall (15 – 45 minutes after the immediate recall), participants are again shown 40 words: the 20 words they were asked to memorize mixed with 20 novel words which are different from the distracters shown during the immediate recall.

On both the immediate and delayed recall, participants are instructed to indicate for each word whether they think they have seen the word before by clicking on one of four buttons; "definitely yes", "probably yes", "probably no", and "definitely no". There is no time limit. Stimuli are equated for frequency, length, concreteness and low imageability.

There is no practice session. Accuracy score is based on the number of correct responses (true positives and true negatives), speed is based on the median reaction time of the correct answers.

Visual object learning test (VOLT)

The Visual object learning test is a measure of spatial memory. First, participants are shown 10 objects (three-dimensional Euclidean shapes) that they will be asked to identify later. Then -the immediate recall- participants are shown a series of 20 shapes: the 10 objects they were asked to memorize mixed with 10 novel shapes. During the delayed recall (15 – 30 minutes after the immediate recall), participants are again shown 20 shapes: the 10 objects they were asked to memorize mixed with 10 novel shapes which are different from the distracters shown during the immediate recall.

On both the immediate and delayed recall, participants are instructed to indicate for each shape whether they think they have seen the object before by clicking on one of four buttons; "definitely yes", "probably yes", "probably no", and "definitely no". There is no time limit.

There is no practice session. Accuracy score is based on the number of correct responses (true positives and true negatives), speed is based on the median reaction time of the correct answers.

3. Complex cognition

Matrix reasoning test (MAT)

The Matrix reasoning test is a measure of nonverbal reasoning. The participants are instructed to click on the option (out of five) that would best fit the missing part of a pattern (arrangements can be 2x2, 3x3 or 1x5). Patterns can be solved based on spatial, design or numerical relations. Items are of increasing difficulty and the test is aborted after five incorrect answers (followed by three bonus questions based on the participants' performance). There is no time limit.

The test is preceded by a practice session. Accuracy is based on the number of correct responses, speed is based on the median response time for the correct responses.

Verbal reasoning test (VRT)

The Verbal reasoning test is a measure of language reasoning. The participant must answer eight verbal analogy problems with multiple-choice answers. There is no time limit.

The test is preceded by a practice session. Accuracy is based on the percentage of correct responses, speed is based on the median response time for the correct responses.

Line orientation test (LOT)

The Line orientation test is a measure of spatial ability. The participant is presented with 24 trials in which they see a pair of lines with different orientations: the participant is supposed to rotate the blue line into parallel orientation to the fixed red line. Participants are instructed to use two buttons to rotate the blue line clockwise or counterclockwise and to use as few clicks as possible. Depending on the item, the line may rotate with 3, 6 or 9 degrees, the size of the blue line may change, and positions along the screen may vary (the distance between the centers of the red and blue line is always the same). There is no time limit.

The test is preceded by a practice session. Accuracy is based on the number of correct responses, speed is based on the median response time for the correct responses.

4. Social cognition

Emotion identification test (EI)

The Emotion identification test is a measure of emotion identification or recognition. Participants are shown a series of 40 faces, and asked to determine what emotion the face is showing. Participants respond to each trial by clicking the button corresponding to the emotion each face expresses: happy, sad, anger, fear and no emotion.

In total there are 40 trials (4 male and 4 female faces for each emotion) consisting of color photographs, balanced for intensity of emotion, age, gender and ethnicity. There is no time limit.

The test is preceded by a practice session. Accuracy is based on the number of correct responses, speed is based on the median response time for the correct responses.

Emotion differentiation test (EDT)

The Emotion differentiation test measures the ability to detect emotion intensity. The subject is presented with a pair of faces. The task is to determine which face is showing more, or a stronger, emotion (anger, fear, happiness, sadness). There are three buttons: one below each face and as a third option, the participant could choose the button in the middle "equal". There is no time limit.

There are 36 trials in total, four show no emotional difference, while the remaining 32 trials have emotion differentials between 10% - 60% (increments of 10%).

The test is preceded by a practice session. Accuracy is based on the number of correct responses, speed is based on the median response time for the correct responses.

Age differentiation test (ADT)

The Age differentiation test measures the ability to detect small visual differences. With the ADT and EDT in the test battery, it is possible to determine to what extent poor performance on the EDT is attributable to the inability to perceive small facial differences rather than a deficiency in emotion perception specifically.

The participant has to choose which face appears older (click on button below the right face) or if both faces appear to be the same age (button in between "same age"). There are 36 trials (18 male; 18 female), in four trials the two faces are identical, in the remaining 32 trials age differential ranges from 10% to 60% (increments of 10%). There is no time limit.

The test is preceded by a practice session. Accuracy is based on the number of correct responses, speed is based on the median response time for the correct responses.

5. Sensorimotor speed

Motor praxis test (MP)

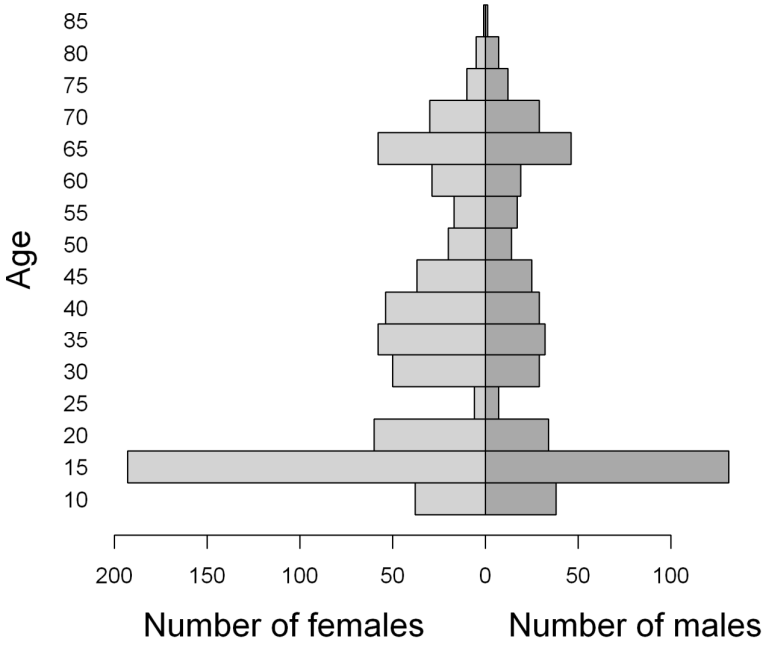
The Motor praxis test measures sensorimotor ability. It is the first test of the battery, so it also enables the participants to familiarize with the computer mouse. Participants are instructed to click the green box in the screen, which moves over different locations on screen and decreases in size. There are five seconds to respond before the next box appears.

The test is preceded by a practice session. Accuracy is based on the number of correct responses, speed on the median response time for the correct responses.

Finger tapping test (TAP)

The Finger tapping test measures motor speed and manual dexterity. Participants are asked to press the space bar with their index finger as often as possible. There are six trials, each of 10 seconds, alternating between their dominant and non-dominant hand. The test is preceded by a practice session for each hand. Speed is calculated as the total number of taps on the 6 trials.

Figure S2. Number of female and male participants per age cohort (of 5 years).



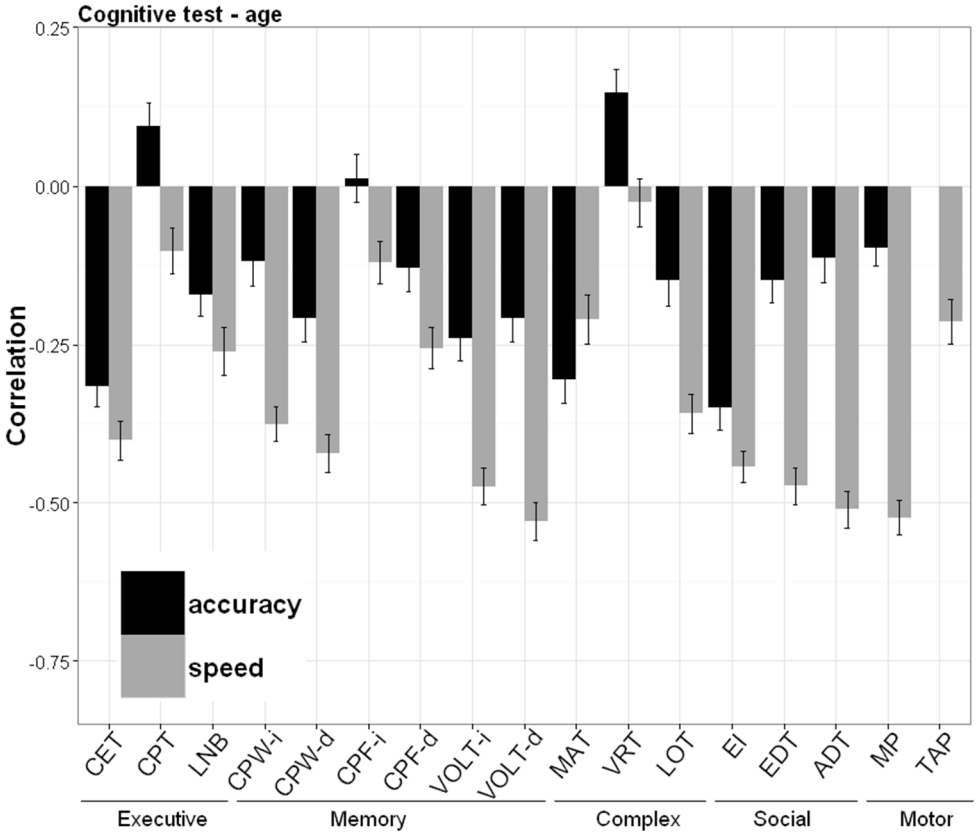
S3. Calculating an IQ score from the CNB tests.

To obtain a testee's (or group's) standardized test score, one can make use of the validation sample's average score (proportion correct, median reaction time, or number of TAPS) and calculate how many standard deviations the testee's score (or group's mean score) deviates from this average. For example, if a testee has an accuracy score of 30.0 on the LNB test, whereas the average LNB accuracy score in the validation's sample is 27.9 with a standard deviation of 2.7 (see Table 1, main text), then the testee's score deviates (30.0-27.9)/2.7 = 0.778 standard deviations from this average. Expressed as a traditional intelligence quotient score with a mean of 100 and standard deviation of 15, this deviation would denote an IQ score of 0.778*15+100 = 111.67, or 112 in whole points. Broader domain IQ scores or a total IQ score can be derived by weighing the tests specific IQ scores by their corresponding factor loadings (Table below), which are based on a single-factor model fitted in the entire validation sample. Note that each IQ score would denote a non-age effect corrected score. Hence, if used in a meta-analysis, for instance, age effects need to be regressed out.

Table S3. Correlations between accuracy scores and psychometric IQ (as measured by the WAIS).

Cognitive function (test name)	TIQ	VIQ	PIQ	g-CNB
Abstraction & flexibility (Conditional Exclusion Test)	0.13	0.23	0.20	0.43
Attention (Continuous Performance Test)	0.18	0.15	0.20	0.40
Working memory (Letter-N-Back Test)	0.09	0.10	0.12	0.43
Verbal Memory (Word Memory Test)	0.27	0.15	0.26	0.53
Verbal Memory - delayed	0.25	0.17	0.26	0.58
Face Memory (Facial Memory Test)	0.09	0.14	0.13	0.51
Face Memory - delayed	0.22	0.25	0.28	0.60
Spatial Memory (Object Learning Test)	0.18	0.31	0.27	0.50
Spatial Memory - delayed	0.18	0.34	0.30	0.50
Nonverbal reasoning (Matrix Reasoning Test)	0.29	0.23	0.31	0.52
Language reasoning (Verbal Reasoning Test)	0.27	0.29	0.33	0.58
Spatial ability (Line Orientation Test)	0.31	0.17	0.30	0.49
Emotion Identification (Emotion Identification Test)	0.37	0.60	0.56	0.69
Emotion Differentiation (Emotion Differentiation Test)	0.57	0.37	0.58	0.46
Age Differentiation (Age Differentiation Test)	0.32	0.36	0.41	0.49

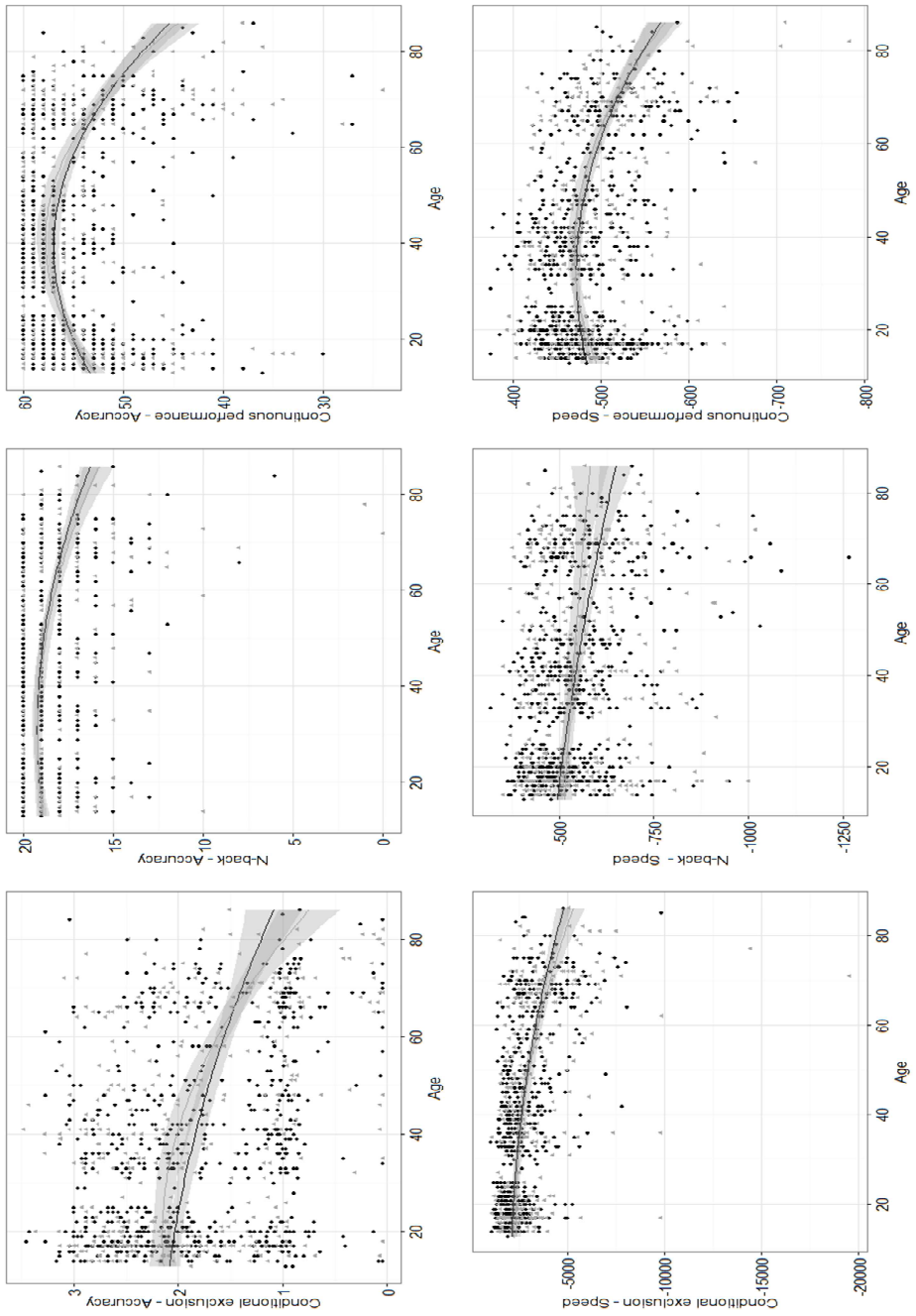
Figure S4. Correlations between the cognitive tests and participants' age (including 95% confidence intervals). Correlations with accuracy scores are given in black and with speed scores in grey.



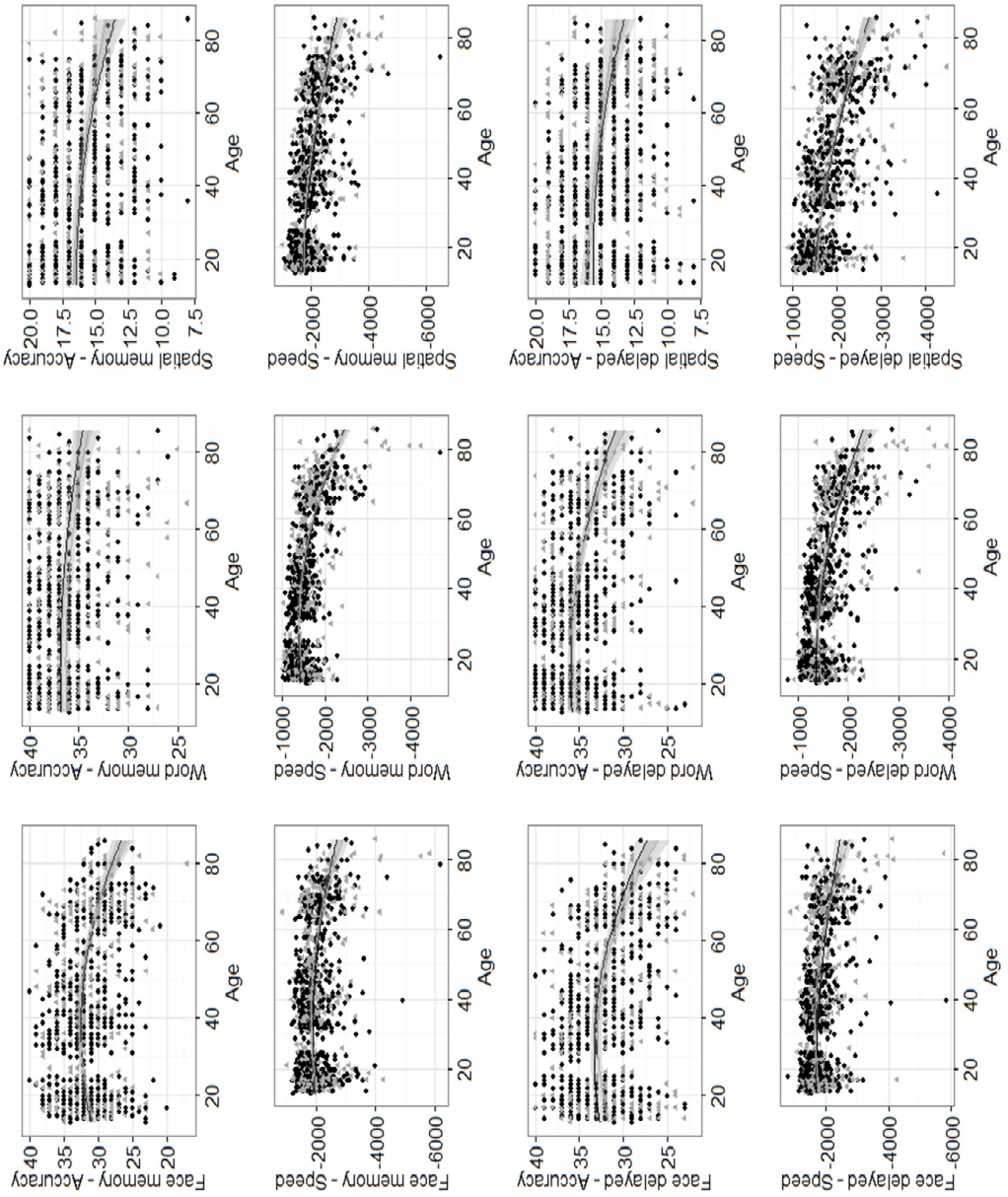
See Table 1 for abbreviations of cognitive tests.
No accuracy score available for TAP.

Figure S5. Illustration of non-linear effects in all cognitive functions.

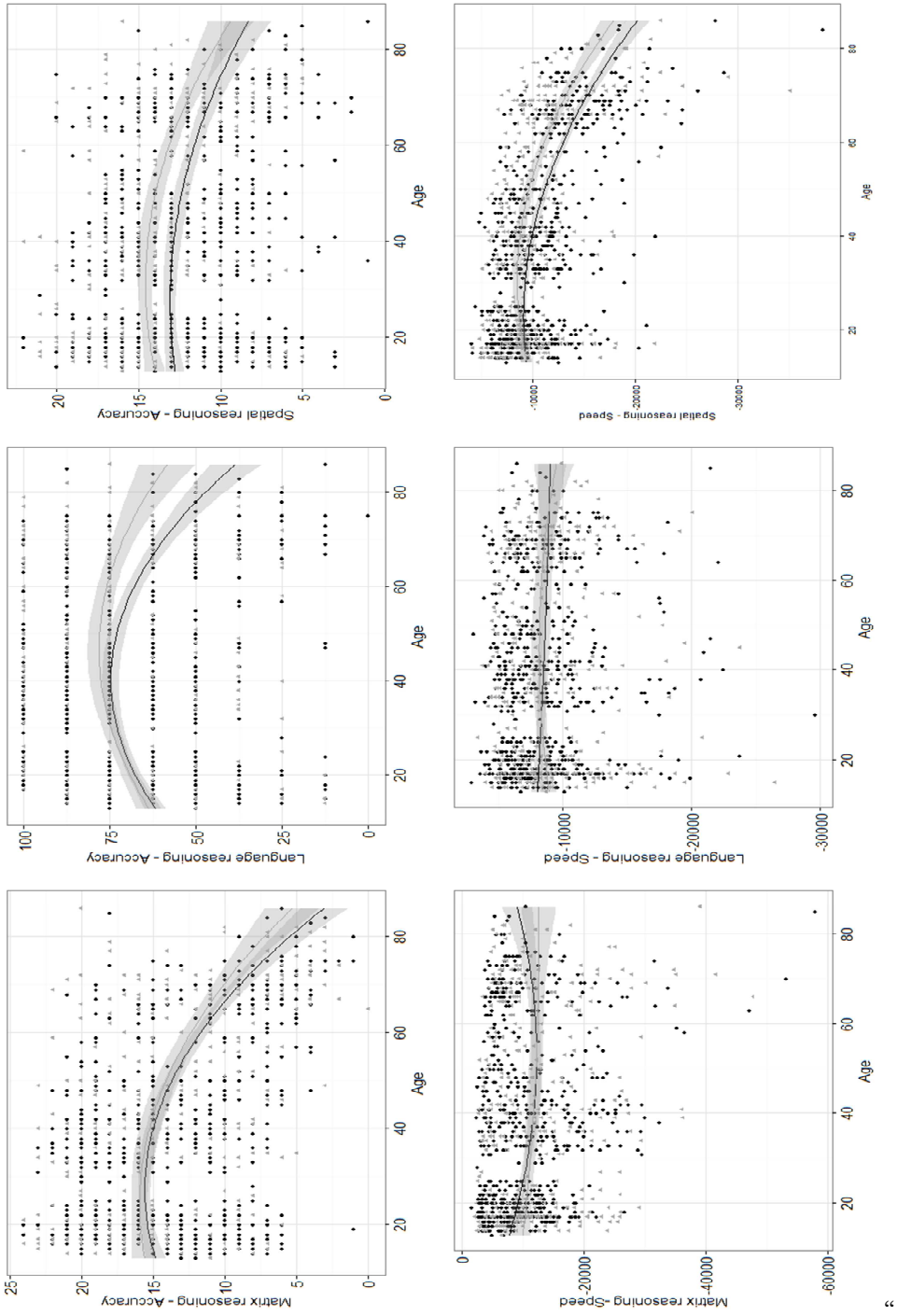
5.1 Executive control (females black ●, males grey ▲).



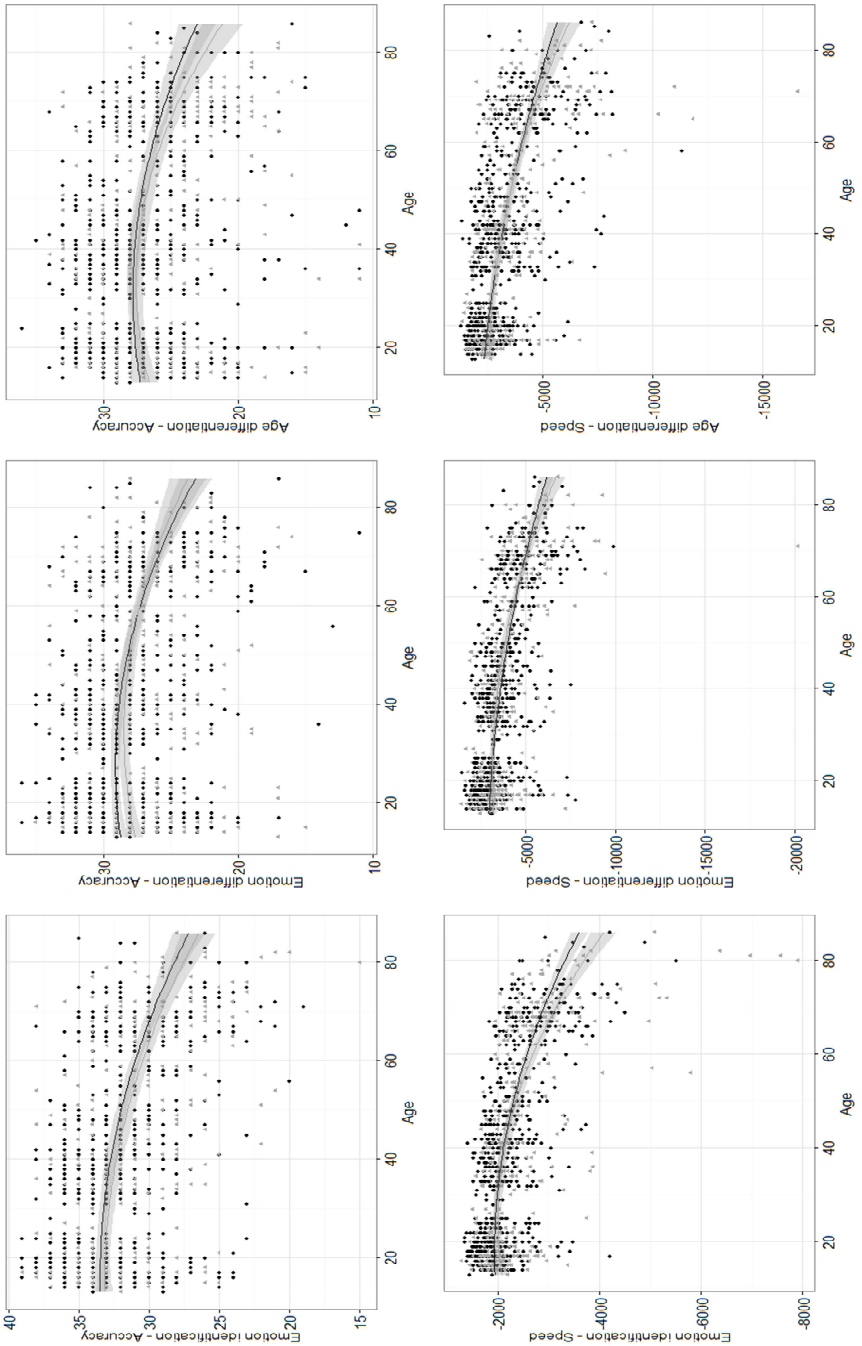
5.2 Memory (females black ●, males grey ▲).



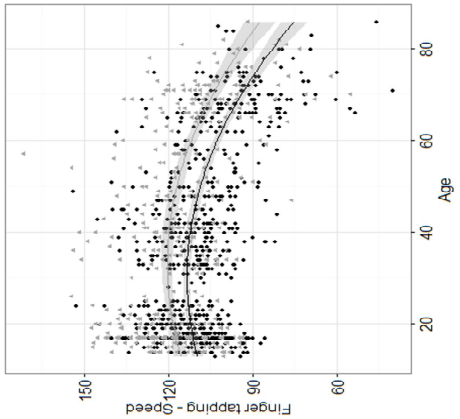
5.3 Complex Cognition (females black ●, males grey ▲).

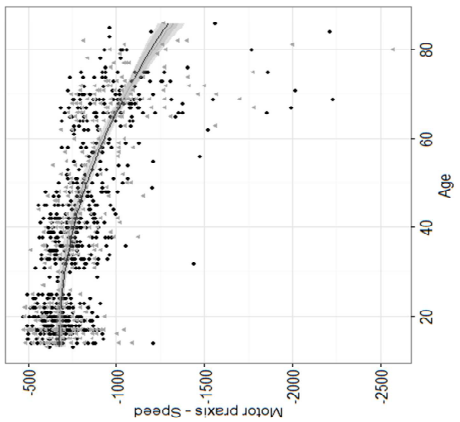


5.4 Social Cognition (females black ●, males grey ▲).



5.5 Sensorimotor (females black ●, males grey ▲).





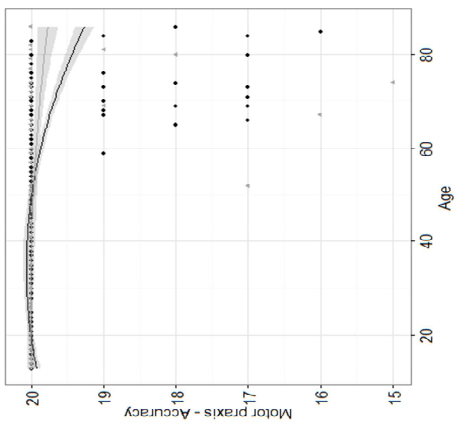


Table S6. Twin correlations of monozygotic (MZ) and dizygotic (DZ) twin pairs. Variance explained by common and unique environmental factors, based on analyses in twins and in all family members. Fit indices for the twin analyses are given (-2 log likelihood), and the p-value of the fit of the saturated model versus the ACE model (6 degrees of freedom).

		Twin correlations		Common environment		Unique environment		Fit indices twin analyses		
Cognitive function		MZ	DZ	Twins	All family members	Twins	All family members	saturated	ACE	*p*
Executive Control										
Abstraction & flexibility	Accuracy	.18	-.07	0	0	88	87	1268.18	1285.28	.01
	Speed	.47	.19	0	11	59	51	4856.29	4865.24	.18
Attention	Accuracy	.34	.09	0	0	58	62	3550.49	3591.77	.00
	Speed	.43	.22	9	0	61	60	3340.40	3348.59	.22
Working memory	Accuracy	.29	.16	5	0	72	78	1994.01	2007.6	.03
	Speed	.35	.24	18	15	67	54	4173.23	4148.76	.20
Memory										
Face Memory	Accuracy	.35	.22	5	3	64	63	3101.48	3110.31	.18
	Speed	.48	.10	0	4	57	60	4447.94	4462.00	.03
delayed	Accuracy	.31	.16	0	0	65	69	3109.02	3119.60	.10
	Speed	.53	.19	0	2	51	55	4349.17	4364.67	.02
Verbal Memory	Accuracy	.33	.00	0	1	73	74	2824.64	2833.53	.18
	Speed	.44	.15	0	0	59	56	3898.54	3903.49	.55
delayed	Accuracy	.16	.05	0	1	84	81	3001.84	3013.57	.07
	Speed	.39	.13	0	0	64	64	4027.91	4044.89	.01
Spatial Memory	Accuracy	.32	.08	0	0	70	69	2643.09	2650.45	.29
	Speed	.36	.11	0	0	67	67	4336.58	4341.37	.07
delayed	Accuracy	.33	.07	0	0	69	70	2662.83	2674.09	.08
	Speed	.40	.09	0	3	65	64	4327.90	4331.20	.77

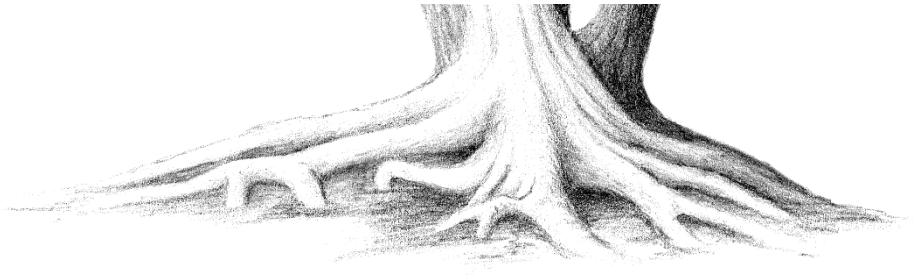
Table S6 - continued.

		Twin correlations		Common environment		Unique environment		Fit indices twin analyses		
Cognitive function		MZ	DZ	Twins	All family members	Twins	All family members	saturated	ACE	*p*
Complex Cognition										
Nonverbal reasoning	Accuracy	.52	.27	1	15	47	45	3451.74	3453.53	.94
	Speed	.49	.21	0	10	54	57	5603.89	5609.65	.45
Language reasoning	Accuracy	.33	.05	0	0	71	63	5241.05	5249.41	.21
	Speed	.38	.20	6	0	63	69	5995.85	5012.20	.01
Spatial ability	Accuracy	.51	.09	0	0	54	51	3170.07	3177.73	.26
	Speed	.40	.22	4	10	61	60	4663.21	4674.01	.25
Social Cognition										
Emotion Identification	Accuracy	.30	.23	7	30	66	56	3003.81	3008.67	.56
	Speed	.40	.22	8	0	62	63	4182.68	4188.64	.43
Emotion Differentiation	Accuracy	.24	.16	17	12	79	71	3126.73	3138.59	.07
	Speed	.44	.24	15	9	61	56	4562.96	4579.85	.01
Age Differentiation	Accuracy	.25	.24	24	15	76	63	3332.47	333.17	.99
	Speed	.44	.14	0	17	60	51	4796.00	4801.86	.44
Sensorimotor										
Sensorimotor Speed	Speed	.43	.33	23	16	58	40	4003.22	4009.18	.43
Motor speed	Speed	.38	.11	0	29	62	40	3542.16	3553.70	.07
General Intelligence										
g-CNB		.70	.29	0	5	30	27	2305.43	2310.39	.55
Total IQ		.76	.51	0	-	25	-	1322.18	1324.44	.89

Chapter 6

Domain dependent associations between cognitive functioning and regular voluntary exercise behavior

This chapter is published as: Swagerman, S. C., de Geus, E. J. C., Koenis, M. M. G., Hulshoff Pol, H. E., Boomsma, D. I., & Kan, K. J. (2015). Domain dependent associations between cognitive functioning and regular voluntary exercise behavior. *Brain & Cognition*, *97*, 32-39.



Introduction

Regular exercise has often been suggested to have beneficial effects on cognitive performance, but empirical findings do not always support this suggestion. As a result, the effectiveness of regular exercise behavior as a means to improve cognitive performance remains a subject of debate, not only among scientists, but also among policy makers. When published findings are summarized, associations between exercise behavior and cognitive performance appear positive on average, but vary considerably in strength (Fedewa & Ahn, 2011; Hindin & Zelinski, 2012; Ploughman, 2008; Singh et al., 2012; Taras, 2005; Trudeau & Shephard, 2008; Verburgh, Konigs, Scherder, & Oosterlaan, 2014). The literature provides four major sources of heterogeneity among study outcomes, the first concerning sample constitution (Singh et al., 2012). Study samples have differed greatly with respect to age, while the association strength between exercise behavior and cognitive performance is considered to differ between children, adolescents and adults (Hillman, Castelli, & Buck, 2005; Tomporowski, Davis, Miller, & Naglieri, 2008; but see Verburgh et al., 2014).

In childhood and adolescence exercise may influence the (rapid and specific) brain changes that take place during development, while in the elderly exercise may prevent (slow or general) deterioration of the brain during aging (Churchill et al., 2002; Fabel & Kempermann, 2008; Greenwood & Parasuraman, 2010; Hillman, Erickson, & Kramer, 2008; Kraft, 2012; Yuki et al., 2012). Furthermore, rates of cognitive decline differ across sexes, which has been linked to the loss of estrogen (Kramer, Erickson, & Colcombe, 2006). Sex may be regarded as a source of heterogeneity in itself as the associations between exercise behavior and cognitive measures in samples consisting of a majority of women tend to be larger than in samples consisting of relatively many men (Colcombe & Kramer, 2003). A second major source of heterogeneity amongst study outcomes concerns the cognitive domain being measured. Recent studies (Colcombe & Kramer, 2003) suggest that cognitive functions are differently susceptible to exercise; executive functions may be more sensitive to exercise than, for example, long-term memory. Empirically however, little is known about how effects of exercise vary across cognitive domains, let alone about how these effects differ in their dependencies on age and sex. Many studies have focused on global cognitive measures, and outcomes thereof, such as academic achievement. This is unfortunate because they do not inform about the sensitivity of specific cognitive functions (Tomporowski et al., 2008). The present study is unique, in that we measured in a single, population representative sample cognitive performance across a wide range of well-defined, specific cognitive domains. The battery we used, the web-based Computerized

Neurocognitive Battery (CNB), consists of 17 cognitive tests and provides measures of accuracy as well as speed in the following cognitive domains: abstraction and mental flexibility, attention, working memory, memory (verbal, face, and spatial), language and nonverbal reasoning, spatial ability, emotion identification, emotion- and age differentiation, sensorimotor speed, and motor speed. Individual differences in these domains are substantially heritable and demonstrate genetic linkage (Almasy et al., 2008). Scores on the CNB are reliable and compare well to scores on traditional pen-and-paper tests in healthy samples as well as in clinical samples (e.g., schizophrenia patients, Gur et al., 2001a; Gur et al., 2001b). While initially constructing the test battery, tests were selected from neuroimaging studies that showed selective activation of specific brain systems in the magnetic resonance imaging (MRI) scanner (Gur et al., 2010). Recently, the CNB tests adapted for administration in the MRI scanner replicated the brain areas that are activated by the CNB's cognitive domains. More specifically, the executive tests activated mainly frontal areas, memory tests involved anterior medial temporal regions, and a test measuring emotion identification activated temporo-limbic regions (Roalf et al., 2014).

A third source of heterogeneity amongst previous results, the definition and reliability of exercise behavior measures, has been discussed extensively in the literature. Studies have varied greatly in the conceptualization of exercise behavior, the broadest conceptualization being the inclusion of all forms of physical activity (i.e. every activity increasing energy expenditure above basal metabolic rate). However, self-reported physical activity corresponds poorly with actual physical activity (Prince et al., 2008). In addition, the idea that common, low intensity forms of physical activity will be sufficient to induce cognitive effects has been questioned; exercise likely needs to be carried out at a moderate to vigorous intensity to have effect on cognitive functioning (Colcombe & Kramer, 2003; Fedewa & Ahn, 2011; Hindin & Zelinski, 2012). It is recommended to focus on relatively vigorous activities, especially leisure time exercise activities: recall is relatively easy and quite accurate as these activities are self-initiated and often clearly defined in time. Indeed, voluntary regular leisure time exercise behavior demonstrated excellent test-retest reliability (de Moor, Boomsma, Stubbe, Willemsen, & de Geus, 2008; Stubbe, de Moor, Boomsma, & de Geus, 2007). In the present study, we will focus on this narrow but well-defined behavior, also because it is often the main target of health-promoting exercise interventions (Kahn et al., 2002).

A fourth source of heterogeneity concerns study design. This is an important source to recognize, because study designs are differently suited to estimate effects of physical activity. In experimental and clinical intervention studies the focus is usually on mean effects as a result of intervention, while the focus of observational studies lies on individual differences in voluntary behavior and on dose-response relationships. Furthermore, intervention studies –experimental studies included- have varied widely in their definition of intervention. In addition, not all intervention studies have been truly experimental; clinical intervention is often performed in non-random samples (Singh et al., 2012; Tomporowski et al., 2008). Another distinction concerns studies investigating the effects of acute physical exercise, and studies that investigate the effects of chronic physical exercise (Verburgh et al., 2014). In the first, the focus is on (short-term) cognitive enhancement right after a single bout of exercise, typically within less than an hour. In the latter, the focus is on (long-term) cognitive enhancement as the result of regular exercise over longer periods, typically weeks or months. Although there is ample evidence for beneficial effects of acute physical exercise (Verburgh et al., 2014), studies into the effects of chronic physical exercise are scarce, hence the call for more research.

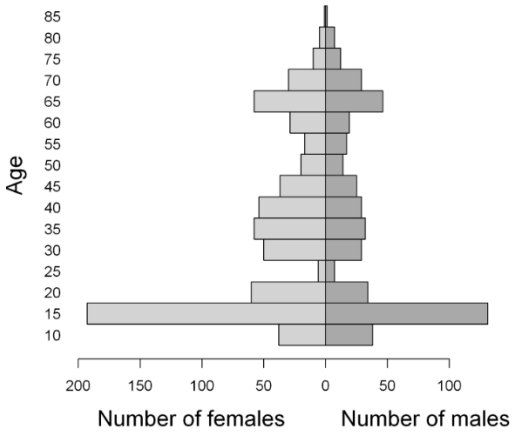
The general objective of the present study is to investigate the chronic dose-response association between voluntary regular leisure time exercise behavior and cognitive performance across a wide range of cognitive domains, while controlling for other sources of heterogeneity. To this end, we first examine whether leisure time exercise associated with accuracy and speed scores, exploring whether and how these associations vary across domains. Next, we explore whether, how, and to what extent these associations vary when accounting for differences in age and sex. We end with a general discussion, in which the results of the present population-based observational study are compared with results from previous (high quality) intervention studies, which typically involve clinical-control designs.

Material and methods

Participants

The subject sample consisted of 472 males and 668 females from the Netherlands Twin Register (NTR) recruited from all over the Netherlands (Boomsma et al., 2006; van Beijsterveldt et al., 2013; Willemsen et al., 2013). The majority (n = 1110) was comprised of twin pairs and their family members (parents, children, siblings, and spouses) who volunteered in NTR projects. The rest (n = 30) was comprised of undergraduate students who piloted in these projects. The participants ranged in age from 10 to 86 years old (M = 37.73, SD = 20.86, see Figure 1).

Figure 1. Age distribution in females and males.



Procedure

Studies and procedures were approved by the Medical Ethics Review Committee of the Vrije Universiteit Medical Center Amsterdam and the Central Committee on Research Involving Human Subjects. The twins and their family members were approached by mail. In case of a positive response, a structured telephone call followed, which was informative about possible exclusion criteria (epilepsy, paralysis). The students were recruited at the university through flyers. They signed up themselves. Data collection took place either at home (n = 536) or in a laboratory (Vrije Universiteit Amsterdam, University Medical Center Utrecht, Amsterdam Medical Center, n = 604).

Cognitive performance was assessed on a 15 inch Macbook laptop, using the web-based Computerized Neurocognitive Battery (CNB, see below). The test administrator was placed behind the participant to be able to read the test instructions out loud and to provide feedback during practice trials. The administrator judged for each test if it was complete and valid (for example based on motivation or attention). On designated timepoints in between tests, the procedure, which lasted 1.30 hours on average, could be paused. Students received study credits, others travel compensation and a gift voucher. All participants signed an informed consent form. For participants under 16 years parents gave additional written consent. All participants received feedback on their performance.

Materials

Cognitive performance

Cognitive performance (accuracy and speed) was assessed by the Dutch translation of the CNB as described by Gur et al. (2010; 2012). It comprises a total of 17 tests that assess performance on 5 neurobehavioral domains of executive control, memory, complex cognition, social cognition, and sensorimotor speed (Table 1). Accuracy was defined as either the percentage or the number of correct responses on a test, whereas speed was defined as minus the median response time (R*-1) in milliseconds for correct responses. Speed performance on the Finger-tapping test (TAP), however, was expressed as the number of taps one can produce within 60 seconds (alternating every 10 seconds between the left and right hand). TAP score thus indicates speed, but motor speed rather than response time. For all cognitive measures it held that higher scores reflected better cognitive performance.

Voluntary regular leisure time exercise behavior

Questions on exercise behavior were collected using a standardized interview (on the same day as the cognitive testing, $n = 894$) or a questionnaire (within 2 weeks of cognitive testing, $n = 246$) with identical questions. The first question was "Do you exercise regularly?". When the answer was affirmative further information was gathered on the type of exercise (for example aerobics classes, soccer, or running) and on the involvement in this type of exercise (months a year, times a week, and average duration of the activity in minutes). Activities were excluded if they are not self-initiated or voluntary, like transportation (walking, biking), or physical education classes in school, as were general physical activities such as gardening. Voluntary exercise activities were only scored when participants had engaged in them for at least three months during the past year (Stubbe, Boomsma, & de Geus, 2005).

Next, we obtained the metabolic equivalent (MET) for each of the reported activities. Here, a MET = 1 corresponds to the rate of energy expenditure of an individual at rest (approximately one kcal/kg/h). Because children and adults differ in the energy cost of activities, MET scores of participants under age 18 were obtained using the Compendium of Physical Activities for Youth (Ridley, Ainsworth, & Olds, 2008), and of older participants using the Ainsworth's compendium of physical activities (Ainsworth et al., 1993). Finally, we computed each individual's weekly METhours by multiplying each activity's MET by the hours per week spent on each activity and by summing these up over the exercise activities. Non-exercisers received a weekly METhour score of 0. Previous

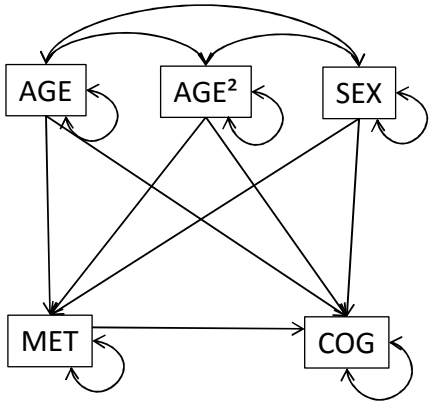
studies have shown that this variable has a high 6-month test-retest reliability of 0.82 (de Moor et al., 2008).

Statistical Analyses

Descriptive statistics, which were calculated using SPSS 21.0 (IBM Corp., 2011), included means and standard deviations of the measured variables and of test administration time (Table 1). Cronbach's alpha coefficients of internal consistency, which are commonly interpreted as indicators of reliability, were also calculated. To explore whether voluntary regular leisure time exercise behavior associated with cognitive performance, we ran for each cognitive variable a univariate regression model, in which cognitive performance was regressed on weekly METhours. Next, in order to statistically control for the effects of age and sex we fitted a multivariate path model (see Figure 2) in which cognitive performance was regressed on weekly METhours, sex and (linear and quadratic terms of) age, while weekly METhours was regressed on sex and age. The quadratic age term was defined as the square of grand mean centered values and was included because inspection of the raw data suggested nonlinear relationships between cognitive performance and age (Figures 3A and 3B provide examples). Sex and age were allowed to intercorrelate (although their intercorrelations were expected not to differ from zero). Cases were excluded from statistical analyses whenever participants were considered to experience too much difficulty (n = 1) or when test performance was judged invalid by the experimenter (~0.9%), for example when computer or mouse issues had occurred, when participants demonstrated a lack of motivation, or when participants reported (noncognitive) impairments such as rheumatoid arthritis or vision problems. Data of children in elementary school (under age 13) were removed from analyses (n = 4).

Because the majority of the cognitive variables were non-normally distributed (skewed) and scores were family clustered, the standard errors of the parameter estimated required correction. Correction was accomplished by analyzing the data in R using package lavaan (Rosseel, 2012) which included the option to use a robust sandwich estimation procedure and family number as cluster variable. This procedure allows for the analysis of nonnormally distributed, continuous outcome variables. Because analyses were carried out for 33 cognitive, possibly related, measures, the Matrix Spectral Decomposition program (Li & Ji, 2005) was used to estimate the number of independent dimensions in the data, which was 23. This yielded a preferred significance level of $\alpha = 0.05/23 \approx 0.002$.

Figure 2. Graphical representation of the multivariate model. The relation between cognitive test performance (COG) and weekly METhours (MET) depends on sex and age.



Results

Descriptive statistics

The descriptive statistics are provided in Table 1. Figure 3C illustrated the complete distribution of weekly METhours across age and sex. The mean weekly METhours in the total sample was 15.6, males scoring higher (20.2) than females (12.3, β = 7.97, p < 0.001), as did young participants compared to older participants (r = -0.23, p < 0.001).

Modeling results

Table 2 and Figure 4 summarize the modeling results. The univariate model yielded standardized regression coefficients that can be interpreted as bivariate correlations between weekly METhours and cognitive performance. With respect to accuracy these ranged from -0.02 (VRT) to 0.14 (MAT) and with respect to speed from -0.01 (reasoning tests MAT and VRT) to 0.18 (TAP), hence from negatively small to positively small. Medians were also small (0.09 for accuracy and 0.07 for speed), yet at α = 0.002 about half of the coefficients were significant. However, the multivariate model, which yielded standardized path coefficients that can be interpreted as partial correlations, demonstrated that direct relationships were small and centered close to 0. Coefficients for accuracy ranged from -0.03 (CPF-d, VRT) to 0.11 (CPT, median = 0.03). And for speed coefficients ranged from -0.05 (EDT, ADT) to 0.06 (TAP, median = -0.02). Only the coefficient between weekly METhours and accuracy on the attention test (CPT) was significant (β = 0.11, se = 0.03, p < 0.001).

Table 1. Cognitive domains and test names, and mean administration duration (Time, in minutes), number of participants who completed the test (N), and the test's mean score (M, and SD), Cronbach's alpha coefficients (α) of accuracy score (percentage or number of correct responses) and speed (median response time, in ms).

					Accuracy			Speed		
Executive Control	Test name	Label	Time	N	M	SD	α	M	SD	α
Abstraction / flexibility	Penn Conditional Exclusion Test [a]	CET	4.9	1125	1.9	0.8	[d]	2813.3	1392.6	[d]
Attention	Penn Continuous Performance Test [a]	CPT	5.3	1125	54.8	5.4	.86	487.7	49.1	.82
Working memory	Letter-N-Back Test [a]	LNB	9.2	1114	18.8	1.8	.77	537.7	118.0	.80
Memory										
Verbal Memory	Penn Word Memory Test [b]									
	- immediate	CPW-i	3.1	1125	36.3	2.8	.62	1564.5	368.2	.92
	- delayed	CPW-d	1.1	1124	35.0	3.3	.64	1541.7	376.6	.91
Face Memory	Penn Facial Memory Test									
	- immediate	CPF-i	3.9	1123	31.4	3.5	.56	1992.7	544.2	.92
	- delayed	CPF-d	1.5	1121	32.1	3.5	.57	1834.2	489.7	.89
Spatial Memory	Visual Object Learning Test [a]									
	- immediate	VOLT-i	2.7	1117	16.0	2.3	.48	1973.8	554.6	.87
	- delayed	VOLT-d	0.5	1115	15.4	2.4	.48	1811.5	519.7	.86
Complex Cognition										
Nonverbal reasoning	Penn Matrix Reasoning Test	MAT	7.8	1129	13.9	5.2	.90	10806.0	6959.8	.88
Language reasoning	Penn Verbal Reasoning Test [a,b]	VRT	1.8	1123	69.2	20.6	.53	8465.8	3332.5	.74
Spatial ability	Variable Penn Line Orientation Test [a]	LOT	5.5	1119	12.9	3.7	.79	10506.8	3861.8	.97
Social Cognition										
Emotion Identification	Penn Emotion Identification Test	EI	2.3	1132	32.1	3.5	.62	2273.4	685.7	.92
Emotion differentiation	Measured Emotion Differentiation	EDT	3.4	1131	28.0	3.5	.69	3721.0	1369.1	.94
Age Differentiation	Age Differentiation Test	ADT	3.0	1122	26.8	3.9	.74	3238.4	1493.5	.94
Sensorimotor										
Sensorimotor speed	Motor Praxis Test	MP	1.8	1130	20.0	0.4	.93	793.2	221.3	.95
Motor speed	Computerized Finger-Tapping Test[a]	TAP	3.5	[c]	[c]	[c]	[c]	110.6	15.1	.96

[a] short test version. [b] different items for children. [c] no accuracy score available for TAP. [d] not amenable for calculating

Table 2. Results from the univariate and multivariate analyses

	Univariate analyses		*Multivariate analyses*							
	Accuracy on		*Accuracy* on							
Cognitive Domain	*Weekly MET-h*		*Weekly MET-h*		*Accuracy* on *Age*		*Accuracy* on *Age*^{2}		*Accuracy* on *Sex*	
	β (SE)	*p*	*β (SE)*	*p*	*β (SE)*	*p*	*β (SE)*	*p*	*β (SE)*	*p*
Executive Control										
Abstraction / flexibility	.13 (.03)	.00	.04 (.03)	.10	-.30 (.03)	.00	-.11 (.04)	.00	.05 (.03)	.12
Attention	.11 (.03)	.00	.11 (.03)	.00	.13 (.04)	.00	-.45 (.04)	.00	.00 (.03)	.97
Working memory	.07 (.03)	.02	.03 (.03)	.40	-.16 (.03)	.00	-.24 (.05)	.00	-.03 (.03)	.35
Memory										
Verbal Memory	.04 (.03)	.18	.03 (.03)	.35	-.11 (.04)	.01	-.09 (.04)	.03	-.10 (.03)	.00
	.10 (.03)	.00	.05 (.03)	.05	-.19 (.04)	.00	-.21 (.04)	.00	-.05 (.03)	.07
Face Memory	.02 (.03)	.44	.00 (.03)	.89	.02 (.04)	.71	-.32 (.04)	.00	-.02 (.03)	.62
	.01 (.03)	.79	-.03 (.03)	.38	-.14 (.04)	.00	-.26 (.04)	.00	-.09 (.03)	.01
Spatial Memory	.08 (.03)	.01	.01 (.03)	.80	-.24 (.04)	.00	-.09 (.04)	.01	.04 (.03)	.17
	.09 (.03)	.00	.02 (.03)	.42	-.20 (.04)	.00	-.07 (.04)	.05	.07 (.03)	.02
Complex Cognition										
Nonverbal reasoning	.14 (.03)	.00	.04 (.03)	.16	-.29 (.04)	.00	-.26 (.04)	.00	.05 (.03)	.06
Language reasoning	-.02 (.03)	.50	-.03 (.03)	.26	.14 (.04)	.00	-.30 (.04)	.00	.12 (.03)	.00
Spatial ability	.11 (.03)	.00	.03 (.03)	.34	-.14 (.04)	.00	-.17 (.04)	.00	.18 (.03)	.00
Social Cognition										
Emotion Identification	.13 (.03)	.00	.06 (.03)	.04	-.34 (.04)	.00	-.16 (.04)	.00	-.07 (.03)	.01
Emotion Differentiation	.05 (.03)	.07	.01 (.03)	.59	-.15 (.04)	.00	-.21 (.04)	.00	-.07 (.03)	.03
Age Differentiation	.04 (.03)	.24	.01 (.03)	.75	-.11 (.04)	.00	-.21 (.04)	.00	-.07 (.03)	.02
Sensorimotor										
Sensorimotor speed	.10 (.03)	.00	.05 (.02)	.00	-.08 (.03)	.00	-.23 (.07)	.00	.04 (.03)	.21

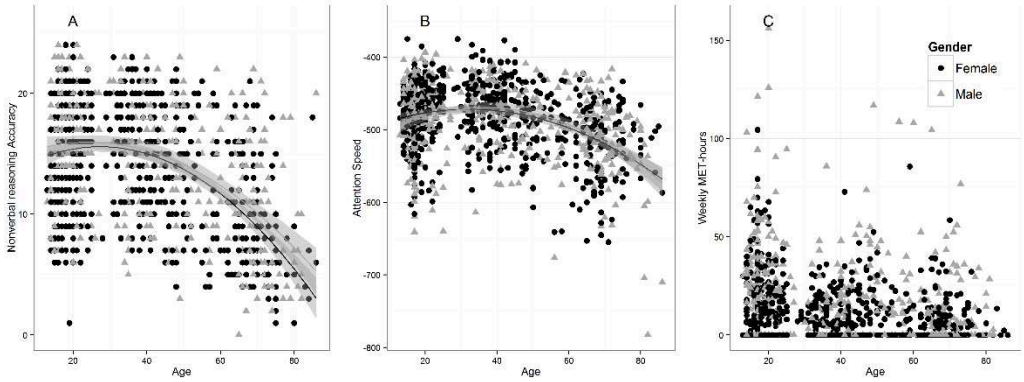
β = standardized regression coefficient, SE = standard error, p = p-value.

Table 2 - continued.

	Univariate analyses		*Multivariate analyses*							
	Speed on		*Speed* on							
Cognitive Domain	*Weekly MET-h*		*Weekly MET-h*		*Speed* on *Age*		*Speed* on *Age*^{2}		*Speed* on *Sex*	
	β (SE)	*p*	*β (SE)*	*p*	*β (SE)*	*p*	*β (SE)*	*p*	*β (SE)*	*p*
Executive Control										
Abstraction / flexibility	.06 (.03)	.03	-.03 (.02)	.16	-.41 (.03)	.00	-.14 (.04)	.00	-.05 (.03)	.09
Attention	.03 (.03)	.29	-.01 (.03)	.75	-.11 (.04)	.00	-.31(.04)	.00	-.01 (.03)	.65
Working memory	.05 (.03)	.13	-.02 (.03)	.62	-.26 (.04)	.00	-.02 (.04)	.62	.03 (.03)	.33
Memory										
Verbal Memory	.12 (.03)	.00	.01 (.02)	.57	-.37 (.03)	.00	-.35 (.05)	.00	-.02 (.02)	.33
	.10 (.03)	.00	-.01 (.02)	.73	-.42 (.03)	.00	-.27 (.04)	.00	-.03 (.02	.15
Face Memory	.06 (.03)	.06	.01 (.03)	.75	-.12 (.03)	.00	-.26 (.05)	.00	.00 (.03)	.90
	.06 (.03)	.03	-.01 (.03)	.68	-.26 (.03)	.00	-.21 (.05)	.00	.01 (.03)	.86
Spatial Memory	.09 (.03)	.00	-.03 (.02)	.22	-.48 (.03)	.00	-.15 (.04)	.00	.00 (.03)	.98
	.12 (.03)	.00	-.01 (.02)	.58	-.53 (.03)	.00	-.12 (.04)	.00	-.01 (.03)	.83
Complex Cognition										
Nonverbal reasoning	-.01 (.03)	.71	-.04 (.03)	.14	-.23 (.04)	.00	.12 (.06)	.05	-.06 (.03)	.04
Language reasoning	-.01 (.03)	.62	-.03 (.03)	.35	-.05 (.04)	.18	-.02 (.04)	.57	-.01 (.03)	.85
Spatial ability	.10 (.03)	.00	-.04 (.02)	.11	-.37 (.03)	.00	-.30 (.04)	.00	.13 (.03)	.00
Social Cognition										
Emotion Identification	.07 (.04)	.06	-.04 (.03)	.20	-.46 (.03)	.00	-.27 (.04)	.00	-.07 (.03)	.01
Emotion Differentiation	.07 (.04)	.04	-.05 (.03)	.10	-.49 (.03)	.00	-.18 (.03)	.00	.00 (.03)	.95
Age Differentiation	.08 (.04)	.07	-.05 (.03)	.13	-.53 (.03)	.00	-.13 (.03)	.00	-.03 (.02)	.23
Sensorimotor										
Sensorimotor speed	.16 (.03)	.00	.02 (.02)	.28	-.52 (.03)	.00	-.25 (.04)	.00	.01 (.02)	.78
Motor speed	.18 (.03)	.00	.06 (.03)	.02	-.19 (.03)	.00	-.34 (.03)	.00	.24 (.03)	.00

β = standardized regression coefficient, SE = standard error, p = p-value.

Figure 3A+B. Examples of the nonlinear relationships between age and cognitive performance (accuracy and speed), including 95% intervals around the quadratic regression lines.
Figure 3C. Illustration of non-normal distribution of weekly METhours against the age of female (black ●) and male (grey ▲) participants.



Discussion

The aim of this paper was to scrutinize the domain dependency of the association between exercise behavior and cognition, while controlling for other major sources of heterogeneity. To this end, we explored in a population based sample, and across a wide range of cognitive domains, the age and sex independent associations using reliable and narrowly defined measures of voluntary regular leisure time exercise behavior.

Univariate analyses confirmed the existence of multiple associations between regular exercise behavior and cognitive performance. At face value, these findings may seem to support the idea of beneficial effects of regular leisure time exercise on cognitive accuracy and speed, however, this interpretation requires some caution. First, in line with results from reviews, the majority of the associations between exercise behavior and cognitive measures were positive, but associations varied in strength; null effects, including ones in the negative direction, were also found. This pattern thus reiterates the heterogeneous findings in the literature and implies that not all cognitive functions may benefit equally from voluntary exercise.

Second, our analyses clearly demonstrate the presence of confounding effects. Sex differences were established in both exercise behavior and cognitive performance and these varied in sign and strength across cognitive domains. Exercise behavior and cognitive performance decreased with age, also replicating previous findings (de Moor, Beem, Stubbe, Boomsma, & de Geus, 2006). The linear and quadratic associations with age were found to vary considerably across cognitive domains. Effects of age, sex, and exercise on

cognitive performance can thus be confounded, while the magnitude of the confounding effect is dependent on the specific cognitive domain.

After regressing out sex and age, and while using a liberal significance level of $\alpha = 0.05$, only four out of 33 relationships between weekly METhours and cognitive performance would reach the level of significance. This is close to the number of expected false positives. Publication bias in previously reported results is thus a serious issue. When proper correction for multiple testing is applied, the association between weekly METhours and cognitive performance may not survive statistical scrutiny. With exception of the Continuous performance test (CPT), none of the standardized regression coefficients in our study was above 0.1 (or below -0.1), therefore any effects must be considered small.

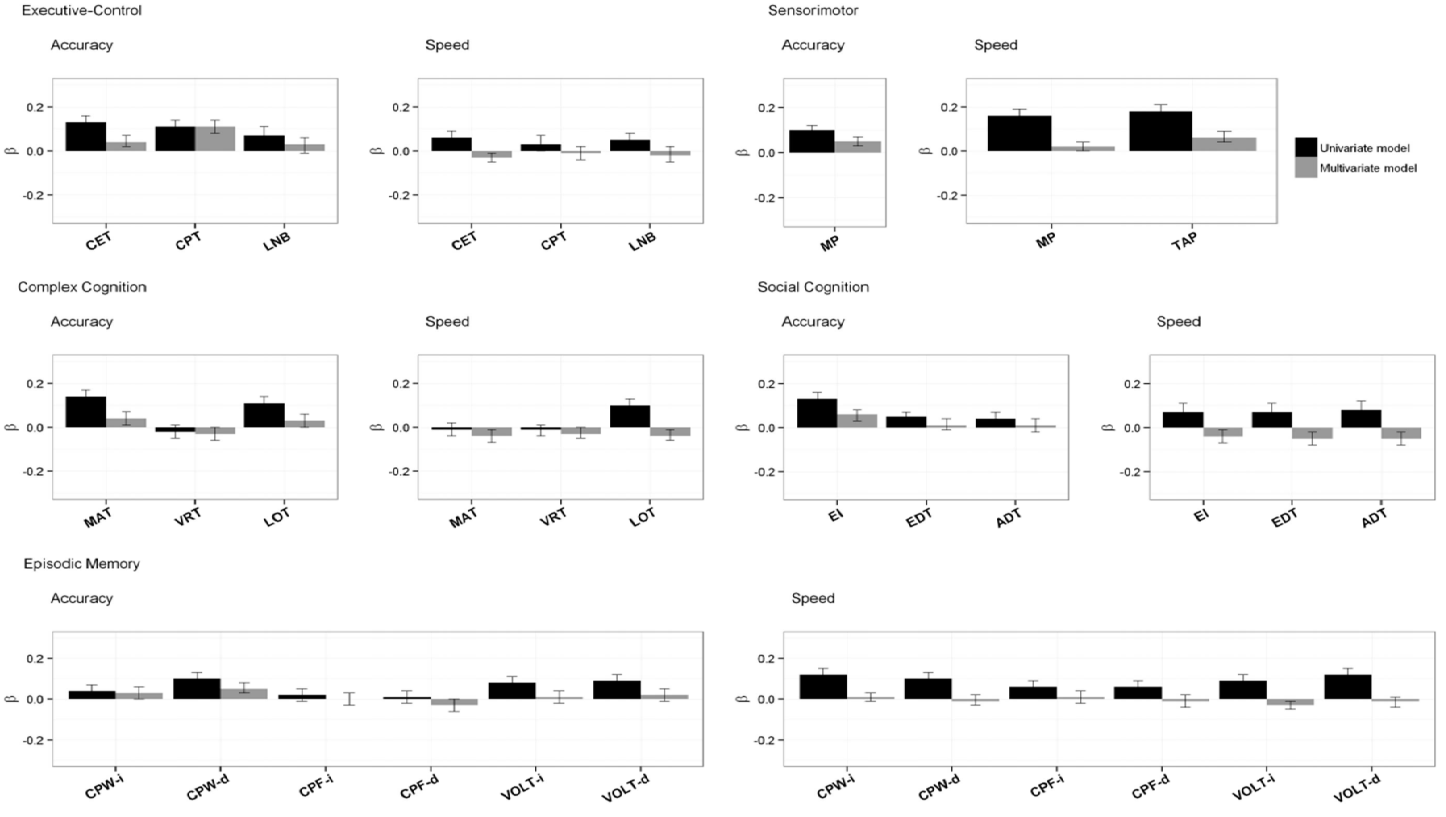
Combining previous and present results, we conclude that only the association between chronic, regular exercise behavior and attentional performance inspires some confidence. Accuracy on the CPT, a widely used neuropsychological test that measures a person's sustained and selective attention, showed the strongest association with voluntary exercise behavior. Multiple clinical studies that explored exercise as a possible treatment option for children with attention deficit / hyperactivity disorder (ADHD) have provided support for a beneficial exercise effect on the ability to focus on relevant stimuli and ignore competing stimuli (Berwid & Halperin, 2012; Pontifex, Saliba, Raine, Picchietti, & Hillman, 2013; Wigal, Emmerson, Gehricke, & Galassetti, 2013). High intensity physical activity in ADHD children may improve their continuous performance test score, for example, irrespective of the effect of the often prescribed drug methylphenidate (Medina et al., 2010). Such clinical studies demonstrate the importance of acknowledging that our results concerning voluntary exercise should not be taken as precluding beneficial effects of exercise on cognition in specific settings. As mentioned in the introduction, study design has been found to be a major source of heterogeneity among previous results as reported in the existent literature (Singh et al., 2012). Experimental studies in which effects of exercise on cognition can be attributed to intervention or treatment have shown larger associations than observational studies in which exercise-related differences between participants may be drowned out by the many other sources of individual differences in cognitive ability, including genetic factors. In part, this may reflect non-specific effects of the participation in an exercise regime; Barnes et al., (2013) found that in a sample of participants with nonclinical cognitive complaints, each of four groups (control and intervention conditions of mental and physical activity) showed increased global cognitive function. Relatively large associations obtained in experimental studies may also be due to the fact that interventions were performed in

vulnerable populations, where exercise may truly have relatively large effects. Elderly with cognitive complains or stroke have shown to benefit substantially from exercise (Barnes et al., 2013; Marzolini, Oh, McIlroy, & Brooks, 2012). Here exercise may protect against brain atrophy, increase brain connectivity, or protect against white matter damage caused by heavy alcohol consumption (Karoly et al., 2013).

Despite inconsistent findings in humans, the effectiveness of exercise has been shown more consistently in animal studies, which have suggested insight into the mechanisms involved in the beneficiary effects of exercise (Lista & Sorrentino, 2010). How these processes translate to human cognition has mainly been discussed in the light of cognitive aging: various plausible pathways have been hypothesized to explain the effects of exercise on cognitive functioning and aging processes. Exercise effects may act through a diverse set of (supra)molecular mechanisms such as angiogenesis due to increased blood flow, neurogenesis and synaptogenesis (both consistently shown in the hippocampus, involved in learning and memory). These mechanisms are controlled by processes that have also been directly associated with exercise: through for example brain derived neurotrophic factor (BNDF), growth factors, neurotransmitters (including glutamate, serotonin, noradrenaline, dopamine and acetylcholine), hormones and second messenger systems (Fabel & Kempermann, 2008; Lista & Sorrentino, 2010; van Praag, 2008). In addition, neuroimaging studies in humans have shown that exercise may induce structural changes in the hippocampus and the frontal and parietal cortex (Erickson et al., 2009; Erickson et al., 2010), as well as functional changes (Colcombe et al., 2004; Voss et al., 2010).

We end by stressing that the question of immediate importance to policy makers should not be the question whether there are associations between exercise behavior and cognitive functioning, but rather whether and how changes in exercise behavior relate to changes in cognitive functioning. Furthering the knowledge about the sources of heterogeneity in the results may be viewed as a first step. In view of the present findings, we suggest that further exploration of the association between changes in voluntary regular leisure time exercise behavior and changes in cognitive functioning is needed. Such explorations should be carried out in homogeneous samples using valid and reliable measures of exercise behavior or other forms of physical activity and neurocognitive functioning. Apart from the use of valid, reliable instruments to measure physical activity, we advance the use of strong research designs – e.g., experimental, longitudinal, or genetically informative (e. g., de Moor et al., 2008) designs – because these are essential to address the crucial question whether physical activity truly is a causal means to improve cognition.

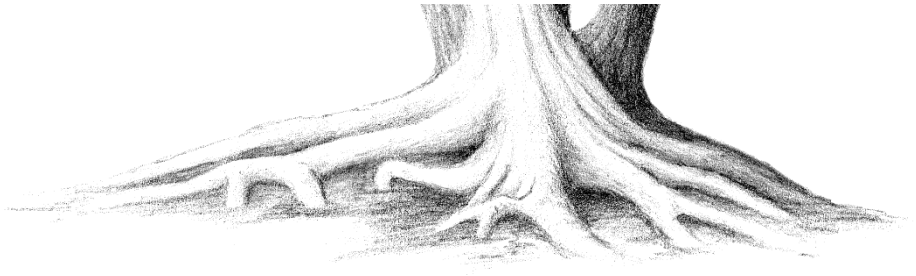
Figure 4. Standardized associations (*β*) between physical activity (in weekly METhours) and cognitive accuracy and speed across the cognitive domains, including 95% confidence intervals. The effect sizes (ES) of weekly METhours in the univariate model (black bars) are generally positive, but vary across cognitive domains. After taking into account confounding effects of sex and age in the multivariate model, effect sizes of direct associations (grey bars) are small and centered around 0. See Table 1 for cognitive domain and full name of the cognitive tests.



Chapter 7

No evidence of causal effects of blood pressure on cognition in the population at large

This chapter is published as: Swagerman, S. C., van Bergen, E., Kan, K. J., Koenis, M. M. G., Hulshoff Pol, H. E., Boomsma, D. I., & de Geus, E. J. C. (2016). No evidence of causal effects of blood pressure on cognition in the population at large. *Twin research & Human Genetics, 19*, 17-26.



Introduction

Healthy cognitive functioning is of high value to all members of society because it is important for general wellbeing, and predicts work performance and successful career paths. To promote cognitive health, a major strategy is to reduce the impact of the 'usual suspects': physical inactivity, smoking, diabetes and high blood pressure. However, for at least one of these risk factors, sustained high blood pressure, the relationship to cognitive functioning is complex. Reviews on this relationship have yielded mixed conclusions (Novak & Hajjar, 2010; Qiu, Winblad, & Fratiglioni, 2005; Reitz & Luchsinger, 2007), in keeping with the rather inconclusive findings in the primary studies (e.g., Harrington, Saxby, McKeith, Wesnes, & Ford, 2000; Lande, Kaczorowski, Auinger, Schwartz, & Weitzman, 2003; Lyngdoh et al., 2013; Tzourio, Dufouil, Ducimetiere, & Alperovitch, 1999). We here address the heterogeneity in these results. To date, whether cognitive functioning actually benefits from the reduction of high blood pressure remains an open question.

A limitation of the vast body of the studies done so far is that they mainly examine the association between blood pressure and cognition in older samples. This is understandable in view of the relatively large number of hypertensives among the elderly (Harrington et al., 2000), and the large societal burden caused not just by neurodegenerative disease but also by normal age-related cognitive decline. In the elderly, there are sensible mechanistic explanations for a negative association between high blood pressure and cognitive functioning. A currently high blood pressure level is the best predictor of prolonged past exposure to high blood pressure. Sustained exposure to high blood pressure increases the risk of atherosclerotic changes in intracerebrovascular vessels that result in a reduction of vessel number and diameter, leading to reduced vascular reserve and perfusion of the brain (Novak & Hajjar, 2010; Pires, Ramos, Matin, & Dorrance, 2013). On the neurological level this results in brain atrophy, white matter lesions, and possibly enhances Alzheimer-like pathology (Schmidt et al., 1995; Skoog, 1998; van Vliet, Westendorp, van Heemst, de Craen, & Oleksik, 2010; Waldstein, 2003).

In spite of the face validity of high blood pressure being detrimental for cognitive functioning, the actual association between blood pressure and cognitive functioning in the elderly is more complex. First, high blood pressure may be beneficial, possibly to compensate for loss of neural cell functioning (Anson & Paran, 2005; Novak & Hajjar, 2010). Second, in older people blood pressure may be lowered as a result of neurological disorders and brain lesions, which may have resulted from high blood pressure in the first place (van Vliet et al., 2010). Third, apart from high blood pressure, low blood pressure has also

been shown to have disadvantageous effects on cognitive functioning in elderly subjects (Hu, Li, Colditz, Willett, & Manson, 2003; Owen, Healy, Matthews, & Dunstan, 2010). Low blood pressure may compromise autoregulatory processes such that elderly are no longer capable of maintaining sufficient blood flow to neuronal tissue. This reduced perfusion causes reduced cortical activation, both during tonic states as well as when cognitive performance demands extra activation (Duschek & Schandry, 2007). Taken together, this indicates that the association between blood pressure and cognition is more complex, possibly because the relation is non-linear across the life span.

Despite the evidence for associations of blood pressure and atherosclerotic processes with cognition, generalization of these results is difficult, because most studies mainly involved elderly subjects. Much less is known about the association between blood pressure and cognitive functioning in samples that are representative of the population at large, which should include adolescent, young adult and middle-aged participants as well. Moreover, whereas general cognitive impairment (usually measured by the Mini-Mental State Examination, also known as MMSE) and memory loss have been the major focus of these studies in the elderly, other cognitive domains are relevant as well. Hypertension is related to brain function through reduced blood flow and metabolism, particularly in frontal, temporal and subcortical parts of the brain (Waldstein, 2003), suggesting possible different sensitivity of brain areas. This may explain why the majority of findings on specific cognitive domains find effects of blood pressure on functions like memory, attention, and executive functions (Duschek & Schandry, 2007; Gifford et al., 2013; Mitchell et al., 2011; Waldstein, 2003), but that learning, verbal, spatial and emotional abilities may be less affected. This means that different cognitive functions that depend on separate areas of the brain can be differentially susceptible to high BP, stressing the need to verify effects across cognitive domains (Birns, Morris, Donaldson, & Kalra, 2006).

A final complication to take into account when studying the association between cognition and blood pressure is that both may be influenced by one or more 'third' factors, which could either have detrimental or beneficial effects on both, or have a detrimental effect on e.g., blood pressure but a beneficial effect on cognition. The latter could lead to the absence of a clear association even in the presence of true causal effects of for example blood pressure on cognition. A clear example of the former 'third' factor influencing both cognition and blood pressure in a detrimental way is age (Deary et al., 2009; Franklin, 1999). Hence, age effects need to be controlled for rigorously. An environmental third factor that could detrimentally impact on both cognitive functioning and blood pressure is parental SES.

As compared to individuals from high SES backgrounds, individuals from low SES backgrounds tend to more often exhibit lifestyles (e.g., smoking) that may influence both cognitive functioning and blood pressure negatively (Adler & Ostrove, 1999). Third factors may also be of genetic origin: pleiotropic genetic variation may impact on common neurobiological pathways that influence both blood pressure and cognitive functioning (Herrera, Pasion, Tan, Moran, & Ruiz-Opazo, 2013; Obisesan, 2009). This might reflect shared common neurobiological pathways, or may indicate there are independent effects on both blood pressure and cognitive functioning (Waldstein, 2003). A design that is optimally suited to address the influence of third factors is a monozygotic (MZ) twin pairs difference design. MZ twins have the same age and sex, share (nearly) 100% of their segregating genes and a large part of their childhood environment. If the MZ twin with a higher blood pressure than the co-twin also has reduced cognitive functioning, this would argue in favor of a detrimental effect of high blood pressure that is causal, as this design controls for the impact of age, shared environment, and genotype (e.g., de Moor et al., 2008).

In the present study, we systematically investigate the association between current levels of systolic and diastolic blood pressure on age-dependent normal cognitive functioning while taking into account the sources that may have led to mixed results across previous studies. We allow the association between blood pressure and cognitive functioning to be non-linear, such that both low and high blood pressure constitute a risk. We further allow the sensitivity to the effects of blood pressure to depend on the cognitive domains. Thirdly, we test the relation between blood pressure and cognition in a design that controls maximally for 'third factors', of genetic and / or environmental origin. To this end we assessed blood pressure and cognitive performance in participants from a large population-based twin-family sample ranging from children to elderly individuals across a wide range of well-defined neurocognitive domains. In a subsample of 123 monozygotic twin pairs, we ordered the twins within pairs according to blood pressure levels and tested if the twin with the higher BP differed from the co-twin on any of the cognitive domains. Because MZ twins share both their genomic information and part of their environment, this analysis controls for genetic pleiotropy and the influence of shared environmental factors, for example (parental) socio-economic status, on both traits.

Materials and Methods

Participants

In total there were 1140 participants, mainly (n = 1110) recruited from 431 families from the Netherlands Twin Register (NTR, Boomsma et al., 2006; van

Beijsterveldt et al., 2013; Willemsen et al., 2013). The other 30 participants were university students. Age of the participants ranged from 10 to 86 years old (668 female, 472 male).

All participants took part in one of two studies in which measurements of blood pressure and cognitive performance were collected, referred to as the twin-family study and the twin-sibling study. The twin-family study sample consisted of 864 family members from 341 families. The majority of these participants were part of a twin pair (438) or siblings of twins (78). The rest were family of these twins and siblings, including parents (160 mothers, 126 fathers), children (9) and partners (53) of twins and siblings. In addition to this family sample, 30 university students participated in the pilot phase and were tested according to the twin-family protocol. The sample of the twin-sibling study consisted of 176 adolescent twins and 70 siblings from 89 families who participated in the third wave of assessment in a longitudinal study on development of brain and cognition (BrainScale study, van Soelen et al., 2012a).

Procedure

Studies and procedures of the twin-family study were approved by the Medical Ethics Review Committee of the Vrije Universiteit Medical Center Amsterdam, and for the twin-sibling study were approved by the Central Committee on Research Involving Human Subjects. First, participants were approached by mail. In a structured phone call interview participants were asked if they were willing to participate, if necessary additional information about the study was provided, and possible exclusion criteria were explored. After establishing an appointment, confirmation letters were sent, which included written informed consent forms and questionnaires. Consent forms were signed and returned at the appointment, after participants were given a full explanation of the procedure. Medication use and demographic data like education were registered, and a reading test was administered.

In the twin-family study, testing took place at the Vrije Universiteit Amsterdam laboratory (n = 358) or at the participants' home (n = 536). A baseline measurement of blood pressure was taken for all participants during the first (n = 885) and second (n = 883) designated break point of the cognitive test battery, and for the more recent and largest part of the participants an additional measurement was taken prior to the start of cognitive testing (n = 764).

In the twin-sibling study, testing took place in the University Medical Center Utrecht. As part of an extensive study which included MRI scans of the brain, the same cognitive test protocol was assessed as in the twin-family study, but only a single measurement of blood pressure was taken.

The administration procedure of the Computerized Neuropsychological Battery was kept similar across studies and locations. The battery was installed on Macbooks to enable offline administration during home visits. The test administrator would read the instructions to the participant (who could read them along on the screen), and monitored whether the participant understood the instructions and practice trials. Test scores were later uploaded to the Pennsylvania servers for data storage and automated validation. Breakpoints were at specific moments, for example after cognitively demanding tests (also shown in Table 1). Test administration was not interrupted at other moments.

Materials

Blood pressure

Blood pressure (BP) was measured from the brachial artery in the arm by a digital blood pressure monitor (Omron). Participants were in a sitting position. Both systolic (SBP) and diastolic (DBP) BP was recorded. Depending on the study sample and study phase (see procedure), BP was measured one, two, or three times. The intercorrelations between DBP measurements ranged from 0.78 to 0.85 and between SBP measurements from 0.87 to 0.91, indicating the reliability of both BP measures.

Cognitive performance

Cognitive performance was assessed through the Dutch translation of the web-based Computerized Neuropsychological Battery (CNB, Gur et al., 2010). In short, the CNB includes a total of 17 cognitive tests, generating 33 measures of cognitive of performance in five broad domains (Table 1): 16 measures of accuracy (the number or percentage of correct responses on the cognitive tests), 16 measures of reaction time (median response time in milliseconds for those correct responses), and one measure of motor speed (number of taps on the finger tapping test, TAP). Reaction times were multiplied by -1 so that high cognitive performance was always denoted by a high test score. We refer to Swagerman et al., (2015a) for a more extensive description of the battery and full descriptives for the performance data.

Statistical analyses

Invalid scores (~0.9%) and data from children under age 13 (n = 4) were excluded from the present analysis, which comprised two sets of statistical analyses: the first set involved data from the complete sample, the other the data from monozygotic twin pairs only. In both sets of analyses, DBP values were averaged over the (one, two or three) measurements. The same held for SBP.

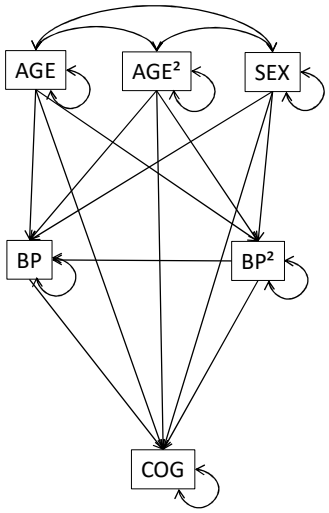
Table 1. Tests and their cognitive measure and domain, given in order of assessment.

Test name	Label	Cognitive measure	Cognitive domain
Motor Praxis Test	MP	sensorimotor speed	sensorimotor
Penn Emotion Identification Test	EI	emotion identification	social cognition
Penn Continuous Performance Test	CPT	attention	executive control
Breakpoint 1			
Penn Facial Memory Test	CPFi	face memory	memory
Penn Word Memory Test	CPWi	verbal memory	memory
Letter-N-Back Test	LNB	working memory	executive control
Penn Facial Memory Test – delayed	CPFd	face memory	memory
Penn Word Memory Test – delayed	CPWd	verbal memory	memory
Breakpoint 2			
Penn Conditional Exclusion Test	CET	abstraction / flexibility	executive control
Measured Emotion Differentiation Test	EDT	emotion differentiation	social cognition
Computerized Finger-Tapping Test	TAP	motor speed	sensorimotor
Penn Matrix Reasoning Test	MAT	nonverbal reasoning	complex cognition
Breakpoint 3			
Visual Object Learning Test	VOLTi	spatial memory	memory
Penn Logical Reasoning Test	VRT	language reasoning	complex cognition
Age Differentiation Test	ADT	age differentiation	social cognition
Variable Penn Line Orientation Test	LOT	spatial ability	complex cognition
Visual Object Learning Test – delayed	VOLTd	spatial memory	memory

Whole sample analyses

For each of the 33 cognitive measures, the association of cognitive performance with BP was estimated by fitting the path model that is diagrammed in Figure 1. That is, each cognitive performance measure was regressed on DBP or SBP and its mean centered square (DBP^2 or SBP^2), while regressing out the possible confounding effects of sex, age, and the (mean centered) square of age (age^2). The coefficients of the direct paths from the BP variables to the cognitive measure were the parameters of interest. Hence, the entire set of analyses yielded 2 (linear and quadratic)*2 (systolic and diastolic)*33 (cognitive variables) = 132 path coefficient estimates, so 132 test statistics. In view of chance capitalization, while acknowledging Bonferroni correction for multiple testing may be too conservative, we corrected the original significance level (0.05) via Matrix Spectral Decomposition (Li & Ji, 2005), which indicated that the data could be described along 23 statistically independent dimensions. This yielded a corrected significance level of: $\alpha = 0.05/23 \approx 0.002$ (see also Swagerman et al., 2015a).

Figure 1. Graphical representation of the path model fitted to the data of the whole sample, showing that cognitive test performance (COG) depends on blood pressure (BP) and (mean centered) blood pressure squared (BP^2), while regressing out the effects of sex, age, and the mean centered square of age (AGE^2).



The path modeling was performed in R using packages lavaan (Rosseel, 2012) and lavaan.survey (Oberski, 2014). We opted for robust sandwich estimation (in view of the data being skewed) and included family number as a cluster variable (in view of the data being family clustered). To estimate the sensitivity of the effects to antihypertensive medication, all analyses were rerun after excluding all individuals who take any type of antihypertensive medication (60 females, 58 males, mean age 69.08).

Monozygotic twin pairs difference analyses

To control for possible confounding effects other than those of age and sex, including shared environmental and genetic factors, additional analyses were performed using the data from the monozygotic (MZ) twins only. Within twin pairs, the MZ twins were ranked on the basis of their blood pressure, such that two groups were formed: a high BP group, consisting of the twins with the relatively high BP, and a low BP group, consisting of the twins with the relatively low BP. Next, we investigated the mean difference in cognitive performance between those groups.

The rationale behind such analyses is as follows. As MZ twin pairs are not only of the same age and sex, but also matched on their genotype (since they are 100% genetically identical) and many environmental variables, mean group differences in cognitive performance between the MZ twins cannot be explained

by the variables on which they are matched. Any mean group differences in cognitive performance between the groups can thus be attributed to the between pair differences in blood pressure (or nonshared causes thereof). If high blood pressure has a negative effect on cognition, the high BP group is expected to display significantly lower cognitive performance scores (on average) as compared to the low BP group; if BP has a positive effect, the high BP group is expected to score higher than the low BP group. If, despite differences in blood pressure, mean differences between cognitive scores are insignificant, one is allowed to conclude that BP does not influence cognitive performance. The ranking, grouping, and subsequent analysis of mean scores was done for SBP and DBP separately.

Two-sided paired t-tests were performed in SPSS version 21 (IBM Corp., 2011) separately for DBP and SBP on the 33 cognitive performance scores. In line with the whole sample analysis, we used a corrected significance level of 0.002 to account for the multiple testing in a set of correlated dependent variables. The 287 MZ twins in total formed 132 pairs of whom data were complete for both twins. Ranks of mean SBP and DBP were consistent in 96 of these pairs. If both twins had exactly equal mean BP values (3 twin pairs had equal mean DBP, 1 pair equal mean SBP), twins were assigned to the high and low BP groups randomly. If one or both twins of a pair used antihypertensive medication, both twins were excluded from the analyses (five pairs concordant, four pairs discordant for medication use). Ultimately, in total the analyses included 123 complete twin pairs.

Results

Descriptive statistics

The mean DBP was 75.49 (SD = 10.98) and the mean SBP 129.51 (SD = 17.02). Mean sex differences in DBP (males: M = 75.83, SD = 11.08; females: M = 75.25, SD = 10.91) were not significant (β = -0.04, se = 0.03 , p = 0.23). Mean sex differences in SBP (males: M = 134.87, SD = 15.87; females: M = 125.76, SD = 16.80) were however significant (β = 0.25, se = 0.03 , p < .001). Both DBP and SBP related significantly with age (β = 0.49, se = 0.03, p < .001; β = 0.48, se = 0.03, p < .001, respectively).

Whole sample analyses

The standardized results of the path modeling, i.e., the direct path coefficients of both the linear and quadratic terms of SBP and DBP on cognitive performance and their test statistics, are provided in Table 2 and summarized graphically in Figure 2. These coefficients can be interpreted as age and sex corrected effects of current blood pressure on current cognitive performance.

Table 2. The age and sex corrected linear and quadratic effects (β; direct path coefficients in Figure 1) of diastolic and systolic blood pressure on cognitive accuracy and speed performance including their standard errors (SE) and p-values (p).

	Diastolic blood pressure						Systolic blood pressure					
	Linear			Quadratic			Linear			Quadratic		
Domain (accuracy)	β	*SE*	p	β	*SE*	p	β	*SE*	p	β	*SE*	p
Executive control												
Abstraction & flexibility	-.005	.032	.876	-.020	.036	.568	.043	.034	.210	-.004	.034	.915
Attention	.030	.031	.325	-.011	.027	.685	.053	.035	.128	-.049	.033	.147
Working Memory	.012	.030	.697	.013	.029	.657	.050	.035	.157	-.016	.035	.651
Memory												
Verbal Memory	.027	.034	.421	-.072	.038	.058	.002	.038	.958	-.012	.035	.737
delayed	.037	.032	.251	-.057	.029	.049*	-.005	.034	.885	-.044	.032	.170
Facial Memory	-.029	.036	.423	-.039	.032	.215	-.018	.036	.616	-.003	.031	.916
delayed	.002	.033	.960	-.029	.030	.336	.010	.036	.786	.017	.035	.633
Spatial Memory	-.053	.032	.092	-.006	.028	.819	-.061	.034	.071	.015	.031	.630
delayed	-.061	.033	.067	-.018	.034	.601	-.020	.036	.568	-.015	.035	.672
Complex Cognition												
Nonverbal Reasoning	-.056	.029	.053	-.021	.024	.377	-.027	.033	.421	-.005	.029	.878
Language Reasoning	-.003	.034	.928	-.034	.029	.240	-.035	.037	.334	.026	.034	.445
Spatial Ability	-.053	.034	.118	.031	.033	.339	.007	.036	.850	.003	.032	.922
Social Cognition												
Emotion Identification	.041	.035	.234	-.040	.031	.186	.060	.035	.085	.001	.029	.981
Emotion Differentiation	.005	.033	.890	.002	.028	.953	.056	.036	.121	-.043	.035	.215
Age differentiation	.053	.032	.100	-.064	.035	.070	.092	.036	.011*	-.073	.036	.043*
Sensorimotor												
Sensorimotor Speed	-.001	.035	.988	-.003	.027	.915	-.030	.033	.359	.032	.030	.290

* = significant at α = .05.

Table 2 - continued.

	Diastolic blood pressure						Systolic blood pressure					
	Linear			Quadratic			Linear			Quadratic		
Domain (speed)	*β*	*SE*	*p*	*β*	*SE*	*p*	*β*	*SE*	*p*	*β*	*SE*	*p*
Executive control												
Abstraction & flexibility	-.060	.030	.048*	.003	.034	.923	-.059	.03	.052	.001	.038	.979
Attention	.023	.034	.501	.013	.027	.634	-.020	.037	.591	.061	.036	.087
Working Memory	.055	.033	.095	-.026	.035	.448	-.008	.038	.822	.037	.037	.320
Memory												
Verbal Memory	.065	.031	.034*	-.037	.025	.136	.061	.032	.061	.001	.025	.967
delayed	.033	.030	.264	-.021	.024	.380	.020	.031	.524	.025	.025	.315
Facial Memory	.071	.036	.045*	-.02	.030	.495	.044	.035	.209	.007	.026	.798
delayed	.048	.034	.162	.003	.031	.915	.017	.035	.628	.031	.027	.238
Spatial Memory	.006	.033	.849	-.006	.030	.849	.013	.033	.693	.020	.026	.449
delayed	.013	.029	.659	.008	.028	.786	-.008	.029	.782	.026	.026	.326
Complex Cognition												
Nonverbal Reasoning	.078	.031	.013*	.034	.026	.189	.049	.038	.194	-.001	.034	.986
Language Reasoning	.107	.035	.002*	.005	.023	.818	.062	.037	.093	.003	.029	.915
Spatial Ability	.033	.028	.242	-.005	.024	.819	-.043	.030	.151	.038	.027	.150
Social Cognition												
Emotion Identification	.114	.037	.002*	-.025	.033	.440	.070	.035	.045*	.033	.022	.145
Emotion Differentiation	.034	.027	.213	.018	.020	.358	-.018	.028	.527	.043	.022	.056
Age differentiation	.030	.027	.267	-.005	.020	.809	-.004	.028	.893	.012	.025	.640
Sensorimotor												
Sensorimotor Speed	.036	.035	.298	-.007	.032	.815	.033	.034	.321	.043	.025	.076
Motor Speed	.016	.027	.560	-.016	.020	.425	.014	.028	.624	.056	.022	.013*

* = significant at α = .05.

Table 3. Showing the mean cognitive performance scores of the high and low diastolic and systolic monozygotic twin groups, the mean differences (Δ) with their standard deviations, and the results of the paired t-tests (t-statistic, t; and p-value, p).

	Diastolic blood pressure						Systolic blood pressure					
Variable	Mean high	Mean low	Δ	Δ (SD)	*t*	*p*	Mean high	Mean low	Δ	Δ (SD)	*t*	*p*
Blood Pressure	77.2	69.0	7.9	6.2	14.141	.000**	129.0	119.2	9.8	8.4	12.995	.000**
Domain (accuracy)												
Executive Control												
Abstraction & flexibility	1.9	2.0	-0.2	0.9	-1.864	.065	1.8	2.0	-0.2	0.9	-2.265	.025*
Attention	55.8	56.3	-0.5	4.2	-1.395	.165	56.2	56.0	0.2	4.2	.576	.566
Working Memory	19.0	19.2	-0.2	1.4	-1.323	.188	19.1	19.1	-0.2	1.4	-1.189	.237
Memory												
Verbal Memory	1.86	2.01	-0.2	0.9	1.376	.171	36.8	36.4	0.4	2.9	1.626	.107
delayed	55.8	56.3	-0.5	4.2	1.122	.264	35.8	35.4	0.3	3.5	1.017	.311
Facial Memory	19.0	19.2	-0.2	1.4	-1.974	.051	31.8	32.0	-0.2	3.7	-.714	.476
delayed	1.86	2.01	-0.2	0.9	-.760	.449	33.0	33.1	-0.1	3.6	-.253	.801
Spatial Memory	55.8	56.3	-0.5	4.2	-1.082	.281	16.1	16.3	-0.2	2.4	-.857	.393
delayed	19.0	19.2	-0.2	1.4	-1.029	.306	15.4	15.7	-0.3	2.6	-1.100	.273
Complex Cognition												
Nonverbal Reasoning	15.1	14.9	0.1	4.3	.233	.817	15.1	14.9	0.2	4.3	.444	.658
Language Reasoning	69.2	65.2	3.6	21.4	1.850	.067	69.3	65.1	3.8	21.4	1.959	.052
Spatial Ability	13.2	13.5	-0.3	3.5	-.978	.330	13.2	13.5	-0.4	3.5	-1.186	.238
Social Cognition												
Emotion Identification	32.8	33.3	-0.4	3.3	-1.423	.157	32.9	33.2	-0.3	3.3	-.871	.385
Emotion Differentiation	28.5	28.4	0.2	4.2	.425	.671	28.4	28.5	-0.1	4.2	-.128	.899
Age differentiation	27.6	27.4	0.2	4.7	.478	.633	27.5	27.5	0.0	4.7	-.057	.954
Sensorimotor												
Sensorimotor Speed	20.0	20.0	0.0	0.2	1.000	.319	20.0	20.0	0.0	0.2	1.000	.319

* = significant at α = .05, ** = significant at α = .01.

Table 3 - continued.

	Diastolic blood pressure						Systolic blood pressure					
Domain (speed)	Mean high	Mean low	Δ	Δ (SD)	*t*	*p*	Mean high	Mean low	Δ	Δ (SD)	*t*	*p*
Executive Control												
Abstraction & flexibility	-2547.3	-2472.5	-74.8	1081.7	-.764	.447	-2466.7	-2553.1	86.4	108.8	.883	.379
Attention	-483.7	-483.2	-0.5	52.2	-.106	.916	-485.2	-481.8	-3.4	52.1	-.714	.476
Working Memory	-529.1	-526.7	-3.7	122.3	-.334	.739	-529.1	-526.7	-3.3	122.3	-.296	.768
Memory												
Verbal Memory	-1478.0	-1453.9	-24.1	237.4	-1.124	.263	-1467.3	-1464.6	-2.8	238.6	-.128	.898
delayed	-1457.0	-1441.3	-15.8	281.3	-.622	.535	-1434.8	-1463.5	28.7	28.2	1.137	.258
Facial Memory	-1957.5	-1864.8	-9.3	498.7	-1.976	.050	-196.2	-1862.9	-1.1	496.8	-2.199	.030*
delayed	-1765.8	-1715.6	-47.7	44.3	-1.183	.239	-1747.1	-1734.6	-11.0	442.8	-.271	.787
Spatial Memory	-184.1	-1862.8	18.2	5.2	.402	.688	-1856.4	-1846.7	-14.3	5.4	-.315	.753
delayed	-1681.5	-1697.3	12.3	426.9	.317	.752	-1701.0	-1678.0	-31.1	425.9	-.803	.424
Complex Cognition												
Nonverbal Reasoning	-1026.5	-11038.6	815.5	6553.9	1.374	.172	-10629.1	-10672.9	78.1	6604.3	.131	.896
Language Reasoning	-9052.8	-8903.5	-136.6	3931.5	-.384	.702	-8955.0	-90.5	59.1	3933.4	.166	.869
Spatial Ability	-9459.6	-997.4	549.5	3126.1	1.933	.056	-9607.3	-9822.8	251.8	3164.3	.875	.383
Social Cognition												
Emotion Identification	-2056.3	-2116.7	6.4	492.1	1.361	.176	-2082.1	-209.9	8.8	495.7	.197	.844
Emotion Differentiation	-3409.1	-3314.1	-95.1	1085.2	-.972	.333	-3415.0	-3308.2	-106.8	1084.1	-1.093	.277
Age differentiation	-2838.5	-2769.3	-69.3	1052.4	-.730	.467	-2815.2	-2792.6	-22.7	1054.5	-.239	.812
Sensorimotor												
Sensorimotor Speed	-712.6	-712.8	0.2	132.8	.019	.985	-722.1	-703.3	-18.7	131.5	-1.580	.117
Motor Speed	114.9	113.2	1.9	12.9	1.574	.118	113.7	114.4	-.6	13.0	-.521	.603

* = significant at α = .05, ** = significant at α = .01.

Accuracy

The effects of BP on accuracy were small and distributed around 0. Linear effects of DBP ranged from -0.061 (spatial memory delayed test) to 0.053 (age differentiation test) with a median of 0.002, and those of SBP ranged from -0.061 (spatial memory test) to 0.092 (age differentiation test) with a median of 0.007. Quadratic effects of DBP ranged from -0.072 (verbal memory test) to 0.031 (spatial ability test) with a median of -0.02 and those of SBP ranged from -0.073 (age differentiation test) to 0.032 (sensorimotor test) with a median of -0.004. None of these effects reached the level of significance ($\alpha = 0.002$). Medication status did not influence the results: nearly identical effect sizes were obtained when the analyses were repeated after removing the 118 individuals taking antihypertensive medication and are therefore not reported.

Speed

Like the effects of BP on accuracy, the effects of BP on speed were small. Linear effects of DBP ranged from -0.060 (abstraction and mental flexibility) to 0.114 (emotion identification test) and were mostly positive with a median of 0.036. Yet of those effects, only the linear effects on language reasoning and emotion identification were significant ($\alpha = 0.002$). This was also true in the repeated analyses after removing individuals taking antihypertensive medication.

Linear effects of SBP were centered around 0 again and ranged from -0.059 (abstraction and mental flexibility) to 0.070 (emotion identification test) with a median of 0.017. None of these coefficients were significant.

Quadratic effects of DBP ranged from -0.037 (verbal memory test) to 0.034 (nonverbal reasoning test) with a median of -0.005, and those of SBP ranged from -0.001 (nonverbal reasoning) to 0.061 (attention) with a median of 0.026. These were insignificant as well.

In conclusion, no systematic linear or quadratic association between blood pressure and cognitive functioning was found. There is some evidence of a linear effect of diastolic blood pressure on the speed of language reasoning and emotion identification, but the DBP effects on cognitive speed were not replicated in other tasks that measure performance in the same domain.

Monozygotic twin pairs difference analyses

The results of the additional MZ twin pair analysis are provided in Table 3. The mean BP of the high DBP and SBP groups were significantly higher than mean BP in the group with their co-twins with low DBP and SBP. The mean differences in cognitive performance were small and for about half of all test scores, the differences were in the positive direction and for the other half in the negative

direction. None of these intrapair MZ differences in cognitive performance remained significant after correction for multiple testing.

From the combined sets of analyses, we may derive that in the population at large associations between blood pressure and cognitive performance are absent when effects of confounding factors like age, sex, genetic pleiotropic and shared environmental factors are taken into account. We conclude that our analyses show no evidence for a causal relationship between (current) level of blood pressure on (current) cognitive functioning.

Discussion

We set out to explore the effects of diastolic and systolic blood pressure on cognitive functioning in a population-based sample with a wide age range. We attempted to explain the inconsistencies in studies thus far by allowing blood pressure effects to be curvilinear and to be different across accuracy and speed measures of five cognitive domains. We accounted for confounding of any association by genetic factors and by many shared environmental factors including parental SES. Our whole sample analyses showed that for the majority of the cognitive tests very little indication was found that, after correction for substantial age-effects, blood pressure was associated with cognitive test performance. The MZ twin pair analyses also did not suggest any effect of DBP or SBP on cognitive performance.

When we look at the analyses in the whole sample in a liberal way, i.e. without any correction for multiple testing, two instances of a possible adverse effect of BP on cognition were found, suggesting a detrimental effect of higher BP on the speed of emotion identification and verbal reasoning. The association with emotion identification is an isolated one as no other studies that we could identify have been conducted on blood pressure and tests of social cognition. Previous findings for blood pressure and language skills have been far from consistent but a meta-analysis of 12 studies found a possible trend that was in keeping with our result (Gifford et al., 2013). In view of the large number of cognitive domains tested, these two nominally significant associations most likely reflect false positives. They were not seen for other tasks within the relevant cognitive domains, and none were replicated for SBP.

Our findings were based on a large dataset of participants spanning a large age range, drawn from the general population. This differs from the bulk of the existing literature that focused on either children but mostly on the elderly. Studies in children are limited by the relatively few children showing hypertension, which is then often complicated in those children by comorbid disorders and congenital disease (Cha, Patel, Hains, & Mahan, 2012; Lande et

al., 2003; Lyngdoh et al., 2013). Studies in the elderly are complicated by comorbidity of high blood pressure with other atherosclerotic risk factors like cholesterol, diabetes, BMI, immune parameters, and medical or psychiatric disorders which can all assert an influence on cognitive functioning (Spauwen et al., 2015; van Vliet et al., 2010; Waldstein, 2003; Walther, Birdsill, Glisky, & Ryan, 2010). These factors can also cause cardiac pathology that itself reduces blood pressure, as seen in heart failure patients, further complicating the association between blood pressure and cognition because too low blood pressure also has disadvantageous effects on cognitive functioning in elderly subjects (Hu et al., 2003; Owen et al., 2010). Moreover, the effects of neurodegeneration and dementia are hard to separate from normal cognitive aging. The Institute of Medicine recently defined cognitive aging as a lifelong process of gradual changes in cognitive function that is highly variable across individuals and within individuals, and across cognitive domains (Blazer et al., 2015). Animal models show that, in contrast to Alzheimer disease or other neurodegenerative disorders, there is no loss of neurons with normal aging, but a gradual change in synaptic structure and function. The distinction is relevant because high blood pressure can compensate the loss of neural cell functioning (Anson & Paran, 2005; Novak & Hajjar, 2010) and this may have a different impact on normal and neurodegenerative cognitive aging. In view of these complexities it is perhaps not surprising that meta-analyses of large randomized controlled trials have not uniformly concluded that antihypertensive medication improves cognitive performance in the elderly (McGuinness, Todd, Passmore, & Bullock, 2009; Novak & Hajjar, 2010).

As a large part of our study sample consisted of adolescents, or young and middle-aged adults, atherosclerotic or neurodegenerative damage will not have been a major confounder, which is a strength of the current study. Repeating analyses in age groups under 30 or over 50 showed virtually the same results (analyses not shown). However, our study also suffered some limitations. The absence of curvilinear effects of blood pressure may mean that we overly complicated our analyses. We still recommend modeling them in future studies, since strong non-linear effects might be present in specific patient groups, as for example found in a subsample of diabetes patients (Spauwen et al., 2015). The largest limitation is the use of only one to three measurements of blood pressure in a sitting condition to characterize blood pressure. This may not be sensitive to detect effects of more complex aspects of blood pressure regulation. A more extensive assessment of blood pressure, for example also taken while lying down or while standing up to test for postural hypotension (e.g., Kuo et al., 2004) might have revealed different results for cognition. Finally, our twin design capitalized

on intra-pair differences in BP in genetically identical individuals, which were modest in size due to the substantial heritability of BP (Hottenga et al., 2005).

In summary, associations between cognitive functioning and DBP and SBP were absent for the vast majority of the 33 cognitive accuracy or speed measures, most notably in a well-matched within-twin pair comparison of MZ twins. The few nominally significant associations between specific test scores and blood pressure most likely reflect chance findings. We conclude that across a wide age range in the population at large, blood pressure level is not associated with cognitive functioning in a clinically meaningful way. This does not detract from the possible value of treating blood pressure to prevent atherosclerotic and neurodegenerative disease in older age, but it also suggests that there are no immediate cognitive benefits from lowering blood pressure per se.

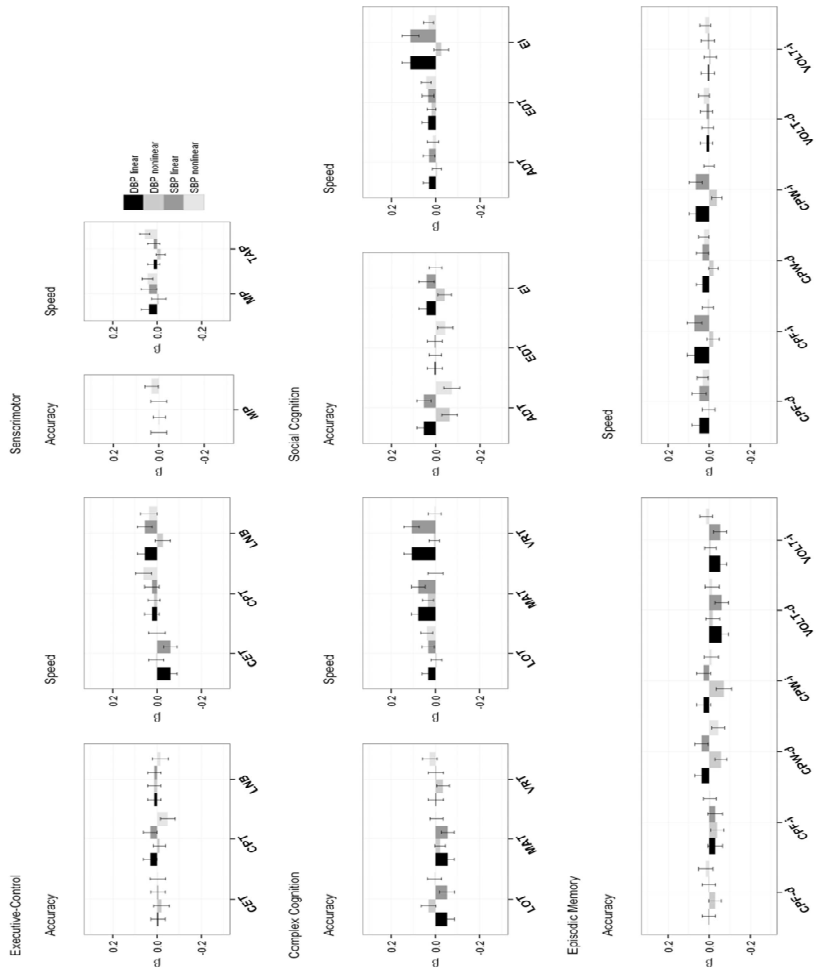
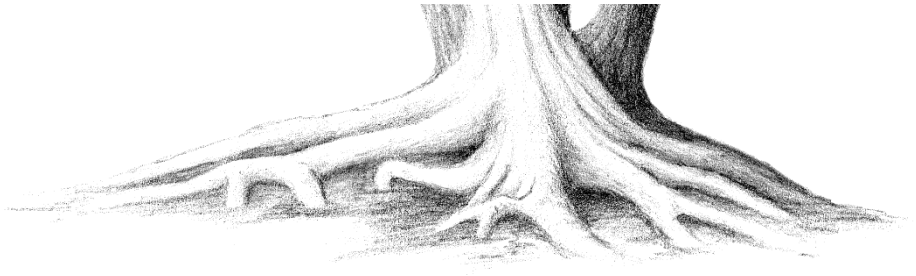


Figure 2. Standardized effect size of linear and nonlinear effects of diastolic blood pressure and systolic blood pressure on cognitive accuracy and speed, ordered by cognitive domain.

Chapter 8

Development and heritability of subcortical brain volumes at age 9 and 12

This chapter is published as: Swagerman, S. C., Brouwer, R. M., de Geus, E. J. C., Hulshoff Pol, H. E., & Boomsma, D. I. (2014). Development and heritability of subcortical brain volumes at age 9 and 12. *Genes, Brain & Behavior*, *13*, 733-742.



Introduction

The heritability of global brain volumes is well established in adults, and also from a number of studies in adolescents and young children (Batouli et al., 2014; Blokland et al., 2012; Peper et al., 2007). Global brain volumes are moderately to highly heritable from birth onwards (Gilmore et al., 2010), increasing in heritability during childhood and adolescence, possibly followed by a decrease (Batouli, Trollor, Wen, & Sachdev, 2014; Giedd et al., 2010; Lenroot & Giedd, 2008; Peper et al., 2009b; van Soelen et al., 2013; Wallace et al., 2006; Yoon, Perusse, Lee, & Evans, 2011).

Regional brain volumes, including the subcortical grey matter structures, may be more sensitive to environmental influences than global brain volumes (Draganski et al., 2004). In particular, plasticity of the hippocampus has been found to be associated with environmental influences in several studies: volume increase due to specific skills training was shown in studies of London taxi drivers (Woollett & Maguire, 2011) and exercisers (Erickson et al., 2011; Schlaffke et al., 2014), whereas stressors like an earthquake have been associated with a decrease in hippocampus volume (Lui et al., 2013). Stress was also found to affect the amygdala, nucleus accumbens, caudate and putamen, all of which have a role in emotion processing, mood regulation, learning and cognitive functions (Davidson et al., 2002; Lucassen et al., 2014; Phelps, 2004; Ring & Serra-Mestres, 2002; Shohamy, 2011).

Subcortical brain structures follow differential developmental patterns from child- to adulthood: decrease (e.g., caudate), increase (e.g., hippocampus) and inverted U shaped trajectories (e.g., thalamus) have been reported (Dennison et al., 2013; Durston et al., 2001; Goddings et al., 2014; Ostby et al., 2009; Wierenga, Langen, Oranje, & Durston, 2014). Developmental changes in total brain volume and cortical thickness have been associated with genetic and environmental factors during the early adolescent years (van Soelen et al., 2012b; van Soelen et al., 2013). However, current knowledge about the extent to which genes and environment influence changes in subcortical brain volumes is much more limited. Recent twin studies in adults and children (see for example Bohlken et al., 2013; den Braber et al., 2013a; Kremen et al., 2010; Yoon et al., 2011) and a comprehensive meta-analysis suggest that heritability for subcortical brain volumes is high. The wide confidence intervals around the point estimates stress the need for further studies (Blokland et al., 2012).

Here, the heritability of 7 subcortical brain structures (thalamus, caudate, putamen, pallidum, amygdala, hippocampus, and nucleus accumbens) is estimated at ages 9 and 12 years in a population based twin sample. The study

is characterized by a longitudinal design, which allows to test for heritability changes over this age span and to test if new genetic factors are expressed at age 12. Differences in puberty status between boys and girls will be small at age 9, but girls may be more advanced at age 12, so we will test for sex differences in heritability estimates. Because the study includes mono- and dizygotic male and female twin pairs, as well as opposite-sex pairs, we can assess both qualitative differences, i.e. test if the same genes are expressed in boys and girls, and quantitative differences, i.e. in the magnitude of genetic and environmental effects.

Methods

Participants

Twins were recruited from the Netherlands Twin Register (NTR, Boomsma et al., 2006; van Beijsterveldt et al., 2013; Willemsen et al., 2013). Twins, aged 9 years, who were born in 1995-1996 with an older brother or sister, aged 10-14 years, were invited to participate in the BrainScale study of brain and cognitive development. This is a longitudinal study in which the NTR, the Brain Center Rudolf Magnus, and the University Medical Center Utrecht collaborate. The sample was largely unselected for phenotype, but children were excluded from participation in case of a pacemaker, metal material in their head, chronic use of medication, a major medical or psychiatric history, participation in special education or physical or sensory disabilities. At the first assessment, 112 twin pairs participated (mean age 9.10, SD = 0.10), and at follow-up 89 pairs came back for the second assessment (M = 12.15, SD = 0.26). At age 9, there were 48 monozygotic (MZ) pairs (23 male / 25 female) and 64 dizygotic (DZ) twin pairs (23 male / 21 female / 20 opposite sex). For demographics see Table 1, and for more details on the sample and study design also see Van Soelen et al., (2012a).

Procedure

The Central Committee on Research involving Human Subjects approved this study. After the test administrator explained the testing procedure and the goal of the research project, both parents and children gave written informed consent. At age 9 twins came to the laboratory at the Vrije Universiteit Amsterdam for cognitive testing and to the University Medical Center Utrecht on a separate occasion for MRI scanning (preceded by a visit to the dummy scanner). At age 12 the cognitive assessment and MRI scans were completed on the same day at the University Medical Center Utrecht. Data on physical development (length, weight, and Tanner phase) were measured by a trained researcher. At age 12 children were offered the option to provide self-report data on Tanner phase (16 girls, 28 boys). Morning urine, saliva samples and cheek swabs were collected at

home on two consecutive days at fixed times and were used for assessment of estrogens, luteinizing hormone (LH), follicle stimulating hormone (FSH), testosterone, and genetic markers (for details see Koenis et al., 2013). Self- and maternal reports of health, lifestyle and behavioral and emotional problems of the children were collected by surveys. MRI scanning was performed on a 1.5-T Philips Achieva scanner on both occasions, using the same scan sequence parameters and image processing procedures (Peper et al., 2008; van Soelen et al., 2013). At both baseline and follow-up image sequences for the whole head were acquired, including a short scout scan for immediate verification of optimal head positioning, and a clinical scan that was used for neurodiagnostic evaluation. A three-dimensional T1-weighted coronal spoiled-gradient echo scan of the whole head (256 x 256 matrix, Echo Time = 4.6 ms, Repetition Time = 30 ms, flip angle = 30°, 160-180 contiguous slices; 1 x 1 x 1.2 mm3 voxels, field-of-view = 256/70%) was acquired for volumetric analysis.

Subcortical structures were segmented automatically by the publicly available Freeserver software package (version 5.1; Fischl et al., 2002; 2004). Our previously manually edited intracranial masks were inserted in this pipeline to compute subcortical structures with a high quality brain mask. Quality control was performed to check segmentation accuracy in outlying volume measurements by visual inspection of the scans for movement effects. Insufficient detail of the subcortical volumes led to excluding participants or specific structures from the analyses (see Supplementary Table S1).

Table 1. Sample characteristics.

	Age 9	Age 12
Total number of twins (girls/boys)	112/112	89/89
Number of participants with complete MRI scan	210	136
Twin pair zygosity (MZ / DZ / DOS)	48/44/20	40/34/15
Mean age of twins in years (sd)	9.2 (0.1)	12.1 (0.3)
Height (centimeter)		
Girls (MZ / DZ / DOS)	136.6/138.8/ 140.6	151.1/153.3/155.1
Boys (MZ / DZ / DOS)	139.5/138.6/140.1	153.5/150.4/151.9
Weight (kilogram)		
Girls (MZ / DZ / DOS)	30.4 / 31.8 / 32.0	43.4 / 44.6 / 41.4
Boys (MZ / DZ / DOS)	31.8 / 31.2 / 31.9	44.5 / 41.9 / 39.4
Tanner stage 1/2/3/4/5 (missing values)		
Girls: Breast development	89/20/-/-/- (3)	10/16/36/17/7 (3)
Pubic hair	91/17/-/-/- (4)	17/17/18/23/5 (9)
Boys: Penis development	100/5/1/1/- (5)	20/37/21/5/- (5)
Pubic hair	96/10/0/0/- (6)	24/31/22/6/- (6)

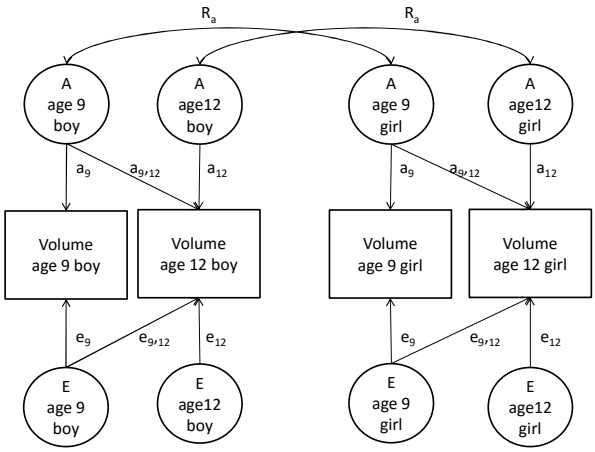
MZ = monozygotic, DZ = dizygotic same sex, DOS = dizygotic opposite sex

Analyses

To estimate heritability, the classical twin model focuses on the difference in resemblance (correlation or covariance) for a particular trait between dizygotic twin pairs (DZ) who share on average half of their segregating genes and monozygotic twins (MZ) who are (nearly) genetically identical. Comparing the cross-twin-within-trait correlations of MZ with DZ twins gives an indication of the sources of variation. Because MZ and DZ twins differ in their genetic similarity, genetic effects are suggested for a trait if the MZ cross-twin-within-trait correlation is higher than the DZ correlation. Additionally, common environmental effects are suggested to also contribute to twin resemblance when the DZ correlation is larger than half the MZ correlation. In longitudinal data, comparing the cross-twin-cross-trait correlations (i.e. brain volume at age 9 in one twin with brain volume at age 12 in the cotwin) gives an indication of the sources causing covariance between traits: the phenotypic correlation between brain volume at two ages is explained by common genetic factors when the MZ cross-twin-cross-trait correlation is larger than the DZ correlation. Longitudinal modeling of all twin data were performed in OpenMx (Boker et al., 2011) by raw-data maximum likelihood, allowing for (any pattern of) missingness in the data. Therefore we did not remove the cotwin if data of the other twin had to be excluded. Excluded participants were evenly spread between zygosity and sex groups. Bivariate analyses were run between the volume data collected at age 9 and 12, separately for the left and right volume of each subcortical brain structure. First, in a saturated model, means, variances and twin correlations, within age and across-age, were estimated for the five sex-by-zygosity groups (MZM, DZM, MZF, DZF and DZMF) and differences in mean volumes between boys and girls and between 9 and 12 years were tested for significance. Next, heritability was estimated in a series of genetic models. In the full longitudinal model, parameters representing the influence of additive genetic factors (A), common environment shared by twins (C) and unique environment (E) were estimated separately for boys and girls. The genetic correlation between opposite-sex twin pairs was estimated freely and changes in the fit of the model were compared to a model in which the correlation was equal (0.5) to the genetic correlation in same-sex DZ pairs. Quantitative sex differences were tested by constraining the influences of A, C and E to be equal for boys and girls. Next, significance of the common and genetic environmental factors was assessed by constraining their influence at zero. Last, the significance of new genetic effects coming into play at age 12 was tested. Figure 1 presents the longitudinal model for 2 twins whose brain volumes were assessed at ages 9 and 12 years, and specifies which parameters were estimated.

Parameter estimation was by raw-data maximum likelihood as implemented in OpenMx and the fit of nested submodels was compared by likelihood-ratio tests, based on the difference in minus twice the log likelihood (-2LL) between two models. The difference has a chi-square ($\chi 2$) distribution with the degrees of freedom (df) equaling the difference in df between the two models. If constraining parameters in a nested model did not result in a significantly worse fit, this more parsimonious model was deemed the best fitting model. All analyses were performed with and without adjustment for intracranial volume (ICV), which yielded similar results. Here we report the results of the analyses without adjustment for ICV. Because tests were done for 14 related traits (left and right volume of 7 brain structures), the Matrix Spectral Decomposition program (matSPd, Li & Ji, 2005) was used to obtain the number of independent dimensions in the data. This was 10, leading to a p-value of 0.005. Correlations between brain volumes and height and weight were calculated in the Statistical Package for the Social Sciences 21.0 statistical package for Windows (SPSS 21, IBM Corp., 2011).

Figure 1. Longitudinal genetic path model for two twins with brain volume data at ages 9 and 12 years.



Observed phenotype data for two twins at two ages are represented in boxes, latent (unobserved) traits are represented by circles: A = genetic factor score at age 9 and 12 ; E = unique environment factor score at age 9 and 12 ; Ra = correlation between factor scores of twins (Ra = 1 for MZ , 0.5 for DZ same-sex, and was estimated in DZ opposite-sex pairs as is shown here); a_9 $a_{9,12}$ and a_{12} are factor loadings representing the influence of the latent factors on the phenotype.

Based on this model the stability of genetic and environmental influences (the genetic and environmental correlations r(g) and r(e)) can be calculated as:

$$r(g) = \frac{a_9 \text{ x } a_{9,12}}{\sqrt{{a_9}^2} \text{ x } \sqrt{{a_{9,12}}^2 + {a_{12}}^2}} \qquad r(e) = \frac{e_9 \text{ x } e_{9,12}}{\sqrt{{e_9}^2} \text{ x } \sqrt{{e_{9,12}}^2 + {e_{12}}^2}}$$

Results

Brain volumes at age 9 and 12 years

Table 1 presents sample characteristics at ages 9 and 12 years. Comparing height, weight and Tanner data between the 2 ages, we see the expected biological maturation. Figure 2 and supplementary Table 1 summarize the volumes of the subcortical structures. The (left and right) thalamus, amygdala, putamen and pallidum were significantly larger in boys than in girls at age 9 and 12; the volume of the nucleus accumbens was significantly larger in boys than in girls at age 9 but not at age 12. Volume of the thalamus, hippocampus, amygdala and pallidum increases between ages 9 and 12 in boys and in girls. In contrast, volume of the caudate and nucleus accumbens decreases in boys and girls, and findings for the putamen are mixed. However, at α=0.005 these differences do not always reach statistical significance (Supplementary Table S1). We also tested whether these changes in brain volume coincide with increasing height and weight but we found no evidence for this (see Supplementary Table S2).

Volumes of the subcortical brain structures between 9 and 12 years old correlate highly for the thalamus, hippocampus, amygdala, putamen and caudate (> 0.70), and moderate (between 0.30 and 0.90) for the pallidum and nucleus accumbens (Figure 2, and Supplementary Table S3).

Genetic analyses

Twin correlations were larger for the MZ twins than the DZ twin pairs, and were relatively similar for male and female twin pairs, suggesting that additive genetic factors explain most of the variance in subcortical brain volume and that there may not be sex differences in heritability (Supplementary Table S2). Indeed, neither qualitative nor quantitative sex differences in heritability were significant, indicating that the same genetic factors, with the same effect, play a role in boys and girls (Supplementary Tables S4-10). Table 2 summarizes for all subcortical brain volumes at age 9 and 12 the proportions of variance accounted for by A, C and E in the full ACE and the nested AE model. In the ACE model, genetic factors explain most of the variance for all brain volumes with exception of the left nucleus accumbens, ranging from 0.43-0.76 at age 9, and from 0.42-0.72 at age 12. For all volumes, an AE model did not fit the data significantly worse than an ACE model, indicating that familial resemblance can be explained by genetic factors and that effects of the common environment are not significant (see Supplementary Tables S4-10). However, in the case of the left nucleus accumbens a CE model (familial resemblance is explained by shared environmental factors) fitted the data better.

Differences in heritability between ages 9 and 12 were small and the genetic correlations (r(g)) over this 3-year interval were 1.0 (see Table 2). Dropping path a_{12}, which represent the influences of new genes as expressed at age 12 (see Figure 1), from the model did not change the fit of nearly all brain volumes (S4-10). This indicates that the same genetic factors are influencing subcortical brain volumes at age 9 and at age 12, and no significant new genetic effects come into play at age 12. In addition to the genotype, the non-shared environment also contributed to stability for most structures (r(e), Table 2).

As was described by de Geus et al. (2007) and van Soelen et al. (2013), a bivariate model allows for estimation of the heritability of change. To estimate heritability of change scores, the genetic variance is obtained as $(a_{9,12} - a_9)^2 + a_{12}^2$, where the first term reflects (de)amplification (the decrease or increase in shared genetic variance over the 3-year time interval) and the second term the emergence of novel genetic effects at age 12 years. Similar expressions can be derived for the environmental variance. As the results of the bivariate models indicated, estimates for $a_{9,12}$ and a_9 were of the same magnitude, and a_{12} tended to be estimated at zero. Thus, the heritability of change scores in brain volume tends to be around zero (see Supplementary Table S11).

Figure 2. Mean volume in ml for the total (left + right) subcortical brain structure volume for boys and girls at ages 9 and 12 (including 95% error bars). The correlations between volumes at age 9 and 12 are given (left / right) .

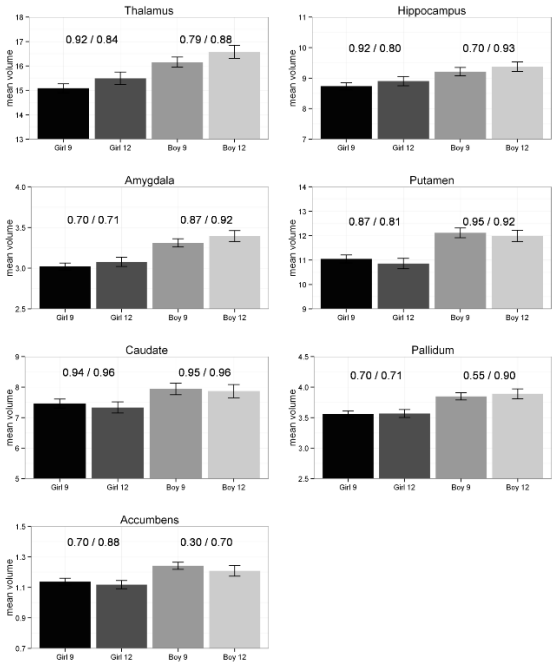


Table 2. ACE and AE model estimates (with 95% confidence intervals) and genetic correlations at age 9 and age 12, covariance explained by shared genetic factors, and fit in the AE model.

		ACE model estimates (95% CI)						AE model estimates (95% CI) and nested fit statistic							
		age 9			age 12			age 9		age 12					
		A	C	E	A	C	E	A	E	A	E	r(g)	%ᵍ	r(e)	p
Thalamus	L	.70	.02	.28	.63	.01	.36	.72	.28	.64	.36	1	93	.17	1
		(.29-.81)	(0-.40)	(.19-.41)	(.27-.78)	(0-.32)	(.22-.54)	(.59-.81)	(.19-.41)	(.46-.77)	(.22-.54)				
	R	.76	0	.24	.72	0	.28	.76	.24	.72	.28	1	90	.32	1
		(.37-.85)	(0-.36)	(.15-.36)	(.31-.82)	(0-.38)	(.18-.44)	(.64-.85)	(.15-.36)	(.56-.82)	(.18-.44)				
Hippocampus	L	.63	.06	.31	.58	.14	.28	.68	.32	.72	.28	1	83	.48	.97
		(.21-.79)	(0-.40)	(.20-.49)	(.15-.82)	(0-.52)	(.17-.45)	(.52-.79)	(.21-.48)	(.55-.83)	(.17-.45)				
	R	.72	0	.27	.69	.01	.30	.71	.29	.70	.30	1	80	.58	1
		(.39-.84)	(0-.26)	(.16-.45)	(.17-.82)	(0-.46)	(.18-.50)	(.51-.83)	(.17-.49)	(.48-.82)	(.18-.52)				
Amygdala	L	.61	0	.39	.72	0	.28	.61	.39	.72	.28	1	83	.40	1
		(.25-.73)	(0-.32)	(.27-.56)	(.35-.84)	(0-.29)	(.16-.48)	(.44-.73)	(.27-.56)	(.52-.84)	(.16-.48)				
	R	.65	.05	.30	.53	.04	.44	.70	.30	.56	.44	1	87	.25	1
		(.19-.80)	(0-.46)	(.20-.45)	(.12-.71)	(0-.42)	(.28-.62)	(.56-.80)	(.20-.44)	(.38-.71)	(.29-.62)				
Putamen	L	.73	.18	.09	.71	.17	.12	.91	.09	.88	.12	1	100	-.04	.84
		(.44-.94)	(0-.47)	(.06-.15)	(.39-.92)	(0-.49)	(.07-.20)	(.85-.94)	(.06-.15)	(.79-.93)	(.07-.21)				
	R	.61	.27	.12	.63	.19	.18	.87	.13	.82	.18	1	97	.14	.57
		(.34-.90)	(0-.53)	(.08-.19)	(.34-.87)	(0-.47)	(.11-.28)	(.81-.91)	(.09-.19)	(.72-.88)	(.12-.28)				
Caudate	L	.50	.24	.26	.58	.21	.21	.74	.26	.79	.21	1	83	.66	.68
		(.14-.82)	(0-.55)	(.17-.40)	(.17-.86)	(0-.57)	(.13-.36)	(.62-.83)	(.17-.38)	(.66-.87)	(.13-.34)				
	R	.43	.32	.26	.42	.33	.26	.75	.25	.75	.25	1	82	.67	.47
		(.06-.81)	(0-.62)	(.16-.41)	(.04-.81)	(0-.64)	(.15-.42)	(.62-.84)	(.13-.38)	(.60-.85)	(.15-.40)				
Pallidum	L	.63	0	.37	.67	.01	.32	.63	.37	.68	.32	1	1	-.01	1
		(.34-.77)	(0-.22)	(.23-.56)	(.25-.80)	(0-.44)	(.20-.51)	(.44-.76)	(.24-.56)	(.49-.80)	(.20-.51)				
	R	.46	.03	.50	.49	.10	.41	.50	.50	.59	.41	1	91	-.11	.99
		(.05-.67)	(0-.32)	(.32-.74)	(.03-.75)	(0-.53)	(.25-.64)	(.28-.67)	(.33-.72)	(.36-.75)	(.25-.64)				
Accumbens	L	.09	.22	.70	.22	.05	.73	.32	.68	.25	.75	1	69	.16	.77
		(0-.51)	(0-.44)	(.49-.89)	(0-.57)	(0-.35)	(.41-.99)	(.09-.53)	(.47-.91)	(.01-.55)	(.45-.99)				
	R	.53	0	.46	.62	0	.38	.53	.47	.61	.39	1	85	.24	1
		(.12-.68)	(0-.37)	(.32-.65)	(.16-.77)	(0-.39)	(.23-.60)	(.35-.68)	(.32-.65)	(.40-.77)	(.23-.60)				

A= additive genetic effects; C= common environment; E= unique environment, r(g)= genetic correlation, %ᵍ = the contribution of shared genetic factors to the covariance between age 9 and 12; r(e)= environmental correlation, p= likelihood-ratio test statistic comparing the AE submodel to the ACE model.

Table 3. Heritability estimates (left / right) from twin studies in healthy children and adults. For each study the number of twin pairs (MZ/DZ) and age range (and mean) of the sample is given.

Children	N pairs	Age	Thalamus	Hippocampus	Amygdala	Putamen	Caudate	Pallidum	Accumbens	Other
Wallace et al., (2010) [4]	107/53	4-19 (12)					85			
Schmitt et al., (2007) *,[4]	127/36	5-18 (11)	88						Basal ganglia: 77	
Yoon et al., (2011) *	57/35	8	59/47			79/77	49/26	81/76		
Peper et al., (2009b, vbm) [1]	45/62	9			83					
This study [1]										
9 years old	48/64	9	72/76	69/73	61/70	91/87	74/75	63/50	33/53	
12 years old	40/49	12	64/72	72/70	72/56	88/82	79/75	68/59	27/61	
Adults	N pairs	Age	Thalamus	Hippocampus	Amygdala	Putamen	Caudate	Pallidum	Accumbens	Other
Kremen et al., (2010)*,[2]	110/92	51-59 (56)	68/60	63/64	63/66	85/84	79/70	66/75	60/48	
den Braber et al., (2013) [3]	176/88	11-56 (29)	80/81	73/78	65/69	86/84	88/86	75/65	65/69	
Bohlken et al., (2013) [3]	50/56	19-55 (30)	81	75	76	80	79	71	49	
Sullivan et al., (2001) *	44/40	68-78 (72)		40						
van Erp et al., (2004) *	23/29	N/A (48)		71						
Panizzon et al.,(2012) [2]	89/68	51-59 (56)		62/66						
Wright et al.,(2002) *	10/10	18-54(27)	0/0	66/71		9/79			Striatum: 33/60	
Brun et al., (2009, vbm) *	23/23	22-25 (24)	25						Basal ganglia: 40	
Hulshoff Pol et al.,(2006, vbm)	54/58	19-69 (31)		80/55						

* Studies are part of the meta-analysis by Blokland et al. (2012). Estimates (left/right) from this meta-analysis were: thalamus 61/52.4, caudate 72.3/64, putamen 78.4/81.6, pallidum 70.7/75.3, hippocampus 58.5/53.2

[1,2,3,4] indicate that analyses are based (partly) on overlapping cohorts.

Note: vmb = heritability estimates from voxel based morphometry. All estimates of other studies are based on volumetric measurements. Basal ganglia include the caudate, putamen, pallidum and nucleus accumbens; striatum includes the caudate and putamen. N/A = age range not available.

Discussion

In this longitudinal twin study we measured volumes of seven subcortical grey matter structures, which play a major role in cognition and emotion. These structures each follow their own pattern of development between 9 and 12 years old. We find high heritabilities for subcortical brain volumes at these ages. No quantitative or qualitative sex differences are found for the heritability estimates, indicating that the same genes, and with the same effect, are expressed in both sexes for these brain volumes. The high correlations between the volumes at age 9 and 12 are due to the stable effects of genetic and environmental influences.

During teenage development, total brain volume increases between the ages of 9 and 12 (van Soelen et al., 2013), but not all subcortical brain structures show the same volumetric increase. In the present study in both girls and boys we find trends of increasing left and right hemisphere volume of the thalamus, pallidum, hippocampus and amygdala between 9 and 12 years of age, while during the same age interval volumes of the caudate, nucleus accumbens, and putamen (bilaterally in boys; right-sided only in girls) decreased.

These results are in line with a growing body of literature that has assessed the volume development of all or most of these subcortical grey matter structures in cross-sectional, longitudinal or mixed-design studies (Dennison et al., 2013; Goddings et al., 2014; Ostby et al., 2009; Wierenga et al., 2014). All these studies show volume decrease of the caudate, nucleus accumbens and putamen, and increases in volumes of the amygdala and hippocampus with development. Results for the thalamus and pallidum are more varied: increases of thalamus volume are reported in the current study and by Ostby et al. (2009), whereas a decrease was found by Dennison et al. (2013). Wierenga et al. (2014) reported a peak volume at 14 years of age followed by a decrease. Similarly, for the pallidum volume increase (our study and Dennison et al., 2013), decrease (Durston et al., 2001; Goddings et al., 2014; Ostby et al., 2009), and inverted U shaped trajectories (Wierenga et al., 2014) are reported. Non-linear trajectories of development, with different peaks for boys and girls, may explain these diverse results. Future studies on brain development need to look beyond the effects of age, and instead take into account the associations of brain development with measurements of body size, hormone levels, or pubertal status (Mills & Tamnes, 2014). Such approaches help us further understand which biological pathways direct brain maturation, see for example longitudinal studies including measurements of body size like height (van Soelen et al., 2013) or studies exploring associations with hormone levels or pubertal status (Koenis et al., 2013; Peper et al., 2009a; Peper, Hulshoff Pol, Crone, & van Honk, 2011).

The heritability estimates from our study are summarized in Table 3, as well as those from all other twin studies that were performed for these seven brain structures. They include studies performed in adults (n=9) and children (n=4), based on nine independent samples (total number of subjects > 1500). Overall, these studies report a wide range of heritabilities for all the subcortical brain structures: thalamus 0-88%; hippocampus 40-80%; amygdala 56-83%; putamen 9-91%; caudate 26-88%; pallidum 50-81%; nucleus accumbens 25-69%. In studies based on childhood samples between 4 and 19 years old, heritability estimates of the thalamus, caudate, putamen and pallidum were high (over 76%, Schmitt et al., 2007; Wallace et al., 2010; Yoon et al., 2011), similar to ours, although lower heritability estimates (26-59%) of the thalamus and caudate at age 8 have also been reported (Yoon et al., 2011). The only studies that have investigated the same seven structures were performed in adult samples, which report heritability estimates in the same range as were estimated in this paper (over 60%, Bohlken et al., 2013; den Braber et al., 2013a; Kremen et al., 2010). Similarly, from their analyses the lower heritability of the nucleus accumbens as compared to the other brain structures is also evident. Although we cannot rule out that accumbens volume is primarily determined by environmental factors, this is possibly a result of measurement error. It might be that the smallest of the subcortical volumes measured in this study is difficult to measure with high precision. This is reflected by the low correlations between volumes over the 3-year interval, as was also shown over a 5-year interval in adult twins (den Braber et al., 2013a). In conclusion, even though heritability estimates may vary between studies, they all illustrate large and stable effects of genetic factors on individual differences in subcortical brain volumes, which does not seem to change substantially to adulthood. Between the sexes, subcortical volumes were on average larger in the males than in the females. Despite the gender differences in average volumes and despite differences in development of sexual characteristics during puberty, we find an absence of significant quantitative and qualitative sex differences in heritability. This finding is in accordance with other studies on heritability of subcortical brain structures and a variety of phenotypes on health and behavior (den Braber et al., 2013a; Vink et al., 2012).

Our sample provides the unique opportunity to assess heritability without confounding effects of age: this study is the first to measure a cohort with only 9 year olds and a follow-up when they were all 12 years old. This thus leaves very little room for effects due to individual differences in age at the time of the scans. The heritability estimates in childhood resemble estimates found in adult samples, which suggests that children may considerably add power in quests trying to find genetic variants influencing brain volume, such as the ENIGMA consortium (Stein et al., 2012; Thompson et al., 2014).

Conclusion

The genome is the most important influence on individual differences in brain volume, both for total volume measures and for most subcortical volumes. Still, there are environmental influences as well. Both genetic and environmental factors need to be identified in follow-up studies aiming to detect genetic variants (in e.g., genome wide association studies) and characterize environmental exposures (e.g., stressors, like life-events). Many studies have focused on global brain development and factors determining individual differences thereof. Subcortical brain structures should be studied next. First of all, they are important for cognitive functions, or play a role in networks that underlie cognitive functions (Aggleton et al., 2010; Aron et al., 2007). During the teenage years, many of these cognitive skills improve (for example executive and social functions, Best & Miller, 2010; Blakemore, 2012; Gur et al., 2012), stressing the importance of healthy brain development during these years. Secondly, during the teenage years there is a high incidence of psychiatric disorders (Lenroot & Giedd, 2006; Paus et al., 2008), many of which are accompanied by (subcortical) brain morphometric changes (Giedd & Rapoport, 2010). The sensitivity of these areas to training, stress, and their involvement in cognitive skills and psychiatric disorders makes it particularly useful to characterize the genetic and environmental causes of (ab)normal brain development of the subcortical grey matter structures.

Table S1. Mean volumes (in ml, with SD) of left (L) and right (R) subcortical brain structures at age 9 and age 12 of girls and boys, and the percentage in volume change (%).

	Girls				**Boys**				
	N 9/12	**9**	**12**	**%**	**N 9/12**	**9**	**12**	**%**	
Thalamus L	106/63	7.56 (.58)	7.85 (.69)	3.8*	101/70	8.20 (.70)	8.44 (.73)	2.9	∩∟
Thalamus R	106/63	7.42 (.53)	7.56 (.66)	1.9	101/70	7.92 (.58)	8.12 (.64)	2.5*	∩∟
Hippocampus L	105/62	4.44 (.39)	4.53 (.42)	2.0*	100/69	4.67 (.42)	4.77 (.42)	2.1	
Hippocampus R	106/62	4.28 (.36)	4.35 (.40)	1.6	99/68	4.52 (.40)	4.60 (.40)	1.8	
Amygdala L	106/63	1.48 (.14)	1.52 (.14)	2.7*	101/69	1.64 (.17)	1.68 (.17)	2.4	∩∟
Amygdala R	106/63	1.53 (.14)	1.57 (.16)	2.6	101/70	1.70 (.17)	1.75 (.18)	2.9	∩∟
Putamen L	106/63	5.59 (.53)	5.62 (.59)	0.5	101/70	6.19 (.63)	6.16 (.57)	-0.5*	∩∟
Putamen R	106/63	5.44 (.53)	5.40 (.56)	-.07	101/70	5.97 (.57)	5.89 (.60)	-1.3	∩∟
Caudate L	106/62	3.67 (.45)	3.63 (.46)	-1.1	100/69	4.02 (.58)	3.94 (.52)	-2.0	
Caudate R	105/63	3.73 (.49)	3.69 (.49)	-1.1	100/68	3.99 (.57)	3.96 (.57)	-0.8	
Pallidum L	106/63	1.85 (.17)	1.88 (.18)	1.6*	101/70	2.02 (.18)	2.06 (.21)	2.0*	∩∟
Pallidum R	106/63	1.67 (.17)	1.71 (.15)	2.4	101/70	1.82 (.21)	1.85 (.23)	1.6	∩∟
Accumbens L	105/61	.54 (.08)	.54 (.08)	0	101/70	.60 (.09)	.58 (.10)	-3.3	∩
Accumbens R	106/62	.61 (.07)	.58 (.08)	-4.9*	101/70	.65 (.08)	.62 (.09)	-4.6*	∩

* indicates that the change in volume between age 9 and 12 is significant

∩ indicates significant difference in volume between boys and girls at age 9

∟ indicates significant difference in volume between boys and girls at age 12

Table S2. Correlations between change in brain volume (left / right) and change in height (centimeter) and weight (kilogram), separately for girls and boys.

		Change in height (Left / Right)	Change in weight (Left / Right)
Girls	Thalamus	-.03 / .07	.12 / -.03
	Hippocampus	.03 / .34*	.00 / .15
	Amygdala	-.09 / -.05	-.03 / -.02
	Putamen	-.03 / .10	.06 / -.13
	Caudate	.03 / -.08	.08 / -.17
	Pallidum	.15 / .02	.10 / -.16
	Accumbens	.09 / .01	.07 / .34*
Boys	Thalamus	-.17 / -.14	.00 / -.04
	Hippocampus	.01 / -.07	.01 / -.22
	Amygdala	.02 / .09	-.15 / .18
	Putamen	-.07 / -.06	-.13 / -.10
	Caudate	.03 / -.15	.10 / -.04
	Pallidum	.14 / .03	.20 / .08
	Accumbens	-.13 / -.05	.05 / -.13

* = significant at α=0.01.

Table S3. The fit of saturated models in -2 log likelihood (-2LL) and Akaike information criterion (AIC). Phenotypic-, twin-, and cross-age correlations in the saturated model are given.

	Thalamus		Hippocampus		Amygdala		Putamen		Caudate		Pallidum		Accumbens	
	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right
Fit:														
-2LL	7.5	1886.54	1617.93	1547.48	1003.33	1061.65	1732.62	1784.31	1677.14	1689.94	1191.24	1225.74	2302.45	2153.71
df	270	270	266	265	269	270	270	270	267	266	270	270	267	269
AIC	1476.28	1346.54	1085.93	1017.48	465.33	521.65	1192.62	1244.31	1143.14	1157.94	651.24	685.74	1768.45	1615.71
Phenotypic correlation volume age 9-12														
girls	.92	.84	.93	.80	.70	.71	.87	.81	.94	.96	.70	.71	.70	.88
boys	.79	.88	.70	.93	.87	.92	.95	.92	.95	.96	.55	.90	.30	.70
Twin correlations, age 9														
MZM	.79	.80	.75	.70	.67	.78	.94	.86	.65	.67	.66	.63	.12	.65
DZM	.60	.24	.48	.38	.32	.17	.29	.05	.56	.70	-.07	.12	.47	.34
MZF	.72	.80	.59	.82	.51	.85	.91	.87	.85	.85	.58	.46	.34	.71
DZF	.11	.43	.47	.25	.54	.24	.47	.70	.18	.09	.04	-.07	.22	.45
DZMF	.35	.51	.14	-.18	.07	.61	.73	.85	.60	.67	.42	.10	.24	-.11
Twin correlations, age 12														
MZM	.59	.70	.66	.73	.87	.65	.96	.86	.72	.64	.76	.79	.70	.53
DZM	-.03	.49	.27	.52	.08	.11	.27	.27	.59	.43	.68	.63	-.09	.16
MZF	.77	.79	.85	.66	.61	.71	.88	.70	.89	.75	.58	.26	.42	.79
DZF	.30	.32	.74	.80	.35	.22	.67	.72	.26	.52	.06	.02	-.16	.60
DZMF	-.11	-.13	.39	-.26	-.13	.41	.58	.64	.65	.64	.57	.47	-.34	.47
Twin correlations, cross age														
MZM	.80	.72	.49	.61	.85	.61	.91	.86	.68	.71	.66	.48	.52	.51
DZM	.31	.21	.33	.47	.20	.09	.62	.10	.52	.58	.15	.28	.28	.14
MZF	.88	.82	.88	.58	.77	.91	.91	.82	.90	.76	.58	.35	.08	.85
DZF	.26	.54	.57	.36	.30	.27	.67	.51	.17	.49	.28	.14	-.06	.69
DZ M_9F_{12}	.36	.31	.42	-.57	-.16	.37	.72	.77	.71	.55	.49	.54	.38	.19
DZ $M_{12}F_9$	-.08	.07	.18	.05	.08	.19	.68	.68	.47	.80	.52	.01	.12	-.32

2-LL = -2 log likelihood; df = degrees of freedom; MZM = monozygotic males; DZM= dizygotic males; MZF = monozygotic females; DZF = dizygotic females; DZMF = dizygotic opposite-sex twin pairs

Table S4. Model fitting results of the bivariate model of thalamus volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		1476.28	2016.28	270			
	1	Age 9, no sex difference	Sat	1490.92	2064.92	287	48.64	17	0
	2	Age 12, no sex difference	Sat	1487.41	2061.41	287	45.13	17	0
	3	Boys, no age difference	Sat	1471.12	2045.12	287	28.84	17	0.04
	4	Girls, no age difference	Sat	1487.37	2061.37	287	45.09	17	0
	ACE	Full ACE		1446.72	2080.72	317			
	1	Ra DZMF =0.5	ACE	1444.98	2080.98	318	.26	1	0.61
	2	No sex difference	1	1437.85	2091.85	327	10.87	9	0.28
	3	CE	2	1443.55	2103.55	330	11.7	3	0.01
	4	AE	2	1431.86	2091.86	330	.01	3	1
	5*	AE, drop a12	4	1429.86	2091.86	331	0	1	1
	6	AE, drop a9,12	4	1481.71	2143.71	331	51.85	1	0
Right	Sat	Saturated		1346.54	1886.54	270			
	1	Age 9, no sex difference	Sat	1358.49	1932.49	287	45.96	17	0
	2	Age 12, no sex difference	Sat	1355.16	1929.16	287	42.63	17	0
	3	Boys, no age difference	Sat	1361.56	1935.56	287	49.03	17	0
	4	Girls, no age difference	Sat	1347.86	1921.86	287	35.33	17	0.01
	ACE	Full ACE		1321.35	1955.35	317			
	1	Ra DZMF =0.5	ACE	1321.39	1957.39	318	2.04	1	0.15
	2	No sex difference	1	1307.83	1961.83	327	4.43	9	0.88
	3	CE	2	1316.85	1976.85	330	15.03	3	0.002
	4	AE	2	1301.83	1961.83	330	0	3	1
	5*	AE, drop a12	4	1299.83	1961.83	331	0	1	1
	6	AE, drop a9,12	4	1356.32	2018.32	331	56.5	1	0

***** *indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

Table S5. Model fitting results of the bivariate model of hippocampus volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		1085.93	1617.93	266			
	1	Age 9, no sex difference	Sat	1081.13	1647.13	283	29.2	17	0.03
	2	Age 12, no sex difference	Sat	1079.86	1645.86	283	27.93	17	0.05
	3	Boys, no age difference	Sat	1085.93	1651.93	283	34	17	0.01
	4	Girls, no age difference	Sat	1090.4	1656.4	283	38.47	17	0
	ACE	Full ACE		1054.36	1680.36	313			
	1	Ra DZMF =0.5	ACE	1052.73	1680.73	314	.37	1	0.54
	2	No sex difference	1	1047.51	1693.51	323	12.77	9	0.17
	3	CE	2	1049.95	1701.95	326	8.44	3	0.04
	4	AE	2	1041.78	1693.78	326	.27	3	0.97
	5*	AE, drop a_{12}	4	1039.82	1693.82	327	.01	1	0.83
	6	AE, drop $a_{9,12}$	4	1082.65	1736.65	327	42.87	1	0
Right	Sat	Saturated		1017.48	1547.48	265			
	1	Age 9, no sex difference	Sat	1015.03	1579.03	282	31.55	17	0.02
	2	Age 12, no sex difference	Sat	1013.14	1577.14	282	29.67	17	0.03
	3	Boys, no age difference	Sat	1016.82	1580.82	282	33.35	17	0.01
	4	Girls, no age difference	Sat	1014.01	1578.01	282	30.53	17	0.02
	ACE	Full ACE		1009.61	1633.61	312			
	1	Ra DZMF =0.5	ACE	1007.67	1633.67	313	.06	1	0.8
	2	No sex difference	1	1009.95	1653.95	322	20.28	9	0.02
	3	CE	2	1018.27	1668.27	325	14.33	3	0.002
	4	AE	2	1003.95	1653.95	325	0	3	1
	5*	AE, drop a_{12}	4	1005.73	1657.73	326	3.78	1	0.05
	6	AE, drop $a_{9,12}$	4	1034.66	1686.66	326	32.73	2	0

***** *indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

Table S6. Model fitting results of the bivariate model of amygdala volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		465.33	1003.33	269			
	1	Age 9, no sex difference	Sat	487.17	1059.17	286	55.84	17	0
	2	Age 12, no sex difference	Sat	484.77	1056.77	286	53.44	17	0
	3	Boys, no age difference	Sat	465.48	1037.48	286	34.15	17	0.01
	4	Girls, no age difference	Sat	471.49	1043.49	286	40.16	17	0
	ACE	Full ACE		435.82	1067.82	316			
	1	Ra DZMF =0.5	ACE	434.22	1068.22	317	.41	1	0.52
	2	No sex difference	1	425	1077	326	8.78	9	0.46
	3	CE	2	429.56	1087.56	329	10.56	3	0.01
	4	AE	2	419	1077	329	0	3	1
	5*	AE, drop a_{12}	4	417	1077	330	0	1	1
	6	AE, drop $a_{9,12}$	4	454.72	1114.72	330	37.73	1	0
Right	Sat	Saturated		521.65	1061.65	270			
	1	Age 9, no sex difference	Sat	554.29	1128.29	287	66.64	17	0
	2	Age 12, no sex difference	Sat	541.43	1115.43	287	53.78	17	0
	3	Boys, no age difference	Sat	516.79	1090.79	287	29.14	17	0.03
	4	Girls, no age difference	Sat	520.3	1094.3	287	32.65	17	0.01
	ACE	Full ACE		496.24	1130.24	317			
	1	Ra DZMF =0.5	ACE	496.28	1132.28	318	2.04	1	0.15
	2	No sex difference	1	484.25	1138.25	327	5.96	9	0.74
	3	CE	2	486.8	1146.8	330	8.55	3	0.04
	4	AE	2	478.29	1138.29	330	.04	3	1
	5*	AE, drop a_{12}	4	476.29	1138.29	331	0	1	1
	6	AE, drop $a_{9,12}$	4	521.97	1183.97	313	45.68	1	0

***** *indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

Table S7. Model fitting results of the bivariate model of putamen volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		1192.62	1732.62	270			
	1	Age 9, no sex difference	Sat	1220.38	1794.38	287	61.76	17	0
	2	Age 12, no sex difference	Sat	1210.25	1784.25	287	51.63	17	0
	3	Boys, no age difference	Sat	1194.37	1768.37	287	35.75	17	0
	4	Girls, no age difference	Sat	1188.63	1762.63	287	30.01	17	0.03
	ACE	Full ACE		1190.56	1824.56	317			
	1	Ra DZMF =0.5	ACE	1190.57	1826.57	318	2	1	0.16
	2	No sex difference	1	1181.06	1835.06	327	8.5	9	0.48
	3	CE	2	1216.17	1876.17	330	41.1	3	0
	4	AE	2	1175.89	1835.89	330	.83	3	0.84
	5*	AE, drop a_{12}	4	1174	1836	331	.11	1	0.74
	6	AE, drop $a_{9,12}$	4	1294.47	1956.47	331	120.58	1	0
Right	Sat	Saturated		1244.31	1784.31	270			
	1	Age 9, no sex difference	Sat	1262.5	1836.5	287	52.18	17	0
	2	Age 12, no sex difference	Sat	1254.23	1828.23	287	43.91	17	0
	3	Boys, no age difference	Sat	1231.97	1805.97	287	21.66	17	0.2
	4	Girls, no age difference	Sat	1238.28	1812.28	287	27.97	17	0.05
	ACE	Full ACE		1206.09	1840.09	317			
	1	Ra DZMF =0.5	ACE	1209	1845	318	4.91	1	0.03
	2	No sex difference	1	1195.55	1849.55	327	4.55	9	0.87
	3	CE	2	1213.33	1873.33	330	23.78	3	0
	4	AE	2	1191.56	1851.56	330	2.02	3	0.57
	5*	AE, drop a_{12}	4	1189.56	1851.56	331	0	1	1
	6	AE, drop $a_{9,12}$	4	1286.13	1948.13	331	96.57	1	0

** indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

Table S8. Model fitting results of the bivariate model of caudate volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		1143.14	1677.14	267			
	1	Age 9, no sex difference	Sat	1138.46	1706.46	284	29.32	17	0.03
	2	Age 12, no sex difference	Sat	1136.18	1704.18	284	27.05	17	0.06
	3	Boys, no age difference	Sat	1126.58	1694.58	284	17.44	17	0.42
	4	Girls, no age difference	Sat	1125.27	1693.27	284	16.13	17	0.51
	ACE	Full ACE		1104.23	1732.23	314			
	1	Ra DZMF =0.5	ACE	1104.85	1734.85	315	2.61	1	0.11
	2	No sex difference	1	1105.95	1753.95	324	19.11	9	0.02
	3	CE	2	1107.94	1761.94	327	7.99	3	0.05
	4	AE	2	1101.46	1755.46	327	1.5	3	0.68
	5*	AE, drop a_{12}	4	1099.46	1755.46	328	0	1	1
	6	AE, drop $a_{9,12}$	4	1156.32	1812.32	328	56.87	1	0
Right	Sat	Saturated		1157.94	1689.94	266			
	1	Age 9, no sex difference	Sat	1158.93	1724.93	283	34.99	17	0.01
	2	Age 12, no sex difference	Sat	1156.61	1722.61	283	32.66	17	0.01
	3	Boys, no age difference	Sat	1157.45	1723.45	283	33.5	17	0.01
	4	Girls, no age difference	Sat	1155.01	1721.01	283	31.06	17	0.02
	ACE	Full ACE		1134.59	1760.59	313			
	1	Ra DZMF =0.5	ACE	1134.94	1762.95	314	2.36	1	0.12
	2	No sex difference	1	1129	1775	323	12.06	9	0.21
	3	CE	2	1128.2	1780.2	326	5.19	3	0.16
	4	AE	2	1125.52	1777.52	215	2.52	3	0.47
	5*	AE, drop a_{12}	4	1123.52	1777.52	327	0	1	1
	6	AE, drop $a_{9,12}$	4	1179.56	1833.56	327	56.04	1	0

***** *indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

Table S9. Model fitting results of the bivariate model of pallidum volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		651.24	1191.24	270			
	1	Age 9, no sex difference	Sat	679.05	1253.05	287	61.8	17	0
	2	Age 12, no sex difference	Sat	665.87	1239.87	287	48.62	17	0
	3	Boys, no age difference	Sat	653.14	1227.14	287	35.9	17	0
	4	Girls, no age difference	Sat	787.34	1361.34	287	170.1	17	0
	ACE	Full ACE		644.48	1278.48	317			
	1	Ra DZMF =0.5	ACE	642.51	1278.51	318	.02	1	0.88
	2	No sex difference	1	630.6	1284.6	327	6.09	9	0.73
	3	CE	2	638.03	1298.03	330	13.43	3	0.004
	4	AE	2	624.6	1284.6	330	0	3	1
	5*	AE, drop a_{12}	4	622.6	1284.6	331	0	1	1
	6	AE, drop $a_{9,12}$	4	660.85	1322.85	331	38.25	1	0
Right	Sat	Saturated		685.74	1225.74	270			
	1	Age 9, no sex difference	Sat	712.49	1286.49	287	60.75	17	0
	2	Age 12, no sex difference	Sat	694.17	1268.17	287	42.43	17	0
	3	Boys, no age difference	Sat	684	1258	287	32.25	17	0.01
	4	Girls, no age difference	Sat	686.59	1260.59	287	34.85	17	0.01
	ACE	Full ACE		673.3	1306.55	317			
	1	Ra DZMF =0.5	ACE	670.83	1306.83	318	.28	1	0.6
	2	No sex difference	1	671.04	1325.04	327	18.21	9	0.03
	3	CE	2	669.83	1329.83	330	4.79	3	0.19
	4	AE	2	665.17	1325.17	330	.13	3	0.99
	5*	AE, drop a_{12}	4	663.17	1325.17	331	0	1	1
	6	AE, drop $a_{9,12}$	4	687.67	1349.67	313	24.5	1	0

** indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

Table S10. Model fitting results of the bivariate model of nucleus accumbens volume at age 9 and 12.

		Model fitted	Against	AIC	-2LL	df	Δ-2LL	Δdf	p
Left	Sat	Saturated		1768.45	2302.45	270			
	1	Age 9, no sex difference	Sat	1772.9	2340.9	284	38.45	17	0
	2	Age 12, no sex difference	Sat	1766.45	2334.45	284	32	17	0.02
	3	Boys, no age difference	Sat	1758.83	2326.83	284	24.37	17	0.11
	4	Girls, no age difference	Sat	1758.85	2326.85	284	24.4	17	0.11
	ACE	Full ACE		1732.7	2360.7	314			
	1	Ra DZMF =0.5	ACE	1731.42	2361.42	315	.72	1	0.4
	2	No sex difference	1	1725	2373	324	11.58	9	0.24
	3*	CE	2	1719.72	2373.72	327	.72	3	0.87
	4	AE	2	1720.16	2374.16	327	1.16	3	0.76
	5	AE, drop a_{12}	4	1718.2	2374.2	328	.01	1	0.84
	6	AE, drop $a_{9,12}$	4	1723.76	2379.76	328	5.6	1	0.02
Right	Sat	Saturated		1615.71	2153.71	269			
	1	Age 9, no sex difference	Sat	1619.23	2191.23	286	37.52	17	0
	2	Age 12, no sex difference	Sat	1615.57	2187.57	286	33.86	17	0.01
	3	Boys, no age difference	Sat	1624.4	2196.4	286	42.68	17	0
	4	Girls, no age difference	Sat	1625	2197	286	43.29	17	0
	ACE	Full ACE		1607.89	2239.89	316			
	1	Ra DZMF =0.5	ACE	1606.93	2240.93	317	1.04	1	0.31
	2	No sex difference	1	1596.32	2248.32	326	7.39	9	0.06
	3	CE	2	1597.95	2255.95	329	7.63	3	0.05
	4	AE	2	1590.32	2248.32	329	.01	3	1
	5*	AE, drop a_{12}	4	1588.32	2248.32	330	0	1	1
	6	AE, drop $a_{9,12}$	4	1621.52	2281.52	330	33.2	1	0

***** *indicates the best fitting model.*

AIC = Akaike Information Criterion; -2LL = -2 log likelihood; df= degrees of freedom; p = p-value.

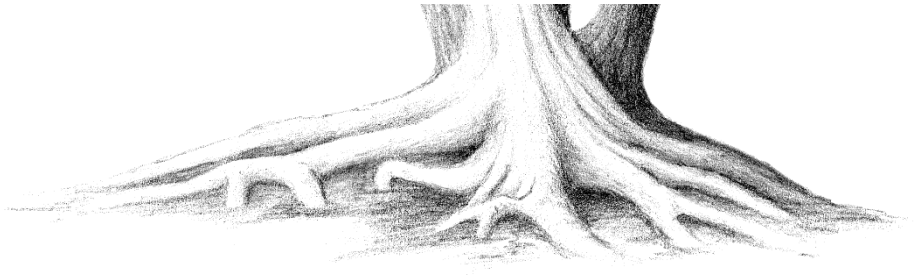
Table S11. ACE model estimate of the heritability of change in brain volume between age 9 and 12.

		A	C	E
Thalamus	Left	0.0014	0.0012	0.9974
	Right	0.0199	0.0	0.9801
Hippocampus	Left	0.0003	0.0555	0.9442
	Right	0.3146	0.0101	0.6753
Amygdala	Left	0.0141	0.0	0.9859
	Right	0.0021	0.0016	0.9963
Putamen	Left	0.0515	0.0005	0.9480
	Right	0.0133	0.0081	0.9786
Caudate	Left	0.0077	0.0616	0.9307
	Right	0.0018	0.0059	0.9922
Pallidum	Left	0.0150	0.0060	0.9790
	Right	0.0013	0.0238	0.9749
Accumbens	Left	0.0378	0.0363	0.9259
	Right	0.0229	0.0078	0.9693

A= additive genetic effects; C= common environment; E= unique environment.

Chapter 10

Nederlandse samenvatting



Gezonde cognitieve functies zijn van belang voor mentale gezondheid en voor het functioneren in ons dagelijks leven. Daarom is het van belang dat er meer bekend wordt over de factoren die van invloed zijn op de hersenen en hersenfuncties die betrokken zijn bij het cognitief functioneren. Dit proefschrift beschrijft factoren die bijdragen aan individuele verschillen in cognitieve prestaties.

Onder cognitieve functies vallen een grote hoeveelheid processen die zowel bewust als onbewust plaatsvinden, zoals aandacht, geheugen, werkgeheugen, problemen oplossen, ruimtelijk inzicht, beslissingen maken, sociaal- en taalbegrip. Sommige mensen ervaren problemen bij het cognitief functioneren. Deze problemen zijn vaak van invloed op de kwaliteit van leven; zo kunnen school- en werk prestaties erdoor worden beïnvloed, maken cognitieve problemen vaak deel uit van psychische stoornissen, en vormen zij een bron van zorgen voor veel ouderen. Daarentegen excelleren andere personen juist in bepaalde cognitieve vaardigheden. De prestatie op de ene functie is niet persé hetzelfde in andere cognitieve domeinen. Individuele verschillen in cognitief functioneren zijn bovendien aanwezig over de gehele levensduur: kinderen leren deze vaardigheden in hun eigen tempo aan, en bij ouderen treedt cognitieve achteruitgang op in meerdere of mindere mate en in verschillend tempo.

Individuele verschillen in cognitieve prestaties kunnen verklaard worden door diverse factoren, de meest bekende zijn leeftijd en sekse. Cognitieve capaciteiten ontwikkelen zich van jongs af aan, het meest duidelijk gedurende de kindertijd. De snelheid van deze ontwikkeling neemt af naarmate een individu ouder wordt, maar ontwikkeling blijft aanwezig gedurende de volwassenheid. Maar gedurende de (late) volwassenheid treedt cognitieve veroudering op: het proces van afname in bepaalde cognitieve domeinen. De leeftijd waarop dat begint en het betreffende cognitieve domein kunnen echter variëren. Bovendien is de adolescentie een periode van belang, want dan is er sprake van belangrijke veranderingen in hersenstructuur en –functie (cognitief en emotioneel). Een bekend sekseverschil betreft het voordeel van vrouwen in verbale vaardigheden, en van mannen in ruimtelijke vaardigheden, maar doorgaans lijken sekse effecten op cognitie klein. Individuele verschillen in cognitie blijven bestaan wanneer wordt gecorrigeerd voor effecten van sekse en leeftijd. De prestatie op deze cognitieve functies hangt af van de activatie van betrokken hersengebieden en netwerken: ook deze activatie vertoont individuele verschillen. Zo blijkt dat intelligentie afhankelijk is van neurale activatie, hersenconnectiviteit, en efficiëntie van netwerken. Dit bevestigt hoe belangrijk een gezonde hersenstructuur en –functie zijn voor normaal cognitief en mentaal functioneren.

Daarnaast spelen genetische factoren een rol bij de individuele verschillen in hersenvolume en functioneren. Zo is intelligentie een hoog erfelijke eigenschap, en ook globale hersenvolumes worden sterk beïnvloed door genetische verschillen. Echter, de erfelijkheid van specifieke hersenvolumes, en ook van specifieke cognitieve vaardigheden, zijn minder vaak het onderwerp van studie geweest, en vertonen mogelijk een lagere erfelijkheid. In tegenstelling tot genetische factoren, kunnen omgevingsinvloeden worden veranderd door interventie. Voor cognitief functioneren zijn omgevingsinvloeden die van belang kunnen zijn mogelijk gerelateerd aan leefstijl en gezondheid. Een voorbeeld van leefstijl is lichamelijke activiteit, waarvan wordt verondersteld dat dit mogelijk de cognitieve problemen bij ouderen en dementerenden kan tegengaan. Een voorbeeld van gezondheid is het verlagen van de bloeddruk, aangezien hoge bloeddruk (hypertensie) zou kunnen leiden tot hersenschade en verminderde cognitieve functies. Dit proefschrift richt zich op al deze vraagstukken.

De studies in dit proefschrift maken gebruik van de gegevens die de afgelopen jaren zijn verzameld bij deelnemers van het Nederlands Tweelingen Register. Deze deelnemers bestaan naast tweelingen uit hun ouders, broers, zussen, kinderen en/of partners. Deze deelnemers hebben meegedaan aan verschillende onderzoeken waar cognitieve testen en vragenlijsten zijn afgenomen, en bij subgroepen zijn onder meer MRI scans van de hersenen gemaakt, is materiaal voor genetisch en hormonaal onderzoek verzameld, en is de activiteit van het autonome zenuwstelsel gemeten. De cognitieve testen bestonden voornamelijk uit de testbatterij van de Universiteit van Pennsylvania: de Computerized Neurocognitive Battery (CNB). Deze testen worden op de computer gedaan en meten binnen korte tijd de prestatie van een grote hoeveelheid cognitieve domeinen.

Tweeling- en familieonderzoek geeft informatie over de relatieve invloed van genen en omgeving doordat bekend is in hoeverre zij deze factoren met elkaar delen. Ten eerste bestaan er twee soorten tweelingen. Identieke (eeneiige) tweelingen worden geboren nadat een eicel zich kort na de bevruchting splitst in twee (genetisch aan elkaar gelijke) individuen. Zo komt het dat zij qua uiterlijk veel op elkaar lijken en altijd van hetzelfde geslacht zijn. Twee-eiige tweelingen komen vaker voor dan identieke, en ontstaan doordat twee eicellen bevrucht worden. Deze individuen zijn genetisch net zo aan elkaar gelijk als andere broers en zussen: zij zijn niet genetisch identiek maar delen gemiddeld de helft van hun genetisch materiaal. Maar 'gewone' broers en zussen worden niet onder dezelfde omstandigheden geboren, zo delen zij bijvoorbeeld niet de prenatale invloeden gedurende hun ontwikkeling in de baarmoeder. Dit is wat de twee-eiige tweelingen waardevol maakt voor dit onderzoek, want zij vormen als het ware de perfecte controlegroep voor de identieke tweelingen.

In tweelingonderzoek wordt de gelijkenis van identieke en twee-eiige tweelingen met elkaar vergeleken. Deze gelijkenis kan worden veroorzaakt doordat de tweeling genetisch materiaal deelt of door gedeelde omgevingsfactoren. Wanneer er een verschil is tussen de mate van gelijkenis tussen de twee soorten tweelingen, dan wordt verondersteld dat deze eigenschap erfelijk is. Als er weinig tot geen verschil is in de gelijkenis tussen de twee typen tweelingen dan zal de bestaande gelijkenis worden veroorzaakt doordat tweelingen omgevingsfactoren met elkaar delen. Tenslotte worden verschillen tussen tweelingen veroorzaakt door de unieke omgeving die zij niet met elkaar delen. Enkele hoofdstukken in dit proefschrift berusten op gegevens die verzameld zijn bij de identieke en twee-eiige tweelingen, en maken gebruik van deze uitgangspunten van het zogenaamde 'klassieke tweeling model'. Andere hoofdstukken maken tevens gebruik van gegevens die zijn verzameld bij familieleden van de tweelingen. Het includeren van familieleden maakt het mogelijk om aanvullende hypotheses te testen, zoals de mogelijkheid dat culturele transmissie een rol speelt bij de gelijkenis tussen familieleden.

In deze these wordt prestatie op cognitief functioneren over het volledige spectrum aan domeinen, en over de hele levensduur onderzocht. De leidende vraag hierin is hoe individuele verschillen kunnen worden verklaard door genetische- en omgevingsfactoren, leefstijlfactoren in het bijzonder. Het eerste deel van dit proefschrift geeft een gedetailleerde beschrijving van de projecten waarop deze hoofdstukken berusten. In **hoofdstuk 2 en 3** worden de participanten, de procedure van de dataverzameling, en de materialen beschreven.

Leesvaardigheid is een belangrijk ontwikkelingsdomein bij kinderen, en is van belang voor de ontwikkeling op andere gebieden. Wanneer er sprake is van problemen op dit vlak leidt dit soms tot de diagnose dyslexie. Ouders met dyslexie hebben een grote kans dat hun kinderen eveneens leesproblemen zullen ervaren. In **hoofdstuk 4** is gekeken of de gelijkenis van familieleden in leesvaardigheid het gevolg is van genetische kwetsbaarheid, of dat dit het gevolg is van de familie-omgevingsinvloeden. Hiervoor is gebruik gemaakt van gegevens van zowel ouders als kinderen. Dit design maakt het mogelijk om de overdracht van ouder naar kind te onderzoeken die niet via genetische overdracht verloopt ('culturele transmissie' genoemd). Bovendien kan in dit design de erfelijkheidsschatting gecorrigeerd worden voor de gelijkenis tussen partners op het gebied van een bepaalde eigenschap. Wanneer partners elkaar selecteren op basis van specifieke eigenschappen (houden van lezen, bijvoorbeeld), dan zijn zij op dit gebied genetisch meer aan elkaar gelijk, wat ertoe leidt dat ook hun kinderen meer op elkaar zullen lijken. Deze genetische gelijkenis leidt tot een overschatting van omgevingsinvloeden en onderschatting

van erfelijkheid. Uit dit hoofdstuk komt naar voren dat individuele verschillen in leesvaardigheid voor het grootste deel worden veroorzaakt door genetische factoren. Deze genetische effecten zijn zowel additief als dominant, samen verklaren zij 64% van de variantie in leesvaardigheid. Scores van ouders en kinderen vertoonden een zekere mate van samenhang, maar er werd geen bewijs gevonden dat deze familiegelijkenis wordt verklaard door culturele transmissie. Dat wil zeggen dat omgevingsinvloeden die worden gedeeld tussen ouders en kinderen geen rol van betekenis spelen bij de familiegelijkenis. Dit hoofdstuk bevestigt allereerst de familiegelijkenis van leesvaardigheid. De oorzaak van deze gelijkenis lijkt van genetische aard en wordt niet significant beïnvloed door de leesomgeving die ouders thuis creëren.

Het volgende deel van dit proefschrift berust op gegevens die zijn verzameld met de Computerized Neurocognitive Battery (CNB). Allereerst diende te worden aangetoond dat deze testbatterij betrouwbare en valide scores oplevert in de Nederlandse populatie. Vandaar dat in **hoofdstuk 5** wordt gestart met een serie analyses die indicatoren van betrouwbaarheid en validiteit vergelijkt met de resultaten die berusten op Amerikaanse steekproeven. In de Nederlandse populatie worden nagenoeg dezelfde gemiddelde scores en intercorrelaties tussen testscores gevonden, en wordt hoge consistentie tussen test-items gerapporteerd (Cronbach's alpha). Bovendien zijn de effectgroottes van leeftijd, sekse en opleiding van dezelfde orde van grootte. Vervolgens werd de mogelijkheid getest of leeftijdseffecten niet-lineair zijn. Dit bleek, naast lineaire effecten, in variërende mate zichtbaar voor vrijwel alle cognitieve functies. Non-lineaire effecten bleken met name sterk in de snelheidsscores van de testen, waarbij de achteruitgang sneller verloopt vanaf de volwassenheid. Aangezien intelligentietesten wereldwijd veelvuldig gebruikt worden, is de mogelijkheid onderzocht of de CNB scores kan opleveren die hier een goede benadering van zijn. Een 'latent factor model' werd gebruikt om te onderzoeken of de variantie tussen CNB accuratesse scores gelijk is aan de variantie tussen IQ subschaal scores. Deze varianties blijken identiek te zijn, al toont de correlatie van 0.82 tussen de CNB factor score en de Totale IQ score aan dat deze benadering vrij goed maar niet perfect is. Dit suggereert echter wel dat de CNB gebruikt kan worden om een inschatting van algemeen intellect te verkrijgen. Dat betekent dat, buiten het wetenschappelijke werkveld, de CNB ook klinische toepassingsmogelijkheden heeft, bijvoorbeeld in de geestelijke gezondheidszorg. Tenslotte is van alle testen de erfelijkheid op twee manieren berekend. Allereerst is de erfelijkheid geschat op basis van enkel tweelinggegevens, en vervolgens gebaseerd op gegevens van alle familieleden. Tweelingen zijn per definitie dezelfde leeftijd, maar wanneer alle familieleden worden geïncludeerd wordt naast de tweeling gelijkenis ook de gelijkenis over generaties geanalyseerd.

Uit deze analyses blijkt dat de cognitieve functies laag tot gemiddeld erfelijk zijn, dit geldt zowel voor de accuratesse- als de snelheid-uitkomstsmaten van de testen. Bovendien zijn de uitkomsten van de tweelinganalyses niet zichtbaar anders dan de uitkomsten van de familieanalyses. Dit suggereert dat gedurende het leven dezelfde factoren van invloed zijn op het cognitief functioneren.

Nadat is gebleken dat de CNB betrouwbare en valide testgegevens oplevert, en dat individuele verschillen in cognitieve prestaties deels erfelijk zijn, kan worden gekeken naar de invloed van enkele omgevingsfactoren. De hoofdstukken 6 en 7 beschrijven de invloed van lichamelijke activiteit en van bloeddruk op cognitieve prestaties. Indien er een relatie bestaat tussen deze variabelen, dan geeft dit mogelijkheden voor preventie, dan wel interventie, met betrekking tot cognitieve achteruitgang.

Er bestaat in de samenleving een breed onderschreven aanname dat regelmatige lichamelijke activiteit een positief effect heeft op het cognitief functioneren. In **hoofdstuk 6** wordt deze relatie nader onderzocht, met aandacht voor enkele factoren die mogelijk bijdragen aan de inconsistente resultaten die tot dusverre gevonden zijn. Allereerst verschilt wat men onder 'lichamelijke activiteit' verstaat. In dit hoofdstuk is het effect van regelmatige beweging gemeten, en niet van een enkele keer. Als bewegingsmaat is gebruik gemaakt van alle sportieve activiteiten die vrijwillig en regelmatig worden uitgevoerd in de vrije tijd, met uitzondering van het dagelijkse wandelen en fietsen. Van deze activiteiten wordt het wekelijkse energieverbruik berekend, afhankelijk van de frequentie, de duur en het type lichaamsbeweging. Initiële analyses suggereerden dat lichaamsbeweging zowel voordelig is voor cognitieve accuratesse als cognitieve snelheid. Wanneer deze analyses echter worden gecorrigeerd voor effecten van leeftijd en sekse blijken deze effecten klein of afwezig. Dat wijst op de mogelijkheid dat de bijdrage van beweging aan cognitief functioneren in de gezonde populatie kleiner is dan in sommige klinische groepen. Bovendien dient er rekening mee gehouden te worden dat sommige cognitieve domeinen gevoeliger kunnen zijn voor de mogelijke gunstige effecten van beweging. Er wordt bijvoorbeeld wel een significant effect gevonden op het domein 'aandacht'.

Vergelijkbaar met de diverse resultaten die tot dusverre zijn gevonden in relatie tot lichaamsbeweging, wordt ook de literatuur met betrekking tot de relatie tussen bloeddruk en cognitie gekenmerkt door inconsistente resultaten. Deze relatie is mogelijk ook complex doordat er nadelige effecten kunnen zijn van zowel hoge als lage bloeddruk, en er in sommige patiëntgroepen wellicht juist voordelige effecten kunnen zijn van hoge bloeddruk. Daarom is in **hoofdstuk 7** gekeken naar zowel lineaire als non-lineaire effecten van

systolische en diastolische bloeddruk (boven- en onderdruk). In hoofdstuk 6 is gebleken hoe belangrijk het is om de effecten op cognitieve scores te corrigeren voor leeftijd en sekse. De effectgroottes van bloeddruk bleken zeer klein. Deze resultaten worden bevestigd in een tweede analyse. Daarin werd de mogelijkheid getest dat bij de relatie tussen bloeddruk en cognitie niet persé van een oorzakelijk verband sprake is. Voor deze analyses werden enkel de gegevens van de identieke tweelingen gebruikt. De tweelingparen werden 'opgesplitst' waarbij in de ene groep de tweelingbroer/zus zat met de hogere bloeddruk, en in de andere groep de tweelingbroer/-zus met de lagere bloeddruk. Aangezien deze tweelingen niet van elkaar verschillen in erfelijk materiaal, leeftijd of sekse, moeten verschillen in hun cognitiescore wel toe te wijzen zijn aan het verschil in bloeddruk. Er zouden aanwijzingen zijn voor een oorzakelijk verband indien de groepen een significant verschil in hun cognitieve prestatie laten zien. Echter, deze resultaten zijn in overeenstemming met de eerste analyses, en geven geen aanwijzing dat er sprake is van een causaal effect van bloeddruk op cognitieve functies.

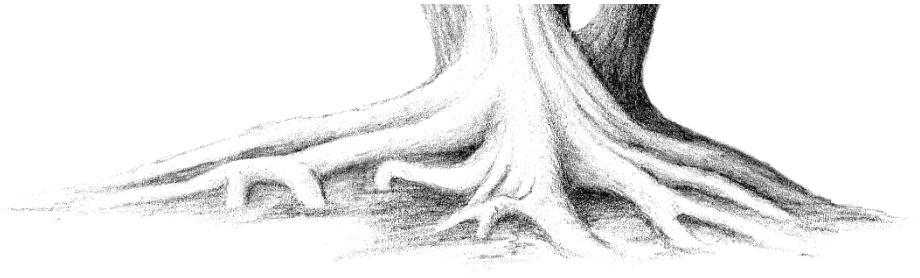
Terwijl voorgaande hoofdstukken zich hebben gericht op hersenfuncties, richt **hoofdstuk 8** zich op de ontwikkeling en erfelijkheid van hersenvolumes bij kinderen tussen 9 en 12 jaar. Vergeleken met globale hersenvolumes is er betrekkelijk weinig onderzoek gedaan naar de erfelijkheid van subcorticale hersenvolumes (hippocampus, thalamus, caudatus, pallidum, putamen, nucleus accumbens en amygdala), en zeker weinig onderzoek bij kinderen. De gegevens waar in dit hoofdstuk gebruik van zijn gemaakt zijn verzameld bij de groep tweelingen en hun broers en zussen die vanaf hun 9e jaar herhaaldelijk hebben meegewerkt aan het Brainscale project, waarin de ontwikkeling van het brein en cognitieve functies gedurende de kindertijd en adolescentie centraal staat. Het volume van enkele hersenstructuren blijkt toe te nemen tussen 9 en 12 jaar, terwijl andere volumes blijken af te nemen, met enkele verschillen tussen jongens en meisjes, en de linker- en rechterhersenhelft. Een toename in volume wordt getoond voor (links en rechts) de thalamus, pallidum, hippocampus en amygdala. Een afname in volume wordt getoond voor de caudatus, nucleus accumbens en putamen (bij jongens links en rechts, bij meisjes alleen rechts). Vervolgens werd van deze hersenvolumes op beide leeftijden de erfelijkheid berekend. Dit werd gedaan in een bivariaat genetisch model, dat het tevens mogelijk maakt om te testen of er op 12 jaar andere genen tot expressie komen dan op 9-jarige leeftijd. Bovendien zijn de analyses apart uitgevoerd voor jongens- en meisjes tweelingen. Door vervolgens te testen of de erfelijkheidsschattingen van jongens en meisjes aan elkaar gelijk kunnen worden gesteld, kan worden onderzocht of sprake is van een interactie tussen sekse en genotype voor hersenvolumes.

In dit hoofdstuk wordt getoond dat de erfelijkheid van alle subcorticale hersenvolumes bij kinderen hoog is op leeftijd 9 en 12, zonder aanwijzingen voor nieuwe genetische effecten op leeftijd 12. Bovendien blijken de genetische effecten, ondanks sekse verschillen in volume en ontwikkelingspatronen, niet verschillend voor jongens en meisjes.

In dit proefschrift zijn enkele factoren aan de orde gekomen die van invloed zijn op diverse cognitieve functies. Effecten van leeftijd, sekse, opleiding en erfelijkheid zijn aan de orde gekomen, deze varieerden in grootte over de verschillende cognitieve domeinen. Bovendien wordt duidelijk hoe belangrijk het is om goed te corrigeren voor de effecten van leeftijd en sekse. Mogelijke factoren die gemanipuleerd kunnen worden, lichamelijke activiteit en bloeddruk, hebben geen grote invloed op cognitieve prestaties in de algemene populatie. Hieruit blijkt dat niet elke persoon in gelijke mate baat zal hebben bij hetzelfde type interventie. Genetische studies kunnen hier een belangrijke rol in spelen door het vinden van de betrokken genen: deze geven een indicatie van de betrokken biologische mechanismen, en kunnen zo helpen om optimale preventie en interventie opties te ontwikkelen voor jong en oud.

Chapter 9

Summary and general discussion



The present thesis examined the influence of genetic and environmental factors on the individual differences in a broad spectrum of cognitive functions, and in subcortical brain volumes. Two modifiable factors that have been hypothesized to influence cognitive functions, i.e. exercise behavior and blood pressure, were studied in detail, while simultaneously recognizing the influence of age and sex on these factors and the outcome traits.

To assess cognitive functions in a large sample with a large variation in age, we translated and validated the well-known Computerized Neurocognitive Battery (CNB). This battery was originally developed by The Brain and Behavior Laboratory of the University of Pennsylvania, with as the main purpose to provide an efficient method to assess performance in a range of cognitive domains that are linked to specific brain systems. This battery has been applied in genetic and treatment studies in English speaking settings. With the Dutch adaptation of the CNB, data were collected in participants registered with the Netherlands Twin Register (see **Chapter 2 and 3**) ranging in age between 10 and 86 years, including a large group of young twins and their siblings who take part in the longitudinal BrainScale project (van Soelen et al., 2012a).

In this last chapter I give a concise summary of the main results, followed by a discussion that integrates these findings and presents their implications for further research.

Summary

Reading

Perhaps the most crucial skill in the early development of cognitive skills is the ability to read fluently and with comprehension. **Chapter 4** focused on reading ability, and the causes of family resemblance in this trait and related disorders, like dyslexia. To explore whether family resemblance is because of transmission of genes from parents to child, or because of shared environmental factors like the household, reading data of twins, their parents and their siblings were analyzed. A model with phenotype information from parents and their twin offspring enables estimation of parameters representing the genetic transmission from parent to child, as well as cultural transmission (that is, transmission through pathways that are not genetically mediated). Individual differences in reading ability were mainly caused by genetic factors, both additive and non-additive (also known as dominant genetic factors), resulting in a broad-sense heritability of 64%. Environmental factors that are shared between parents and children did not contribute to familial resemblance: no evidence was found for cultural transmission from parents to their offspring. In addition, from the parent data, it was clear that there was significant assortative

mating for reading ability, as there was a spouse correlation of 0.38. This study confirms the widely accepted phenomenon that reading (dis)ability tends to "run in families", but is one of the first to study the nature of the transmission from parents to children. The results of this chapter show that this resemblance is due to the transmission of genes, and that there is no additional contribution from the home-literacy environment parents create.

Computerized Neurocognitive Battery

The Computerized Neurocognitive Battery (CNB) was developed by the Brain and Behavior Laboratory of the University of Pennsylvania, and was translated into Dutch by the Department of Biological Psychology, Vrije Universiteit Amsterdam in close collaboration with drs Ruben and Raquel Gur. Chapters 5, 6 and 7 are based on cognitive data collected using the CNB.

First, in **Chapter 5**, reliability and validity of the Dutch translation of the CNB were established, by comparing several outcome measures to those obtained in the U.S. (Gur et al., 2010). Mean accuracy scores on the tests obtained in the Dutch sample were comparable to the U.S. sample, as were intercorrelations between test scores. Further, high Cronbach's alpha's were reported across tests. The (small) deviations in mean scores and intercorrelations from measures in the U.S. sample were most likely due to the use of shortened tests in the Dutch sample, as well as the wider age range including more elderly participants. Validity was confirmed by similar effects of sex and age in the Dutch and US samples. I further explored these age effects by also including non-linear effects of age into a regression model. Linear and non-linear age effects showed a decline in the performance across tests that accelerated around age 50, though in varying degrees. The decline and its acceleration at higher ages were most notable for speed outcomes. Further, cognitive decline as a result of linear age effects was strong for accuracy on e.g., abstraction and mental flexibility, nonverbal reasoning and emotion identification. Non-linear age effects were strong, besides for speed, for accuracy in attention and working memory. Other domains, like verbal reasoning, also showed strong non-linear effects with a late peak around 40-50 years with relative sparing afterwards.

Further validity was shown by the positive associations between cognitive accuracy and speed with educational attainment, both of participants themselves and of their parents. As a last part of the validation procedure, a latent factor model showed that the variance across accuracy measures of the CNB tests was identical to the variance among traditional general intelligence scales. The correlation of 0.82 between the CNB factor scores and Total IQ indicated that performance on the CNB can be used to obtain an approximation

of general intelligence. This suggests that the CNB has purposes beyond research, and may be meaningful in the clinical neuropsychological setting as well.

The heritability of the CNB's cognitive test performance was estimated for all tests. In the first set of analyses data of mono- and dizygotic twin pairs were analyzed, who are of the same age by definition, and it was estimated to what extent their resemblance was due to shared genes, or to common environmental influences shared by offspring growing up in the same family. The other part of the sample with CNB data consisted of family members of twins: parents, siblings, and children of twins and siblings. Therefore, in the second set of analyses data of all family members were analyzed, where cross-generation resemblance was analyzed simultaneously with the resemblance in twin pairs. Overall, estimates based on twin data closely resembled those that were estimated from family data, demonstrating that, where twins form a perfectly controlled design because of equal environmental factors like age and prenatal environment, heritability estimates from multi-generation data do not differ from those based on twin data. This indicates that it is likely that the same genes for cognition are expressed across the lifespan. Family resemblance was to the largest part due to shared genetic factors, and less so due to shared environmental factors, with moderate estimates with wide ranges for both accuracy (1-52%) and speed (14-50%). The latent intelligence factor extracted from the CNB tests was 70% heritable.

After establishing the influence of genetic factors on the cognitive domains of the CNB, the next two chapters focused on the influence of exercise and blood pressure. Unlike sex and age, these factors can to some extent be modified. If clear relationships exist between these two examples of modifiable influences and cognitive functioning, these would present suitable targets for prevention and intervention of cognitive decline or deterioration.

Chapter 6 focused on a generally accepted phenomenon of beneficial effects of regular exercise on cognitive function. Several sources of heterogeneity were addressed in this chapter, including the definition of the phenotype. This chapter made use of a well-defined and reliable measure of voluntary regular leisure time exercise behavior, so the effect of chronic exercise, as opposed to a single bout, was studied. That is, for each participant the average energy expenditure per week (weekly METhours) was calculated, based on the type of exercise and frequency and duration derived from interview data. In addition, effects were studied across cognitive domains to be able to detect whether certain domains were more sensitive to effects of regular exercise than others. Finally, after initially applying univariate regression models between weekly METhours and

cognitive accuracy and speed, these analyses were repeated while correcting for the significant effects of sex, age, and age^2 that were detected in Chapter 5. The univariate models suggested mainly positive associations between weekly METhours and cognitive functions, but after correction for sex and age these associations were small to absent. Even though the relationship may seem intuitive, and is generally accepted by lay people and professionals alike, results of this chapter suggest that in the existing literature, effects of age, sex and exercise on cognitive performance may be confounded.

The associations between cognitive functions and blood pressure (BP) are suspected to be complex, since harmful effects of both high and low BP have been reported, and high BP may be beneficial in specific clinical (e.g., elderly) samples. Therefore, in **Chapter 7** linear and nonlinear effects of diastolic and systolic BP across cognitive domains were studied. In these analyses, cognitive function and BP were corrected for effects of age (linear and quadratic) and sex. Secondly, the possibility that any association between BP and cognition is not necessarily causal was tested. This was done through analyzing data of the monozygotic twins, comparing cognitive test scores of the twins with the higher BP to the co-twins with the lower BP. As monozygotic twins are of the same age and sex, and can be assumed genetically identical, any difference in their cognitive scores must be due to the difference in BP. Causality would be indicated when the MZ twin with the higher blood pressure than the co-twin showed reduced cognitive functioning. Both types of analyses in this chapter failed to provide evidence for a causal effect of blood pressure on cognitive functioning.

Subcortical brain volume

Heritable individual differences in cognitive functioning have been linked to differences in total brain volumes and to measures in cortical structure (Brouwer et al., 2014; Jung & Haier, 2007; Posthuma et al., 2002). Recently, in a large imaging-genetics study genes for subcortical structures have been identified (Hibar et al., 2015), clearly adding to the validity of heritability estimates of subcortical structures in adults (e.g., den Braber et al., 2013a) and in children (e.g., Yoon et al., 2011). However, studies in children are scarce. **Chapter 8** is based on data from a longitudinal study in which magnetic resonance imaging (MRI) scans of the brain were made in addition to cognitive testing. Brain volume data of seven subcortical brain volumes (thalamus, hippocampus, amygdala, putamen, caudate, pallidum and nucleus accumbens) of twins at age 9 and 12 of age were analyzed. First, changes in volume between ages were investigated. Increases in volume were seen for left and right hemisphere volumes of the thalamus, pallidum, hippocampus and amygdala, while volumes of the caudate, nucleus accumbens, and putamen (bilaterally in boys; right-sided only in girls)

decreased. In a bivariate genetic model, effects of genes and environment were estimated at both ages, showing that heritability of all volumes is high from childhood onwards with no evidence for new genetic effects at age 12. Even though the brains of boys and girls show slightly different volumes and develop in a different pattern, genetic effects were similar for boys and girls.

Discussion

This thesis combines cross-sectional and longitudinal data to explore to what extent genetic and environmental factors explain individual differences in cognitive functions throughout life. To do so, multiple indicators of functioning across a wide selection of cognitive domains were obtained in a large twin-family based sample. This effort does not stand alone, as previous twin studies have addressed the genetic architecture of cognition (Polderman et al., 2015). A relatively large amount of these twin studies on cognition are based on measures of general intelligence, as assessed by psychometric IQ tests. These studies have consistently found that heritability of intelligence increases from childhood into adulthood, up till about 80% (Haworth et al., 2010), consistent with findings from several longitudinal studies in Dutch children, which reported increasing heritability of intelligence from (young) childhood into adolescence of about 30 to 60% (Bartels, Rietveld, van Baal, & Boomsma, 2002; van Soelen et al., 2011). Similar increases are seen for verbal and performance IQ (Hoekstra, Bartels, & Boomsma, 2007).

As has become apparent in this thesis, however, different cognitive functions show different sensitivity to sources of genetic and environmental variance. Moreover, the genetic variance is much smaller than that of summary measures like IQ. We found hardly any influence of genetic factors on abstraction and mental flexibility, and most other cognitive functions tested by the CNB showed only small to moderate heritability. These results are congruent with those of a number of previous twin studies addressing the heritability of cognitive test performance on reaction time tasks, working memory tasks, memory tests, attention tests, and WAIS-intelligence subscales and -tests (Kan et al., 2013; Kremen et al., 2011; Kremen, Eisen, Tsuang, & Lyons, 2007; Polderman et al., 2009). The overarching conclusion is that tests of separate cognitive functions are less heritable than more general measures of intelligence. There are two major reasons for this finding that are not mutually exclusive. First, following the logic of 'generalist genes' (Plomin & Kovas, 2005), there may exist a number of genetic variants that have a small but directionally consistent effect on multiple basic cognitive functions. As IQ scores reflect the synergy of all these functions, i.e. the sum of their main effects and any possible interaction terms, the relatively small genetic contribution to each basic cognitive function adds up

to a larger genetic contribution to the so-called 'g-factor'. Likewise, assuming that the measurement errors of each of the tests of basic cognitive functions are uncorrelated, any summary measure of the joint performance on all tests (like IQ) may have a reduced measurement error compared to individual tests. This would lead to higher heritability estimates for IQ as measurement error is part of E.

Our findings showed that research on general cognitive ability can be served by a detailed neurocognitive battery as the latent factor derived from the subtests gives the same answers as using full scale IQ. However, the main aim for developing such batteries has of course not been to replace IQ tests. The raison d'être for these tests is their clinical use in the neuropsychological setting. Many more traditional tests of cognitive functioning are developed to be sensitive, and most ideally specific, to dysfunction that is part of a certain disorder or disease. They were therefore most often not developed to be sensitive to individual differences in cognitive performance. The Computerized Neurocognitive Battery (CNB) was shown in Chapter 5 to be able to validly measure individual differences in a multitude of cognitive domains.

Originally, the CNB cognitive domains were selected because they correspond to specific brain systems. This link between cognitive function and brain system would provide clinical utility since it provides endophenotypes, or biomarkers, of psychopathology. This has been validated in two ways. First, Roalf et al. (2014) confirmed the association of the CNB domains to different brain systems. Although execution of the tests in the five neurobehavioral functions showed some overlap in activated brain areas, there was also test-specific activation of brain systems consistent with other neuroimaging studies. The test of mental flexibility and abstraction activated mainly frontal areas, the attention test activated the frontal-parietal network, the episodic memory tests activated frontal and temporal regions, and the emotion test activated temporo-limbic regions.

The robustness of the structure of the cognitive domains of the CNB is further attested by using different ways of analyzing the performance data. All chapters in this thesis report on analyses that are performed separately for speed and accuracy, but it is also possible to analyze efficiency scores, which are calculated as follows: accuracy / log(speed). Efficiency scores may be most directly comparable to traditional cognitive tests, where speed and accuracy scores are often confounded. However, analyzing them separately would be preferable, as the relation between accuracy and speed may differ per test: they may interact and show a trade-off where a better accuracy score requires longer deliberation before responding. Efficiency scores of tests belonging to the same

cognitive domain showed factor loadings that aggregated within their domains, varying between .45 and .79 (Moore, Reise, Gur, Hakonarson, & Gur, 2015). However, there was a high correlation between the Complex cognition factor and the Executive control factor (.94), suggesting that they are basically the same construct. This was replicated when accuracy scores were analyzed. However, speed scores provided two factors, one for tests that require deliberation, and one for tests that require fast responses and vigilance (attention, working memory, and motor test). Interestingly, correlations between factors were quite high (between .64 and .78), suggesting the presence of an underlying factor that influences all tests. This corresponds to the common factor that was found in Chapter 5.

Sex differences and cognitive functioning across the age range

Sex differences are present across cognitive domains, but they are often of small magnitude, as is clear from the data reported in Chapter 5. While sex differences in certain domains are present in childhood already (Gur et al., 2012) it is unlikely that large and global sex differences exist in cognitive functioning. Prudence dictates that sex is treated as a covariate in cognition research, but generally results will not be strongly affected if one fails to do so. This contrasts sharply with the differences in cognitive functioning found across the life span. This thesis clearly shows that age is a factor that strongly determines cognitive test performance. Moreover, as was shown in Chapter 5, the age effect has significant linear as well as non-linear components for most cognitive domains.

A first source of the deviation from a linear age effect is seen in the period from childhood to adolescence. Accuracy in these domains shows a clear increase in children age 8 to young adults age 21 (Gur et al., 2012). Improvement of cognitive function was most pronounced for executive functions (attention specifically) and motor speed. However, memory was quite good at young ages already and showed relatively minor improvements after age 8. Cognitive development in puberty and adolescence is accompanied by changes in the brain. Specific brain structures develop at a different speed (Gogtay et al., 2004; Lebel & Beaulieu, 2009) and it is possible that the differences between domains of cognitive development are related to differences in timing of local brain development. The developmental patterns of cortical thickness (Brouwer et al., 2014; Burgaleta, Johnson, Waber, Colom, & Karama, 2014; Schnack et al., 2015; Shaw et al., 2006), gray matter density of the cortex (Ramsden et al., 2011) and the white matter network (Koenis et al., 2015) are associated with the level of intelligence, and depend on the brain region. These developmental patterns could be crucial for healthy development, as deviations from this pattern have

been associated with psychiatric disorders (Giedd et al., 2015; Greenstein et al., 2006; Paus et al., 2008; Rapoport & Gogtay, 2008).

There may be valuable information in the individual differences in this maturation process. This was illustrated by Erus et al. (2014) who derived a brain development index in participants up to age 22. This index can identify subjects with brain maturation delay or those who are ahead of their chronological age. Interestingly, subjects with a brain development index that is higher than their age showed better cognitive processing speed on CNB tests, rather than better accuracy.

Whereas cognitive function shows increasingly better performance during the childhood and adolescent years, performance levels peak relatively early in adulthood followed by a decrease towards older ages. This decrease is a second source of the non-linear age effect as it shows acceleration for some but not all of the cognitive functions. Accelerated loss is most apparent for accuracy on the executive function domains, whereas verbal reasoning shows less decrease into old age, probably caused by an increasing number of words existing in the lexicon. It is tempting to explain the non-linear effect of age in the elderly population as a superposition of normal 'linear' cognitive aging and pathological aging in a subset of the elderly, the size of which grows with increasing age. Naturally occurring brain atrophy is suggested to start in early adulthood, and is reflected in altered function, structure and perfusion of the brain (Tarumi & Zhang, 2015). This pathological aging can be related to the various forms of dementia that take an increasing toll on Western societies with a vast burden to these individuals, their families and society. Even before the onset of dementia, declines in memory cause many elderly significant stress and worry. Different types of dementia's are sometimes directly related to neurological abnormalities, like neurofibrillary tangles in Alzheimer, or to vascular damage in frontotemporal dementia. Where loss of neurons is a part of Alzheimer pathology, this does not occur in normal aging. Human and animal studies suggest that normal aging co-occurs with variations in synaptic integrity and plasticity, possibly in networks that are involved in memory and executive functions. Understanding cognitive aging is complicated for several reasons: even at older age the plasticity of neurons enables learning of new or improved skills; general decline may become (temporarily) interrupted by for example stress or medication; or because certain people may more easily find compensatory strategies for their problems (Blazer et al., 2015). This may explain the relatively recent interest in mild cognitive impairment (MCI), an interim clinical diagnostic phase sometimes preceding onset of dementia, but not necessarily so. Because of the impact of even mild cognitive problems on every-day life and wellbeing, biomarkers are required that will predict which persons

will continue to develop severe cognitive problems. Thus far, sensitive biomarkers that can reliably diagnose and predict deviations from normality have been difficult to find.

A thorough understanding of normal cognitive and brain development throughout life will have great clinical utility, as it may provide "growth charts" that may form an instrument to detect deviations or delays from normality. Such understanding may prove crucial for optimizing detection of, and intervention on, impaired cognitive functioning. It is therefore important to know which factors contribute to stability of these traits, and which factors may provide opportunities for change. The present thesis is based on data of both a longitudinal and a cross-sectional study. Ideally, these two types of studies should be combined more often. The strength of the BrainScale study is that it includes a sample of twins that are nearly of the same age, thus minimizing effects of age. Longitudinal designs track individual change and stability over time, and are therefore optimally suited to study cognitive and brain development. However, longitudinal testing of cognitive performance might lead to decreased reliability because of possible test-retest effects. This would mean that the fact that one has previous experience with test administration will influence their scores on a consecutive occasion (Salthouse, 2009). On the other hand, cross-sectional studies cannot control for cohort effects like the Flynn effect (Hofer & Sliwinski, 2001). The Flynn effect is the phenomenon where intelligence in the population increases with time. Both designs have their merits and their disadvantages, and by combining them, a more comprehensive understanding of sources that cause both inter-individual and intra-individual differences will be obtained. An attractive alternative is the parent-offspring design, which can be seen as a 'short-cut' to longitudinal studies and which is well suited, as I showed for reading, for traits that are heritable.

Genetics of brain structure

It is now well known that heritability of global measures of brain (e.g., total brain volume) and cognition (e.g., intelligence) is high. In this thesis I show that measures of subcortical brain structures are highly heritable too and that this is already true in childhood (Chapter 8). This is just a beginning. The field of behavior genetics has in the past decades developed tools to move beyond the first crucial step of estimating the contribution of genetic factors to a variety of traits (Polderman et al., 2015). It is now feasible to study interactions of the genome with sex, age and environmental exposures (GxE interaction). These studies address the difficult questions of genotype-environment covariance, as arising for example from cultural inheritance, decomposing the covariance or comorbidity among traits and studying the longitudinal stability and change in

phenotypes as a function of genes and environment and their interplay (de Kort et al., 2014).

To this new twin methodology the genomic era has added many tools that allow us to go beyond sheer biometric modeling of (latent) genetic effects. Genetic (or genome-wide) association (GWA) studies measure associations of phenotypes with genetic markers covering the entire genome with the aim of identifying the actual causal variants. Because these tests involve large amounts of markers, GWA studies depend on large sample sizes to overcome the large burden of multiple tests. Until recently, performing GWAS on MRI data was difficult because this expensive way of data collection results in relatively small sample sizes. For this reason, the Enigma consortium was established, providing protocols for centers around the world to be able to perform a GWAS on their data (Thompson et al., 2014). After this, results are pooled, resulting in sample sizes large enough to obtain reliable results. One of the efforts undertaken within the ENIGMA consortium was a GWAS on the subcortical brain structures, obtaining several genetic variants related to volume of the hippocampus and putamen (Hibar et al., 2015; Stein et al., 2012).

Genetics of cognitive performance

Specific cognitive functions do not show heritability estimates similar to intelligence or academic achievement. Heritability estimates of cognitive domains are mostly low to moderate (Chapter 5), where reading ability was relatively high (Chapter 4). Heritability of reading and the cognitive domains of the CNB were obtained from samples with a wide age range, and include family pedigrees. These types of analyses thus assume that sources of variance are of equal magnitude for all members in a family, regardless of age. In Chapter 5 heritability was estimated in a twin group as well, in which similar estimates were obtained. This confirms that estimates of genetic and environmental factors are not biased due to this assumption. This provides further opportunities, as this shows that heritability analyses do not necessarily have to include twins, but can be based on nuclear families in a reliable way as well.

For intelligence, a greater number of GWA studies have been performed (e.g. Benyamin et al., 2014; Davies et al., 2011). Several genetic variants were replicated across studies (often expressed in the brain) and across related phenotypes, like educational achievement and school performance (Rietveld et al., 2013; Ward et al., 2014). These genome-wide association studies offer great possibilities to understand the biological pathways to behavior and disorders, even though most behavioral phenotypes are most likely affected by a great number of genes, all with small effect sizes, and related to environmental factors (Davies et al., 2011).

Modifiable effects on cognitive performance

The relatively moderate heritability estimates of performance on tests in the five neurobehavioral functions of the CNB make clear that a moderate to large part of the variance in these functions is caused by environmental factors, leaving room for intervention on the part of the environmental factors that can be modified by behavioral or pharmacological approaches. In Chapter 6 and 7 two such variables were investigated: exercise behavior and blood pressure, both of which are viable targets for intervention. To explore the possibility of beneficial effects of exercise and low blood pressure on cognition, a multitude of cross-sectional studies has tested the association between these variables. This has led to conflicting results. After addressing major sources of heterogeneity in the findings, this thesis found no evidence that regular voluntary exercise in leisure time or low blood pressure are associated with benefits for cognitive performance across the five cognitive domains tested.

It is important to note here that these null-findings do not preclude possible beneficial effects in specific samples. Chapter 6 replicated previous findings: that individuals with attention deficit hyperactivity disorder may benefit from regular exercise as performance on the attention test was positively associated even after taking age and sex into account. In addition, in the elder part of the population, the subset of individuals with beginning pathology, which was very small in our sample, may still prove responsive to physical activity intervention. Several studies have shown that regular and aerobic exercise attenuate the decline in cognitive performance, linked to a protective effect on brain structure and function (Colcombe & Kramer, 2003; Erickson et al., 2011; Steves, Mehta, Jackson, & Spector, 2015). Tarumi and Zhang (2015) further describe how aerobic exercise benefits brain functions, through improvements in arterial pressure regulation (less risk of stiffening of the aorta, atherosclerosis and high blood pressure), blood flow homeostasis (better perfusion of the brain) and metabolic waste clearance (preserving blood supply). Finally, the cognitive effort that is part of exercise activities may itself also account for cognitive improvements (van der Niet et al., 2015). An added complexity is that little is known about the optimal dose of exercise to protect cognition. While it is likely that exercise should be of moderate to vigorous intensity in order to cause substantial benefits, too strenuous or insufficient recovery time may have adverse effects, causing atrophy and lesions in brain matter (Tarumi & Zhang, 2015).

Concluding remarks

This thesis has addressed influences of age, sex, education, heritability and environmental factors across a range of cognitive domains. Potential

determinants of cognitive performance that could be modifiable, exercise behavior and blood pressure, did not show effects on cognition in this general population sample. Clearly, not every person may benefit equally to the same type of intervention. Genetics may offer suggestions for interventions, by suggesting which pathways are involved when networks of genes are found in GWA studies. To create optimized, and maybe even personalized, prevention and intervention options for the young and old, genetically sensitive studies should ideally be combined with longitudinal studies. These are optimally suited to understand the factors that cause individual differences in trajectories of development and aging.

While the majority of cognitive and neurobiological studies is aimed at understanding abnormal development and behavior, disease and disorders, studies in healthy population samples are equally important. This thesis has provided substantial first input for a normative database on cognitive functioning across the lifespan, which will hopefully will be extended by future endeavors using the CNB. This database will help us address a number of pressing questions on the complex effects of genetic and modifiable factors on cognitive functioning, and the intermediate role of brain structure and function.