

Exercise behavior and mental health

A genetic perspective

Marleen de Moor

Reading committee

Prof. dr. P. (Pim) Cuijpers
Prof. dr. R. (Richard) van Dyck
Dr. D.M. (David) Evans
Dr. H.H.M. (Hermine) Maes
Prof. dr. W. (Willem) van Mechelen
Prof. dr. T.D. (Tim) Spector
Dr. J.H. (Janine) Stubbe

Paranimphs

Drs. F.A. (Froukje) Demant
Drs. M.M. (Maartje) Schaap-Nout

Acknowledgements

We thank all participating twin families.

This work was supported by the Netherlands Organization for Scientific Research (NWO): Resolving cause and effect in the association between exercise and well-being (904-61-193); Database twin register (575-25-006), Twin-family database for behavior genetics and genomic studies (480-04-004), Genetic basis of anxiety and depression (904-61-090); Center for Medical Systems Biology (NWO Genomics); Spinozapremie (SPI 56-464-14192); Centre for Neurogenomics and Cognitive Research (CNCR-VU); Genomewide analyses of European twin and population cohorts (EU/QLRT-2001-01254); Genome scan for neuroticism (NIMH R01 MH059160). SNP genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health. Statistical analyses were partly carried out on the Genetic Cluster Computer (<http://www.geneticcluster.org>) which is financially supported by NWO (480-05-003).

Financial support for the publication of this thesis by the Center for Neuroscience and Cognitive Research (CNCR) is gratefully acknowledged.

ISBN: 978-90-9023881-4

Printed by: PrintPartners Ipskamp

Cover design: Laura Teekens (laura@studiotEEKENS.nl)

Lay-out: Marleen de Moor & Laura Teekens

Copyright © Marleen de Moor, 2008, Amsterdam

VRIJE UNIVERSITEIT

Exercise behavior and mental health
A genetic perspective

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Psychologie en Pedagogiek
op vrijdag 13 februari 2009 om 10.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Maria Helena Margaretha de Moor
geboren te Uithoorn

promotoren: prof.dr. D.I. Boomsma
prof.dr. J.C.N. de Geus

‘Although we may never know whether a particular explanation of individual differences is ‘true’ we may, given a suitable experiment, decide whether it is false.’

Lindon J. Eaves (1977). Inferring the causes of human variation.

The Journal of the Royal Statistical Society Series A (General), 140 , 324-355

Voor mijn lieve oma's

Table of contents

1. General introduction	9
PART I	
Exercise behavior and mental health	
2. Regular exercise, anxiety, depression and personality: A population-based study	29
3. Exercise participation and self-rated health: Do common genes explain the association?.	45
4. Testing causality in the association between regular exercise and symptoms of anxiety and depression	59
PART II	
Genetics of exercise behavior	
5. Genome-wide linkage scan for exercise participation in Dutch sibling pairs	83
6. Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs	99
7. Genome-wide association study of exercise behavior in Dutch and American adults	115
8. Exercise participation in adolescents and their parents: Evidence for genetic and generation specific environmental effects	139
9. Summary and general discussion	161
Nederlandse samenvatting (Dutch summary)	175
References	183
Appendices	
A Overview of all exercise-related questions that have been included in the surveys from 1991 to 2004	205
B Mx script for parent-offspring model with sex and generation differences in variance decomposition	221
C Expected covariances among parents and offspring based on parent-offspring model specified in appendix B	249
List of publications	255
Dankwoord	261



1

General introduction



There is a long history of studies on the relationship between exercise and mental health. In one of the first studies dating back to the early 1900s (Franz & Hamilton, 1905), it was reported that two patients with depression showed improvement in mood after participation in moderate exercise. Studies that included more subjects started to appear in the late 1970s (Greist et al., 1979; Morgan & Horstman, 1976) and the first large-scale population-based study on the relationship between exercise and depression, including over 1,900 subjects, is from 1988 (Farmer et al., 1988). Thereafter, the scientific literature has been enriched with an explosive number of studies examining the relationship of exercise with depression, anxiety and other mental health problems.

Broadly, these studies can be divided into two types of studies: experimental and observational studies. Evidence from experimental studies in clinical populations suggests that exercise can relieve symptoms of anxiety and depression in individuals diagnosed with a depressive or anxiety disorder (Brosse et al., 2002; Byrne & Byrne, 1993; Dunn et al., 2005; Lawlor & Hopker, 2001), with effects comparable to anti-depressant medication (Blumenthal et al., 2007). Further, a recent review on the dose-response relationship between exercise and depression suggests that low doses of exercise are also effective to treat individuals suffering from depression (Teychenne et al., 2008). Evidence from experimental studies in non-clinical populations is somewhat more ambiguous. Some trials reported improved well-being after training in previously sedentary subjects (Barbour & Blumenthal, 2002; Brosse et al., 2002; Steptoe et al., 1989) but others failed to note such effects (De Geus et al., 1993; King et al., 1989).

These training studies in volunteers suffer, by necessity, from the shortcoming of selectively attracting participants who are willing to engage in an exercise training program and, since most studies failed to take selective drop-out into account, are vulnerable to selective attrition, where only those who experienced positive psychological effects complete the study.

Observational studies in large population-based samples conducted thus far have well-established that regular exercise is negatively associated with symptoms of anxiety and depression (Brown et al., 2005; Camacho et al., 1991; Cooper-Patrick et al., 1997; Farmer et al., 1988; Kritz-Silverstein et al., 2001; Mobily et al., 1996; Norris et al., 1992; Rhodes & Smith, 2006; Stephens, 1988; Steptoe et al., 1997; Strawbridge et al., 2002; van Gool et al., 2003; Weyerer, 1992; Wise et al., 2006). Some studies use cross-sectional designs (Norris et al., 1992; Rhodes & Smith, 2006; Stephens, 1988; Steptoe et al., 1997) while other studies use longitudinal designs (Brown et al., 2005; Camacho et al., 1991; Cooper-Patrick et al., 1997; Farmer et al., 1988; Kritz-Silverstein et al., 2001; Mobily et al., 1996; Strawbridge et al., 2002; van Gool et al., 2003; Weyerer, 1992; Wise et al., 2006). Most of these studies find that regular exercise at baseline is associated with less depression and anxiety at follow-up. Some studies, however, do not find evidence for a longitudinal association (Cooper-Patrick et al., 1997; Kritz-Silverstein et al., 2001; Weyerer, 1992), and one study finds a longitudinal association in White women but not in men and Black women (Farmer et al., 1988).

Observational studies are vulnerable to confounding. If a “third” underlying factor causes both exercise and well-being, this creates an association between exercise and well-being, but the association does not reflect a causal effect. This confounding is not limited to cross-sectional designs, but could also occur in longitudinal designs. If the confounder influences exercise at baseline and symptoms of anxiety and depression later in time, there will appear to be “prospective evidence” for a causal effect of exercise, whereas, in fact, there is none. Confounding can be remedied by adding an actual measure of the confounder to the study. Obviously, this requires all possible confounders to be known, which is unlikely to be the case. To date, most studies only corrected for a limited number of measured confounding factors and it remains uncertain whether the population association truly reflects a causal effect of exercise on fewer symptoms of anxiety and depression. Potentially confounding factors that have not been taken into account in previous studies include genetic factors. Correcting for genetic factors in the association between exercise and symptoms of anxiety and depression may be crucial, since it is well-known that individual differences in both adult exercise behavior and (symptoms

of) anxiety and depression are partly accounted for by genetic factors.

The heritability of exercise behavior and symptoms of anxiety and depression has been well established. Studies conducted in adult twins aged between 19 and 60 years (Beunen & Thomis, 1999; Eriksson et al., 2006; Kujala et al., 2002; Lauderdale et al., 1997; Stubbe et al., 2006a) show that variation in exercise behavior is accounted for by additive genetic and non-shared environmental factors, with additive genetic factors explaining between 35 and 83% of the variance in exercise. Recent twin studies in adolescence (Carlsson et al., 2006; Maia et al., 2002; Stubbe et al., 2005) show that variation in adolescent leisure time exercise behavior is explained by a combination of additive genetic, shared environmental and unique environmental factors. In a study of adolescent twins aged 13 to 20 years (Stubbe et al., 2005), it was observed that exercise behavior in young adolescents (up to 16 years) is largely determined by shared environmental factors. The influence of these factors rapidly wanes when adolescents become young adults and genetic factors start to appear. After age 18, broad heritability estimates for exercise behavior are as high as 80% (De Geus et al., 2003; Stubbe et al., 2005). Variation in (symptoms of) anxiety and depression is accounted for by a combination of additive genetic and unique environmental factors, with additive genetic factors explaining between 37 and 50% of the variance (Boomsma et al., 2000; Hettema et al., 2001; Sullivan et al., 2000).

In the first part of my thesis, I aim to examine the population association between exercise behavior, symptoms of anxiety and depression and related traits such as self-rated health and general personality traits by testing whether the population association derives from a causal effect of exercise on these traits or whether 'third underlying factors', such as genetic factors, influence both exercise behavior and self-reported physical and mental health. In the second part, I will initiate a second strategy to test whether there are overlapping genetic factors that influence both exercise behavior and anxious depressive symptoms. The idea is to first find the actual genetic variants that influence voluntary exercise behavior and test whether these overlap with the genes implicated in anxiety and depression. In psychiatric genetics, large numbers of studies have been conducted that aimed to identify the genes that affect anxiety, depression and related personality traits such as neuroticism. There has been some progress but solid replication of findings is still lacking (Clement et al., 2002; Fullerton, 2006; Lopez-Leon et al., 2008). Again, this is not for lack of trying; thousands of papers have appeared on this subject and over 100 candidate genes for anxiety or depression are now listed (Sullivan et al.,

2008). In contrast, the number of gene mapping studies for exercise behavior or related physical activity phenotypes is very small (Cai et al., 2006; Loos et al., 2005; Lorentzon et al., 2001; Rankinen et al., 2006a; Salmen et al., 2003; Simonen et al., 2003a; Simonen et al., 2003b; Stefan et al., 2002; Winnicki et al., 2004). Only a handful of candidate genes exist to date. To remedy this, I have performed two genetic linkage studies and a genome-wide association study, to identify novel chromosomal regions and genetic variants that may be related to exercise behavior.

Genetic linkage is the phenomenon that alleles at loci close together on the genome tend to be inherited together, because the chance on a recombination event for two loci close together is very small (Ferreira, 2004; Neale et al., 2007; Posthuma et al., 2003). Therefore, these close loci are said to be 'linked'. Genetic linkage studies make use of this fact by studying the cosegregation of a marker with a trait or disease within families. For example, in pairs of siblings, it can be tested whether an increased sharing of a marker identical-by-descent (IBD) predicts concordance, or resemblance, for the trait under study (Haseman & Elston, 1972). This can be done by multi-point analysis using multiple markers along the genome. If enough markers are used (300 to 400 in sib pair studies) a quantitative trait locus (QTL) has a probability to be close to a marker and IBD at the marker is informative about IBD at the QTL. Microsatellites, which are markers consisting of repeating units that are often highly polymorphic, are frequently used in genome-wide linkage studies, since they are highly informative to determine IBD status. Two genome-wide linkage studies have been conducted for past year physical activity and daily physical activity, one in an adult sample (Simonen et al., 2003b) and one in a sample of children (Cai et al., 2006). Using microsatellite markers typed in sibling pairs and their parents, these studies show suggestive and significant linkages on the chromosomal regions 2p22-p16, 4q28.2, 7p11.2, 11p15, 13q22-q31, 15q13.3, 9q31.1, 20q13.1 (Simonen et al., 2003b) and 18q12-q21 (Cai et al., 2006).

Association studies can be carried out with markers in candidate genes or with genome wide markers. Six studies tested for association of genetic markers (insertion-deletion polymorphisms and single nucleotide polymorphisms (SNPs)) with exercise or physical activity phenotypes using a candidate gene approach. These studies found significant association of exercise behavior or physical activity with the calcium sensing receptor (CASR) gene (Lorentzon et al., 2001), the leptin receptor (LEPR) gene (Stefan et al., 2002), the dopamine 2 receptor (DRD2) gene (Simonen et al., 2003a), the aromatase (CYP19) gene (Salmen et al., 2003), the angio-

tensin-converting enzyme (ACE) gene (Winnicki et al., 2004), and with the melanocortin-4 receptor (MC4R) gene (Loos et al., 2005). These genes are not located in or near the regions detected by the linkage studies, with the exception of the MC4R gene on chromosome 18q12-21, a region for which significant linkage has been found (Cai et al., 2006). Thus, replication of linkage and association studies is needed.

In the last years, it has become feasible to test for association with a phenotype for hundreds of thousands of SNPs, covering a large part of the common genetic variation along the genome (Hirschhorn & Daly, 2005). Important developments have been the completion of the International HAPMAP project (Altshuler et al., 2005), which documents patterns of genomic variation and linkage disequilibrium for different populations, and technical advances in developing dense genotyping chips for tens of thousands of SNPs (Barrett & Cardon, 2006). Genome-wide association studies combine the strength of a linkage study in that an exploratory genome-wide search can be made with the strength of the candidate gene association studies in terms of power to detect relatively small effects, which is extremely useful when studying polygenic traits such as exercise behavior (Hirschhorn & Daly, 2005).

To summarize, in this thesis I aim to study the population association between exercise behavior and symptoms of anxiety and depression from a genetic perspective. By taking a genetic approach in studying the association, it becomes possible to more thoroughly discern association from causation. To this end, I used a large sample of twins and their family members who are registered at the Netherlands Twin Register (NTR) with longitudinal data (1991-2004) on regular exercise and symptoms of anxiety and depression. In the remaining part of this chapter, I will first describe the NTR sample in some detail, as well as the measures of exercise behavior, symptoms of anxiety and depression, and some related personality traits that were analyzed in this thesis. Next, I give a description of the genotyping procedures for two subsets of the sample that were used to conduct linkage and association genetic studies for exercise behavior. I will close this chapter with a brief outline of the contents of the remaining parts of this thesis.

Data collection in the Netherlands Twin Register

Sample description

This thesis is based on data from an ongoing longitudinal study on lifestyle, health and personality in adolescent and adult twins and their family members registered at the Netherlands Twin Register (NTR) (Boomsma et al., 2002; Boomsma et al., 2006b). Since 1991, every two to three years twins and their families receive a survey sent by mail containing standardized personality questionnaires and questions about physical health and lifestyle (exercise behavior, smoking and alcohol use). In 1990 and 1992, twins and their families have been recruited by asking all city councils in the Netherlands to provide the names and addresses of twins aged between 13 and 22 years old. In later years, additional twins and their families were recruited by new efforts of contacting city councils (mainly from larger cities), by advertisements in the media, by announcements in the yearly information bulletin of the NTR and by contacting the Dutch Twin Club (Boomsma et al., 2002).

The total number of individuals that ever participated in survey research of the NTR is 21,426 individuals from 5,573 families. There are 9,646 twins, 82 triplets, 3,160 non-twin siblings, 2,897 fathers, 3,331 mothers, 1,973 spouses of twins and siblings and 337 children of twins and siblings. The sample contains 4,430 twin pairs for which both twins participated at least once in the study. Of these, 694 are monozygotic male (MZM), 507 are dizygotic male (DZM), 1,327 are monozygotic female (MZF), 776 are dizygotic female (DZF), 1,045 are dizygotic opposite-sex (DOS) pairs and 81 pairs are of unknown zygosity. Table 1.1 gives an overview of the participation of twins and their family members at the NTR across the different surveys. Table 1.2 provides an overview of the longitudinal participation of twins and their family members at the NTR. Note that twins could participate in each survey, but parents and siblings were invited for a maximum of 5 surveys, spouses for 2 surveys and offspring of twins only in the last survey. Also note that new twin families could enter the study at any point in time between 1991 and 2004, implying that these participants were also unable to participate in all 7 surveys.

Table 1.1. Participation of twins and their family members at the NTR per survey

	1991	1993	1995	1997	2000	2002	2004
Twins	3,384	4,217	3,409	3,218	4,560	4,472	4,660
Triplets	2	7	4	13	50	51	49
Siblings*	–	–	1,478	1,517	1,474	1,454	1,479
Fathers	1,438	1,774	1,572	1	2	1,266	1,229
Mothers	1,607	1,920	1,688	4	1	1,529	1,580
Spouses**	–	–	–	–	708	1,527	992
Offspring***	–	–	–	–	–	–	337
Total	6,431	7,918	8,151	4,753	6,795	10,299	10,326

* In 1995, families have been invited who also took part in 1991 or in 1993 and at most 1 or 2 sibs per family were invited.

** In 2000, only spouses from twins between 25 and 30 years old were invited. In 2002, all twins also received a survey for their partners. In 2004, spouses were selected who also participated or registered in 2000 or 2002.

*** In 2004, offspring of twins and siblings were invited if they were 18 years old or above.

Table 1.2. Longitudinal participation of twins and their family members at the NTR

	1x	2x	3x	4x	5x	6x	7x
Twins	2,809	1,929	1,830	1,106	880	705	387
Triplets	28	27	22	1	1	2	1
Siblings	1,153	752	538	454	263	–	–
Fathers	895	739	477	452	334	–	–
Mothers	1,089	791	539	520	391	1	–
Spouses	914	864	195	–	–	–	–
Offspring	337	–	–	–	–	–	–
Total	7,225	5,102	3,601	2,533	1,869	708	388

Measurement of exercise behavior

All 7 surveys contain questions about exercise behavior, as shown in Table 1.3. In each survey we asked the participants whether they regularly participate in exercise. If affirmative, further questions were asked about the number of years they had been exercising, type of exercise activities, the frequency of participation in exercise and the duration of an average exercise session (see Table 1.3). The Ainsworth's Compendium of physical activity was used to recode each reported exercise activity into a metabolic equivalent (MET) score. One MET corresponds to the rate of energy expenditure of an individual at rest, which is approximately one kcal/kg/h (Ainsworth et al., 2000). The MET score is thus a measure of the intensity level of the exercise performed. To compute exercise phenotypes, we only used activities that exceeded a minimum intensity criterion of 4 METs. Also, the participants had to have been engaged in one or more of these activities for at least 10 months in the recent year. Using all activities meeting these criteria, we compute the weekly METhours for each participant, which is defined as the product of the MET score with the hours per week engaged in exercise. In this thesis, we used different measures of exercise behavior: a dichotomous measure exercise participation (available in all surveys) which divided participants into non-exercisers and exercisers based on a 4 METhours cut off, a categorical exercise measure, where subjects were divided into non-exercisers, moderate or vigorous exercisers, and the continuous measure METhours (available in all surveys except 1997 and 2000).

Table 1.3. Measurement of exercise behavior

	1991	1993	1995	1997	2000	2002	2004
Exercise participation (yes/no)	X	X	X	X	X	X	X
Type of exercise	X	X	X	X		X	X
Years of exercise participation	X	X	X			X	X
Frequency (times per week)	X	X	X		X	X	X
Duration (minutes per time)	X	X	X		X	X	X

Note: Appendix A contains a full list of all exercise-related questions that have been included in the surveys from 1991 to 2004.

Test-retest correlations were computed in a sample of 200 individuals who filled in survey 2004 (in December) and who were sent a shortened version of this survey containing the same questions on exercise behavior half a year later in June 2005. The tetrachoric test-retest correlation of exercise participation was 0.91 (Stubbe et al., 2007) and the Pearson test-retest correlation of METhours was 0.82 (De Moor et al., 2008) indicating good reliability for both measures of exercise behavior.

Tracking of exercise behavior

The tracking of exercise behavior from adolescence into adulthood over a maximum period of 13 years was determined using all individuals who participated once or more in the study and who were in between 13 and 50 years old at each survey. The Pearson correlations for METhours between different ages of participants aged 13-30 years across surveys are given in Table 1.4. The Pearson correlations for METhours between different ages of the participants aged 31-50 years are given in Table 1.5. The familial relatedness of the subjects has not been taken into account when computing these correlations, but this does not affect the point estimates, and the bias in standard errors is minor (Rebollo et al., 2006).

The tracking of exercise behavior over short periods of time (two to five years) is moderate to high for all age groups, with longitudinal correlations ranging from 0.42 to 0.62. There is a clear decrease in correlations as the time interval increases; longitudinal correlations over a period of 13 years are low (between 0.12 and 0.24). These tracking coefficients are comparable with other studies reporting moderate tracking of exercise behavior from adolescence into adulthood (Malina, 1996; Twisk et al., 2000; van Mechelen et al., 2000). If exercise behavior is not stable across the life span it will be important to test whether changes in exercise participation are accompanied by parallel changes in mental health problems such as anxiety and depression (this is the topic of part I of this thesis). It also highlights the necessity to always include age as a covariate when studying the genetic architecture of exercise in family members of different age, such as for example in siblings and parents of twins (see part II of this thesis).

Table 1.4. Longitudinal tracking of METHours for participants aged 13-30 years (Pearson correlations, number of individuals given in brackets)

	13-14	15-16	17-18	19-20	21-22	23-24	25-26	27-28
13-14								
15-16	0.56 (726)							
17-18	0.42 (322)	0.54 (956)						
19-20	–	0.43 (490)	0.58 (940)					
21-22	0.36 (169)	0.08 (82)	0.46 (503)	0.61 (830)				
23-24	0.26 (406)	0.31 (368)	0.43 (74)	0.52 (427)	0.65 (649)			
25-26	0.23 (389)	0.30 (509)	0.33 (490)	0.49 (112)	0.43 (115)	0.59 (477)		
27-28	0.14 ^{ns} (141)	0.30 (236)	0.34 (647)	0.34 (457)	0.39 (59)	–	0.58 (506)	
29-30	–	0.20 (236)	0.35 (538)	0.32 (581)	0.35 (440)	0.46 (53)	–	0.62 (450)

Note: – = correlation not reported because less than 50 individuals available, ns=non-significant correlation at $\alpha=0.05$

Table 1.5. Longitudinal tracking of METHours for participants aged 31-50 years (Pearson correlations, number of individuals given in brackets)

	31-35	36-40	41-45
36-40	0.42 (253)		
41-45	–	0.46 (514)	
46-50	–	0.44 (274)	0.63 (1,212)

Note: – = correlation not reported because less than 50 individuals available. For participants aged 31-50, mean METHours of each individual during five years was computed, because there were not enough data points to reliably compute the longitudinal correlations across 2-year periods.

Measurement of symptoms of anxiety and depression and other personality traits

An overview of all measures used in this thesis to assess symptoms of anxiety and depression and related personality traits is given in Table 1.6. We used five different measures to assess anxious and depressive symptoms: the anxious depression scale of the Young Adult Self-Report (YASR, 16 items in 3 categories) (Achenbach, 1990), the Spielberger Trait Anxiety Inventory (STAI, 20 items on a 4-point Likert-type scale) (Spielberger et al., 1970), the Beck's Depression Inventory (BDI, 13 items in 4 answering categories) (Beck et al., 1961), the neuroticism scale of the Amsterdamse Biografische Vragenlijst (ABV) (Wilde, 1970), which is similar in content to the neuroticism scale of the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1964), and the somatic anxiety scale of the ABV. The neuroticism scale consisted of 30 items and the somatic anxiety scale of 17 items (3 answering categories). These scales are all interrelated, with correlations ranging between 0.43 and 0.77 (Boomsma et al., 2000; Middeldorp et al., 2006). They further show good predictive validity for the DSM-IV-based disorders of major depressive disorder, dysthymia, generalized anxiety disorder, panic disorders and social phobia, as obtained from a clinical diagnostic interview (CIDI) in a subsample of 1,240 participants (Middeldorp et al., 2006).

Table 1.6. Measurement of anxious and depressive symptoms and personality

	1991	1993	1995	1997	2000	2002	2004
YASR Anxious depression	X		X	X	X	X	
STAI Anxiety	X	X		X	X	X	
BDI Depression		X		X			
ABV Neuroticism	X	X		X	X	X	X*
ABV Somatic anxiety	X	X		X	X	X	
ABV Extraversion	X	X		X	X	X	X*
SSS Thrill and adventure seeking	X	X		X	X	X	
SSS Experience seeking	X	X		X	X	X	
SSS Boredom susceptibility	X	X		X	X	X	
SSS Disinhibition	X	X		X	X	X	
YASR Social problems	X		X	X	X		
Self-rated health	X	X	X	X	X	X	X

Note: only those measures that were used in this thesis are included, for a full list of measures used in surveys from 1991 to 2000 see (Boomsma et al., 2002), for the 2002 survey see Stubbe (2005) and for the 2004 survey see (Distel et al., 2007). * Not based on ABV but on the NEO-FFI short-form (Costa & McCrae, 1992).

To assess variation in normal personality, we additionally used the extraversion scale from the ABV (21 items in 3 categories) and the Sensation Seeking Scale (51 items with a 4-point Likert-type scale, divided over the 4 subscales Thrill and adventure seeking, Experience seeking, Boredom susceptibility and Disinhibition). Finally, we also used the social problems scale from the YASR (8 items in 3 categories) and a five category item to assess self-rated health (Eriksson et al., 2001). All these measures are of potential interest in relation to exercise behavior (Allison et al., 2005; Jack & Ronan, 1998; Norman et al., 2002; Potgieter & Venter, 1995; Simonen et al., 2004).

Genotyping

Two non-overlapping subsets of the sample of adolescent and adult twins and their families registered at the NTR have been genotyped. The first subset consisted of 2,818 twins, siblings and parents who were genotyped by either the Mammalian Genotyping Service in Marshfield for a 400 microsatellite 10 cM genome scan (N=2,399 individuals) or by the Molecular Epidemiology Section, Leiden University Medical Centre, using a 10 cM Applied Biosystems Human Linkage Set v2.5 MD10 with some additional markers (419 microsatellites in 985 individuals). DNA was extracted from either whole blood or buccal swabs following standard protocols (Meulenbelt et al., 1995; Miller et al., 1988). The genotype data from both screens were combined. The data from this subset of individuals were used for the genome-wide linkage analysis described in chapter 5.

The second subset consisted of 1,860 unrelated individuals registered at the NTR who were genotyped as part of the Genetic Association Information Network (GAIN) initiative (Manolio et al., 2007), of which 1,703 served as controls and 160 as cases in a genome-wide association study for major depressive disorder (GAIN-MDD) (Boomsma et al., 2008; Sullivan et al., 2008). DNA was extracted from frozen whole blood samples using the Puregene DNA Isolation kit (Gentra, Minneapolis, MN, USA) according to the manufacturer's protocols. Genotyping was conducted by Perlegen Sciences, using a ~600k high-density oligonucleotide array-based platform. A total of 599,156 genotyped SNPs from 98.5% of all individuals participating in the GAIN-MDD study were returned. After careful quality control, 435,291 SNPs were left for analyses, of which 427,024 were autosomal SNPs. SNPs were excluded because of gross mapping errors (1,487 SNPs), duplicate errors (1,143), Mendelian inconsistencies (536 SNPs), minor allele frequency <0.01 (41,495 SNPs) and missing genotypes >0.05 (156,673 SNPs), or a combination of these reasons. More details on geno-

typing procedures and quality control checks in the GAIN-MDD sample can be found in (Sullivan et al., 2008). The data from this subset of individuals were used for the genome-wide association analysis reported in chapter 7.

Outline of this thesis

This thesis consists of two parts, reflecting the two aims of this thesis:

Aim 1: to describe the population association between exercise behavior and symptoms of anxiety and depression and to test whether this association is causal or derives from a set of common genetic factors.

Aim 2: to further characterize the genetic basis of exercise behavior in adolescents and adults by applying both linkage and association methods, and advanced structural equation modeling techniques.

Chapters 2 to 4 address the first aim. In chapter 2, the association of exercise is explored with symptoms of anxiety and depression, social problems and with the general personality traits sensation seeking, extraversion and neuroticism. It is tested whether these associations depend on gender and age. In chapter 3, I applied a bivariate genetic model (Neale & Cardon, 1992) to examine whether the association between self-rated health and exercise participation is explained by a common genetic factor or whether the association is causal. In chapter 4, I applied a new behavioral genetic method to test for causality between exercise and anxious and depressive symptoms. This method exploited the longitudinal nature of our sample while using all twin pairs to take possible genetic confounding into account.

Chapters 5 to 8 address the second aim of further characterizing the genetic basis of exercise behavior. In chapter 5, a genome-wide linkage scan for exercise participation is reported. The aim of this scan is to identify for which genomic regions genetic resemblance in pairs of siblings is predictive for resemblance in exercise participation. One of the hypothesized mechanisms through which genetic variants may have an impact on exercise participation is through their effect on exercise ability. In chapter 6, a genome-wide linkage scan for exercise ability in British sibling pairs, measured as peak performance level, is reported in order to test whether the linkage regions previously identified for exercise behavior

are also found for exercise ability or whether new linkage regions appear. There was no information on exercise ability in the NTR, but I was able to use a dataset on 4,488 British female twins (of which 700 pairs were genotyped) from the TwinsUK Adult Twin Registry (Spector & Williams, 2006), available through the GenomeUtwinn collaboration (Peltonen et al., 2003). In chapter 7, a genome-wide association study of exercise behavior in two samples of Dutch and American adults is reported. This enabled us to study whether the previously identified linkage regions and candidate genes from previous studies are replicated and to identify new genetic variants that contribute to regular exercise behavior.

In chapter 8, I explored to what extent the exercise behavior of parents influences, through genetic or cultural transmission, the exercise behavior of their adolescent offspring. A parent-offspring design including parents and siblings of adolescent twins was used in order to determine whether the shared environment in adolescents is best explained by the influence of parental exercise behavior on their offspring's exercise behavior, by environmental factors specific to the adolescent generation or by the effects of assortative mating. The parent-offspring model is extended to account for sex and generation differences in variance decomposition of exercise. Finally, in chapter 9, the results from this thesis are summarized and their implications are discussed.





PART I
Exercise behavior and mental health



2

Regular exercise, anxiety, depression and personality: A population-based study

This chapter is published as:

De Moor, M.H.M., Beem, A.L., Stubbe, J.H., Boomsma, D.I., & De Geus, E.J.C. (2006). Regular exercise, anxiety, depression and personality: A population-based study. *Preventive Medicine*, 42 (4), 273-279.

Abstract

Objectives To examine whether regular exercise is associated with anxiety, depression and personality in a large population-based sample as a function of gender and age.

Methods The sample consisted of adolescent and adult twins and their families (N=19,288) who participated in the study on lifestyle and health from the Netherlands Twin Registry (1991-2002). Exercise participation, anxiety, depression and personality were assessed with self-report questionnaires.

Results The overall prevalence of exercise participation (with a minimum of 60 minutes weekly at 4 METs (Metabolic Energy Expenditure Index)) in our sample was 51.4%. Exercise participation strongly declined with age from about 70% in young adolescents to 30% in older adults. Among adolescents, males exercised more, whereas among older adults, females exercised more. Exercisers were on average less anxious (-0.18 SD), depressed (-0.29 SD) and neurotic (-0.14 SD), more extraverted (+0.32 SD) and were higher in dimensions of sensation seeking (from +0.25 SD to +0.47 SD) than non-exercisers. These differences were modest in size, but very consistent across gender and age.

Conclusions This study corroborates and extends previous findings: regular exercise is cross-sectionally associated with lower neuroticism, anxiety and depression and higher extraversion and sensation seeking in the population.

The relationship between exercise behavior and mental health has been examined by many researchers (Byrne & Byrne, 1993; Folkins & Sime, 1981; Gauvin & Spence, 1996; North et al., 1990; Salmon, 2001; Scully et al., 1998). However, population studies on the association between exercise and mental health are scarce. To our knowledge, only two large population studies have examined the cross-sectional association in both males and females across a wide age range (Farmer et al., 1988; Weyerer, 1992). Although the instruments to measure depression have greatly differed, and exercise was based on a single question only, these two studies converge on the main finding that regular exercise is associated with less depression in the population. Both studies only examined the relationship between exercise and depression. The association of exercise with other prominent aspects of mental health, such as anxiety, remains unexplored in population-based samples, as does the association with personality traits.

Both anxiety and depression are known to be correlated with personality, most prominently with neuroticism, but also with extraversion (Costa, Jr. & McCrae, 1980; Middeldorp et al., 2006). Because personality traits can index the risk for anxious and depressive psychopathology, they could potentially mediate the co-morbidity between exercise and anxiety and depression. Personality traits like sensation seeking seem independent of anxiety and depression, but predict lifestyle factors such as smoking or drinking (Koopmans, 1997; Vink et al., 2003) and possibly also exercise behavior. The aim of the present paper, therefore, is to examine the association of exercise with anxiety, depression and personality in a large

population-based sample. By using data from multiple surveys in the same population, we obtained a sample size that allowed us to study these associations as a function of gender and age.

Materials and methods

Study population

This study was part of an on-going study on lifestyle and the health of adolescent and adult twins and their families (i.e., their siblings, parents and spouses) in the Netherlands (Boomsma et al., 2002; Vink et al., 2004). All subjects are voluntary participants in the Netherlands Twin Registry (NTR) and receive surveys on lifestyle and health every two years. Questionnaires were collected in 1991, 1993, 1995, 1997, 2000 and 2002. The procedure by which subjects were recruited is described in detail elsewhere (Boomsma et al., 2002).

The total number of participants in the study on lifestyle and health was 19,469. We excluded seven subjects from our analyses because they were younger than 10 years, and 174 subjects because exercise behavior was unknown. This resulted in a total sample of 19,288 subjects. There were 7,342 subjects (38.1%) who returned a questionnaire on only one wave, 4,705 subjects (24.4%) who participated twice in the study and 7,241 subjects (37.5%) who participated three or more times. Only twins had the opportunity to participate at all occasions. Parents and siblings could participate on four waves and spouses only twice. The information of all subjects at all time points was used resulting in 43,888 observations.

There were 8,773 males (45.5%) and 10,515 females (54.5%). The mean age was 33.0 years ($SD=14.4$). Participants were divided into age groups, each with a range of 5 years. Regarding variables such as socio-economic status, smoking behavior and religious background, the sample used in this study is fairly representative of the Dutch population (Boomsma et al., 2002).

Measures

Exercise participation was measured in detail in each survey. First, the participants were asked whether they exercised regularly ('Yes' or 'No'). If the participants responded affirmative further information on type, frequency and duration of exercise was gathered to calculate their MET-scores. The MET is an index for metabolic energy expenditure. A MET-score of 1 equals the rate of energy expended when at rest (1 kcal/kg/h). Subjects were classified as exerciser if they exercised at least 60 minutes per week

with a MET-score of 4 or more. Since the focus was on self-initiated exercise, physical education at school by adolescents was not counted as exercise. More details are given by Stubbe et al. (2005).

Four variables reflecting anxiety or depression and six personality traits were measured. Table 2.1 summarizes which variables were measured in each survey in subjects with complete information on age, gender, and exercise participation. Anxiety and depression were measured with the short version of the Beck's Depression Inventory (BDI, 13 items, average internal consistency over waves $\alpha=0.80$) (Beck et al., 1961), the anxious depression scale from the Young Adult Self Report (YASR, 16 items, $\alpha=0.85$) (Achenbach, 1990) and the Spielberger State-Trait Anxiety Inventory trait version (STAI, 20 items, $\alpha=0.91$) (Spielberger et al., 1970). Additionally, the social problems scale from the YASR was used (8 items, $\alpha=0.54$).

Six personality traits were measured: neuroticism, extraversion and the four dimensions of sensation seeking. Neuroticism and extraversion were measured with the Amsterdamse Biografische Vragenlijst (Wilde, 1970), a Dutch questionnaire based on the Eysenck Personality Questionnaire (EPQ) (Eysenck & Eysenck, 1964). The neuroticism scale consisted of 30 items ($\alpha=0.89$) and the extraversion scale of 21 items ($\alpha=0.84$). The four dimensions of sensation seeking were measured with the Dutch translation of the Sensation Seeking Scale (SSS) from Zuckerman (Feij et al., 1997; Feij & Van Zuilen, 1984; Zuckerman, 1971). The dimension 'Thrill and adventure seeking' contained 12 items ($\alpha=0.87$), 'Experience seeking' 14 items ($\alpha=0.69$), 'Boredom susceptibility' 13 items ($\alpha=0.66$) and 'Disinhibition' 12 items ($\alpha=0.79$).

Table 2.1. Measurement of anxiety, depression and personality over waves, Netherlands twin family study on lifestyle and health, 1991-2002

	Year of measurement						Total Observations	Total Families
	1991	1993	1995	1997	2000	2002		
Exercise Participation	6,366	7,852	8,120	4,732	6,628	10,190	43,888	4,923
Depression (BDI)		7,745		4,705			12,450	2,864
Anxious Depression (YASR)	3,315*		3,303*	4,685	6,533	9,975	27,811	4,767
Anxiety (STAI)	6,284	7,759		4,693	6,600	10,089	35,425	4,908
Social Problems (YASR)	3,315*		3,303*	4,685	6,533		17,836	4,226
Neuroticism (ABV)	6,269	7,733		4,694	6,568	9,852	35,116	4,898
Extraversion (ABV)	6,281	7,760		4,694	6,573	9,857	35,165	4,898
Thrill Adventure Seeking (SSS)	6,304	7,774		4,695	6,551	9,968	35,292	4,900
Experience Seeking (SSS)	6,295	7,763		4,694	6,527	10,063	35,343	4,901
Boredom Susceptibility (SSS)	6,293	7,757		4,692	6,531	10,068	35,341	4,902
Disinhibition (SSS)	6,299	7,765		4,691	6,533	10,074	35,362	4,902

* Only twins filled in the YASR questionnaire on these waves.

Statistical analyses

Linear mixed modeling was used to test for differences in means on anxiety, depression and personality traits between exercisers and non-exercisers. Linear mixed models allow for modeling statistical dependencies among observations by including random effects, such as dependencies due to nested sampling. In our sample, the data were nested at two levels: individuals were nested within families and repeated measurements were nested within individuals.

The effects of exercise participation, gender and categorized age (including interaction effects) on the anxiety, depression and personality variables were modeled by including them as fixed effects in the model. The means of the repeated measurement were modeled by including a constant as a fixed effect for each measurement occasion. The model included a family effect that varied randomly over families. Within family covariances were estimated by the single variance component associated with this effect. Variances and covariances among the repeated measurements were estimated by the unrestricted residual variances and covariances. We used the statistical package S-Plus for the analyses (S-Plus, 2001). As expected, both the variances of the family effect on anxiety, depression and personality and the correlations between repeated measurements over time within the same individual significantly differed from zero. This indicates that the correction for family structure and repeated measurements was necessary.

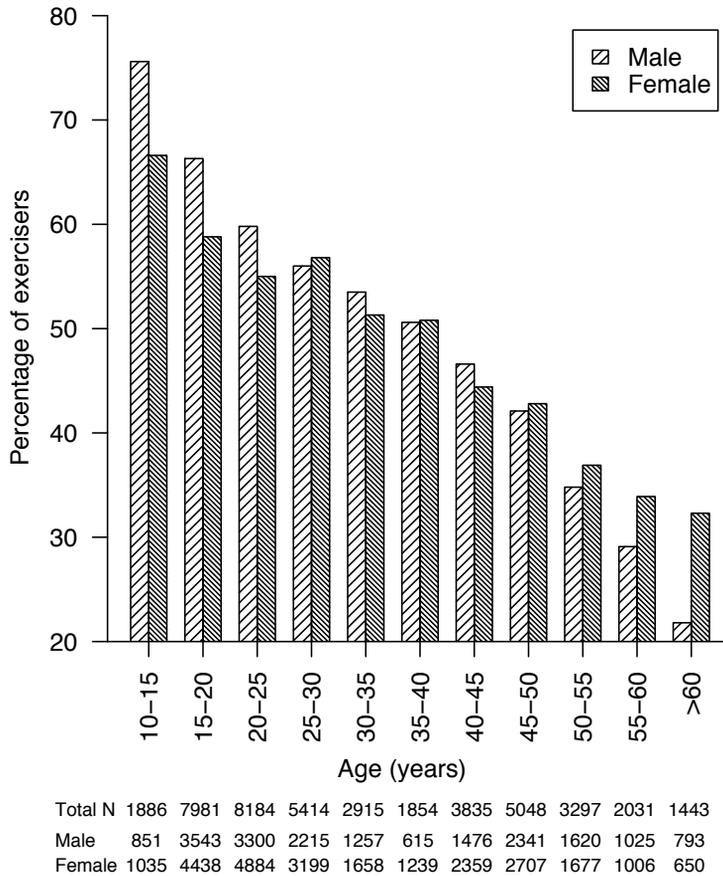
The analyses were performed for all ten dependent variables. The overall significance level was set to alpha 0.01. To correct for multiple testing we used the Bonferroni method to obtain the significance level used for every test. With ten dependent variables and eight different fixed effects, this led to the use of an alpha of 0.00013 for each fixed effect.

Results

Prevalence of exercise participation

Figure 2.1 depicts the prevalences of exercise participation in the Netherlands across gender and age groups. The overall prevalence of exercise participation (in percentage exercising subjects) for this sample is 51.4%. The overall prevalence for men is 52.0% and for women 50.9%. There is a clear decline in the prevalences over age for both males and females: older subjects exercise less. In the young age (10 to 25 years) more men than women exercise, in the older age (45 years and older), more women exercise than men.

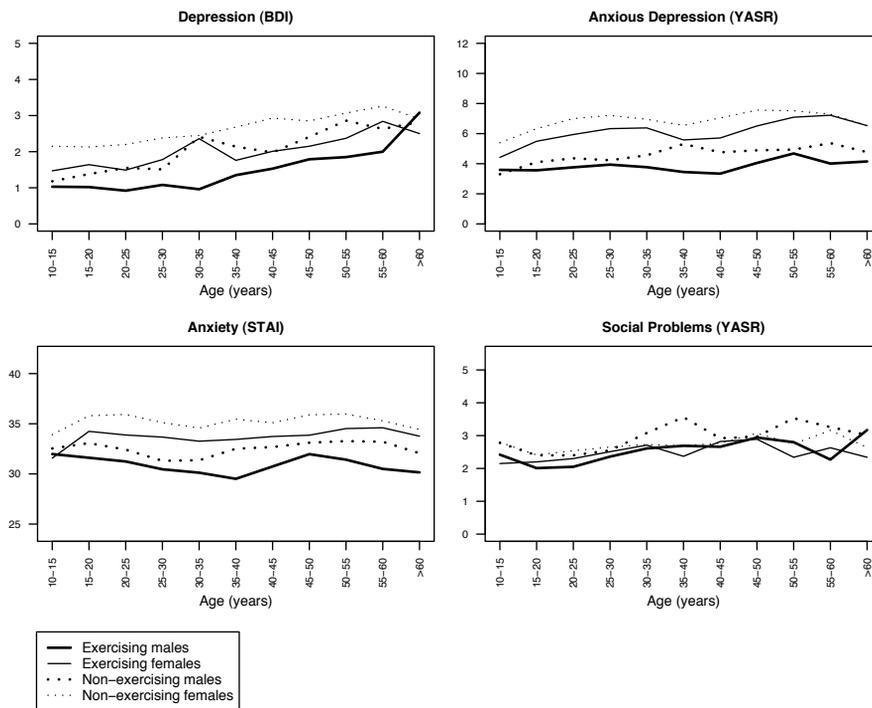
Figure 2.1. Prevalence of exercise participation across gender and age, Netherlands twin family study on lifestyle and health, 1991-2002



Effects of exercise participation, gender and age on anxiety and depression

In the upper part of Table 2.2, the results of the significance tests of the effects of exercise, gender and age on anxiety and depression are given, based on approximate F-tests. The upper part of Table 2.3 contains the observed means for anxiety and depression as a function of exercise. Figure 2.2 shows the observed means of anxiety and depression for the exercise and gender groups over age. All main effects are significant, except for the effects of gender and age on social problems. Exercisers on average score lower on anxiety, depression and social problems than non-exercisers. Expressed in pooled standard deviations, exercisers are 0.29 SD lower in depression, 0.18 SD lower in anxious depression, 0.19 SD lower in anxiety and 0.15 SD lower in social problems.

Figure 2.2. Anxiety and depression as a function of exercise, gender and age, Netherlands twin family study on lifestyle and health, 1991-2002



Note: To allow for comparison between graphs, values on the y-axes range from mean - 1 SD to mean + 1 SD for all variables

Table 2.2. Effects of exercise participation, gender and age on anxiety, depression and personality, Netherlands twin family study on lifestyle and health, 1991-2002

		Main effects			2-way interactions			3-way interaction
		Exercise	Gender	Age	Ex* Gender	Ex* Age	Gender* Age	Ex* Gender* Age
	df	1	1	10	1	10	10	10
Depression (BDI)	F	32.5*	49.6*	23.1*	0.0	1.4	1.0	1.0
Anxious Depression (YASR)	F	23.7*	740.7*	6.9*	0.2	0.8	3.2	1.9
Anxiety (STAI)	F	71.0*	523.5*	11.9*	0.2	0.9	5.1*	0.9
Social Problems (YASR)	F	16.1*	5.3	3.3	0.5	1.0	2.3	1.0
Neuroticism (EPQ)	F	46.4*	833.2*	40.4*	2.2	1.4	2.0	2.6
Extraversion (EPQ)	F	115.9*	6.8	93.4*	0.0	1.2	4.4*	0.5
Thrill Adventure Seeking (SSS)	F	167.6*	2657.8*	975.2*	5.8	0.9	20.7*	1.4
Experience Seeking (SSS)	F	9.4	438.0*	80.7*	1.2	2.3	10.1*	1.4
Boredom Susceptibility (SSS)	F	4.3	206.5*	38.1*	0.8	2.1	5.2*	1.1
Disinhibition (SSS)	F	14.6*	3328.8*	390.8*	0.1	1.7	10.8*	1.2

* $p < .00013$

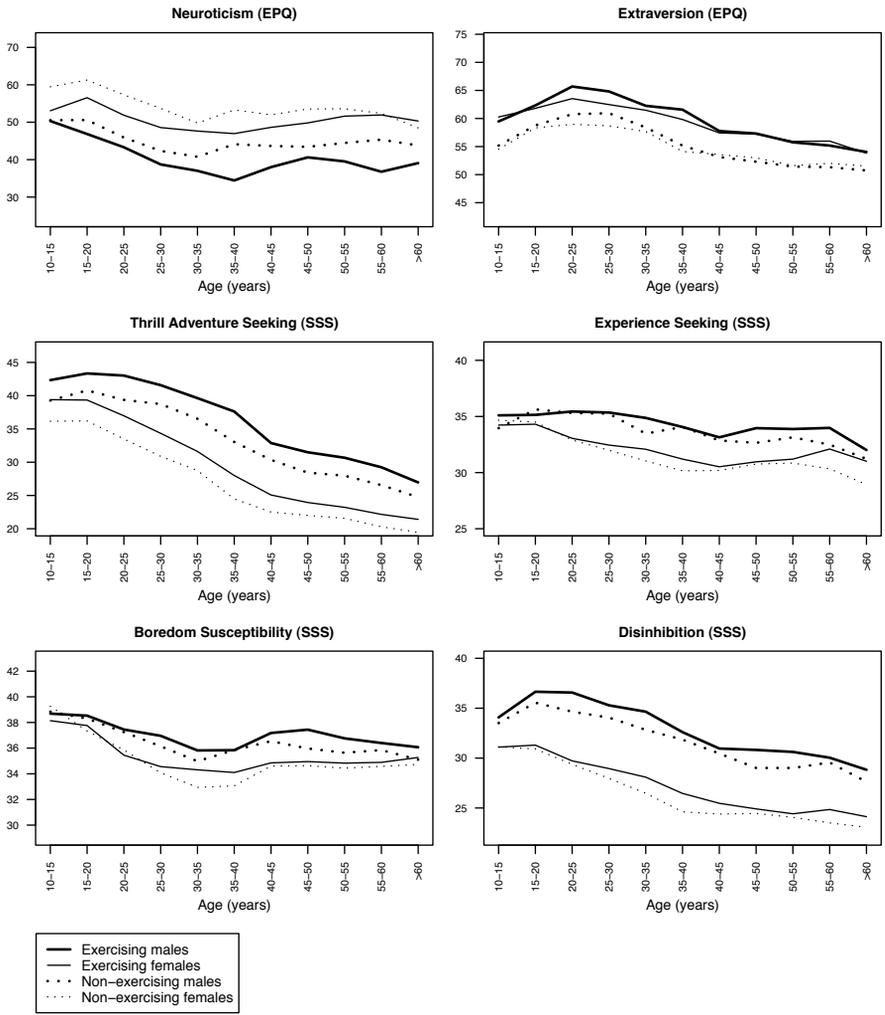
Effects of exercise participation, gender and age on personality

In the lower part of Table 2.2, the results of the significance tests of the effects of exercise, gender and age on personality are shown. The lower part of Table 2.3 gives the observed means of the personality traits as a function of exercise. Figure 2.3 presents the observed means of personality for the exercise and gender groups over age. Main effects of exercise, gender and age on personality are found for all personality measures, except for an effect of exercise on experience seeking and boredom susceptibility and an effect of gender on extraversion. Exercisers on average score higher on extraversion (+0.32 SD), thrill and adventure seeking (+0.47 SD) and disinhibition (+0.25 SD) and lower on neuroticism (-0.14 SD).

Table 2.3. Anxiety, depression and personality as a function of exercise participation, Netherlands twin family study on lifestyle and health, 1991-2002

Exercise	No		Yes	
	N	M (SD)	N	M (SD)
Depression (BDI)	5,738	2.3 (3.1)	6,712	1.5 (2.4)
Anxious Depression (YASR)	12,586	6.0 (5.2)	15,225	5.1 (4.6)
Anxiety (STAI)	17,088	34.2 (9.1)	18,337	32.6 (8.2)
Social Problems (YASR)	7,186	2.6 (2.1)	10,650	2.3 (2.0)
Neuroticism (EPQ)	16,908	50.5 (25.2)	18,208	47.2 (23.7)
Extraversion (EPQ)	16,945	55.7 (17.2)	18,220	61.1 (16.1)
Thrill Adventure Seeking (SSS)	16,996	30.2 (11.3)	18,296	35.5 (11.2)
Experience Seeking (SSS)	17,032	32.6 (8.3)	18,311	33.5 (7.6)
Boredom Susceptibility (SSS)	17,025	35.6 (7.5)	18,316	36.4 (7.3)
Disinhibition (SSS)	17,044	28.9 (8.2)	18,318	30.9 (8.0)

Figure 2.3. Personality as a function of exercise, gender and age, Netherlands twin family study on lifestyle and health, 1991-2002



Note: To allow for comparison between graphs, values on the y-axes range from mean - 1 SD to mean + 1 SD for all variables

Discussion

The primary findings from this investigation are that exercisers are on average less anxious and depressed (effect sizes from -0.18 to -0.29 SD), less neurotic (effect size -0.14 SD), more extraverted (effect size $+0.32$ SD), higher in thrill and adventure seeking (effect size $+0.47$) and higher in disinhibition (effect size $+0.25$ SD) than non-exercisers. Although the differences between exercisers and non-exercisers are small (Cohen, 1969), they are very consistent across gender and age. Strikingly, the associations with exercise participation were independent of the well-known main effect of age on exercise prevalence, that is, all associations held for 20-year olds when exercise prevalence is still 70% but also for 60-year olds when exercise prevalence has dropped below 30%.

The association of exercise with anxiety and depression corroborates previous reports. Lack of exercise was found to be cross-sectionally associated with depression in population samples with a broad age range (Farmer et al., 1988; Weyerer, 1992) and in samples consisting of young (Steptoe et al., 1997; Steptoe & Butler, 1996) or older adults (Kritz-Silverstein et al., 2001; Strawbridge et al., 2002). In a sample of adolescents, however, Allison et al. (2005) found that regular exercise was associated with better social functioning, but not with less anxiety and depression.

To date, only a few other studies have addressed the relation between exercise and personality, although none in a large population-based European sample. Our results are concordant with these previous reports. In a large Japanese population study (Arai & Hisamichi, 1998) exercisers scored higher on extraversion and lower on neuroticism. In a sample of South African students (Potgieter & Venter, 1995) adherers to exercise score significantly higher on extraversion and lower on neuroticism than subjects who drop out from an exercise program. Two of three small studies examining exercise and sensation seeking (Franken et al., 1994; Jack & Ronan, 1998; Potgieter & Bisschoff, 1990) reported higher sensation seeking in exercisers (Jack & Ronan, 1998; Potgieter & Bisschoff, 1990).

We classified subjects of all ages as exercisers only when they met a minimum criterion of 60 minutes at 4 METs weekly. This is a rather strict criterion in comparison to previous studies on exercise and depression (Camacho et al., 1991; Farmer et al., 1988; Weyerer, 1992). However, the prevalences of exercise participation in our sample are quite consistent with prevalences for the Dutch population in 1999 estimated by Statistics Netherlands (50%; Van der Gils et al., 2002) and the Social and Cultural Planning Office of the Netherlands (65%; Van der Meulen, 2003), although

in both of these studies a broader classification of exercise was used. Both studies found the difference between males and females in main prevalence of exercise to be negligible and they observed the same strong decline with age. The consistency of the pattern of prevalence of exercise participation in our study with these reports indicates that our sample is representative for the Dutch population.

Cross-sectional analyses, as presented here, cannot inform us on the causal structure between exercise, personality, anxiety and depression. The most direct test of causality is an experimental design in which exercise behavior is manipulated, preferably in a combined training/detraining design (De Geus et al., 1993). Many training studies on different aspects of mental health have been performed. Careful meta-analyses and qualitative reviews of these studies yield an incoherent picture (Craft & Landers, 1998; de Geus et al., 1993; Long & Vanstavel, 1995; Salmon, 2001). A fundamental problem with training studies may be that they use selected samples, in which two groups of subjects are always underrepresented: those who already exercise vigorously and those who are persistently sedentary and refuse to participate (or drop out early). This makes the results of training studies inherently difficult to generalize to the population, because they may fail to fully capture the source of the association between exercise and mental health in the population at large.

An alternative approach to causality, therefore, is the use of prospective analyses in large population-based samples. A prospective association between lack of exercise at baseline and depression or anxiety at follow-up was found in some population studies (Camacho et al., 1991; Farmer et al., 1988; Strawbridge et al., 2002), but was absent in other studies (Allison et al., 2005; Cooper-Patrick et al., 1997; Kritz-Silverstein et al., 2001; Weyerer, 1992). These studies did not, unfortunately, examine the reverse causality, where depression or anxiety at baseline may predict reduced exercise participation at follow-up. Such reverse causality may be particularly plausible as an explanation for the association between exercise behavior and neuroticism and extraversion. These traits show stable individual differences from an early age onward (Caspi & Roberts, 2001) that may precede active choices for voluntary leisure time exercise behavior.

Even if we accept a model where personality influences exercise behavior, exercise may still have a direct causal effect on anxiety and depression. Effectively, this would mean that the risk conveyed by personality on mental health acts through its influence on exercise behavior. In a different scenario exercise may have an indirect causal effect by moderating the “normal” effects of personality on mental health. Various other

scenarios can be conceived and none can be ruled out based on the present data. Finally, all three variables -exercise, personality, and mental health- may derive from “third” underlying factors such as socio-economic background or pleiotropic genes affecting the (neuro)biology crucial to these variables simultaneously.

By showing a small but robust cross-sectional association of exercise with anxiety, depression and personality that holds in both genders and across age, this study paves the way for future investigation of the causal structure of the associations of exercise with personality and anxiety and depression in population-based samples.



3

Exercise participation and self-rated health: Do common genes explain the association?

This chapter is published as:

De Moor, M.H.M., Stubbe, J.H., Boomsma, D.I., & De Geus, E.J.C. (2007). Exercise participation and self-rated health: Do common genes explain the association? *European Journal of Epidemiology*, 22 (1), 27-32.

Abstract

The purpose of this study is to investigate whether there is an association between exercise participation and self-rated health and whether this association can be explained by common genes and/or common environmental influences. In a sample of 5,140 Dutch adult twins and their non-twin siblings from 2,831 families, exercise participation (sedentaries, light or moderate, vigorous exercisers) and self-rated health were assessed by survey. To investigate the etiology of the association, bivariate genetic models using structural equation modeling were applied to the data. The correlation between exercise participation and self-rated health is significant but modest ($r=0.20$). Exercise participation and self-rated health are both heritable (around 50% of the variance of both phenotypes is explained by genetic factors). The genetic factors influencing exercise participation and self-rated health partially overlap ($r=0.36$) and this overlap fully explains their phenotypic correlation. We conclude that the association between exercise and self-rated health can be explained by genes predisposing to both exercise participation and self-rated health. These genes may directly influence both phenotypes (pleiotropy). Alternatively, genes that affect exercise or self-rated health may indirectly influence the other phenotype through a causal relationship. We propose that identification of the genes that cause differences in exercise behavior will help resolve the issue of causality.

In the last decades it has become increasingly clear that persistent regular exercise during adulthood protects from both physical health problems (i.e., cardiovascular disease) and mental health problems (i.e., anxiety and depression) (Berlin & Colditz, 1990; Camacho et al., 1991; Farmer et al., 1988). Not surprisingly, some studies also report an association between exercise participation and self-rated health (Leinonen et al., 2005; Norman et al., 2002; Rutten et al., 2001; Simonen et al., 2004). People who do not participate in exercise at a regular basis are more likely to evaluate their health as poor. Self-rated health can be described as a general self-evaluation of a person's health, reflecting both physical and mental aspects of health (Eriksson et al., 2001; Idler & Benyamini, 1997). Studies show that self-rated health measured with a single question is a very good predictor of morbidity and mortality (Burstrom & Fredlund, 2001; Idler & Benyamini, 1997). An additional advantage of the self-rated health measure is that it is easy to include in large-scale epidemiological survey studies at relatively low costs.

If the overall judgment of a person's health is negative this might be one of the causes of a sedentary lifestyle. On the other hand, a sedentary lifestyle could also cause a poor rated health, because a sedentary lifestyle increases the risk for true physical and/or mental health problems. These explanations are not mutually exclusive, in that there may also be a reciprocal causal relationship between exercise and self-rated health. Perceived lowering of health may reduce exercise frequency which in turn may further influence the perception of health. Although often overlooked, the association between exercise and self-rated health could also be explained

by a third underlying variable, such as common genetic or environmental factors that predispose to a sedentary lifestyle as well as poor self-rated health. It is, for instance, well-known that individual differences in adult exercise behavior are heritable (Beunen & Thomis, 1999; Kujala et al., 2002; Maia et al., 2002; Simonen et al., 2004) and there is increasing evidence that individual differences in self-rated health may also be heritable (Svedberg et al., 2001; Svedberg et al., 2005). Hence, a common set of genes may influence both traits (pleiotropy).

As far as we know, there exists only one study that investigated the etiology of the association between exercise participation and self-rated health using a genetic approach (Simonen et al., 2004). In a small sample of 300 male twin pairs aged 35-70 years, lifetime exercise and self-rated health were significantly related (odds ratio 1.62) and this association could fully be explained by the overlap of genetic factors influencing both phenotypes. It is not known, however, whether these results also hold for females and younger adults.

The purpose of the present study, using structural equation modeling techniques, is to investigate 1) whether there is an association between exercise participation and self-rated health in a large sample of young and middle-aged twins and their siblings and 2) whether the association between exercise participation and self-rated health can be explained by common genes, other familial factors (i.e., shared environmental influences) or unique environmental factors.

Methods

Subjects

This study is part of an on-going study on lifestyle and health in twin families that are voluntarily registered with the Netherlands Twin Registry (NTR) (Boomsma et al., 2002). Since 1991, every two to three years participants receive questionnaires on lifestyle and health (i.e., health status, exercise participation, smoking behavior, alcohol use, and personality). In this study, we focus on data on exercise participation and self-rated health obtained from twins and their siblings collected in 2002. In total, 5,950 twins and siblings participated in the 2002 survey. Twins and siblings aged 18-50 years were selected (N=5,200). We excluded twins with unknown zygosity (N=23), subjects for whom both data on exercise participation and self-rated health was missing (N=3) and genetically unrelated and half siblings (N=25). In large families, only the first two brothers and the

first two sisters of a twin pair were selected. This excluded 4 brothers and 5 sisters from large families. The final sample consisted of 5,140 twins and siblings from 2,831 families. There were 3,950 twins, 465 brothers and 725 sisters. There were 1,370 complete and 1,210 incomplete twin pairs. Complete twin pairs are pairs where measurements are available for both twins, whereas in incomplete twin pairs data are available for only one twin. Data from incomplete pairs were retained in the analyses to improve the estimation of the prevalences.

Zygoty of same-sex twins was determined by DNA typing for 27.5% of the same-sex twin pairs (Willemsen et al., 2005). For the other same-sex twin pairs zygoty was based on eight items on physical similarity and the frequency of confusion of the twins by parents, other family members and strangers. Agreement between zygoty based on these items and zygoty based on DNA was 97%. There were 366 monozygotic male (MZM), 223 dizygotic male (DZM), 859 monozygotic female (MZF), 493 dizygotic female (DZF) and 639 dizygotic opposite-sex (DOS) twin pairs (both complete and incomplete twin pairs). Sex and age of the participants was always known, exercise participation was missing for 16 participants (0.3%) and self-rated health was missing for 63 participants (1.2%). The mean age of all participants was 30.5 (SD=7.0).

Measures

Exercise participation was assessed with multiple questions. Firstly, the question 'Do you exercise regularly?' (Yes or No) was asked. If the answer was affirmative, further information on type of exercise, frequency and duration was gathered. MET scores (Metabolic equivalents) were assigned to the different types of sports using the Ainsworth's Compendium of physical activity (Ainsworth et al., 2000). One MET equals the rate of energy expended when at rest (1 kcal/kg/h). The respondents were classified in three distinct ordered categories. The first category consists of respondents who did not participate in sports (i.e. no MET-score could be assigned); these respondents were classified as sedentary. The second category consists of light and moderate exercisers. Light and moderate exercisers are respondents who are involved in some type of exercise but who do not meet the criterion for vigorous exercise (exercising at least 60 minutes weekly, two or more times a week, with an intensity of 6 METs or higher). The third category consists of respondents who meet the above criterion for vigorous exercise. The resulting measure of exercise participation (three categories: sedentaries, light or moderate exercisers and vigorous exercisers) was used in the analyses.

Self-rated health was assessed with the single question 'How, in general, is your health?' (Eriksson et al., 2001). Participants could respond with Bad, Poor, Fair, Good or Excellent. Because the prevalences of self-reported bad, poor or fair health were low, the data from the first three and the last two categories were pooled together. This resulted in a measure of self-rated health with two categories defined as Less than good and Good/Excellent health.

Statistical analyses

Structural equation modeling was used to estimate the correlation between exercise and self-rated health, the relative contribution of genetic factors to variation in exercise participation and self-rated health (i.e., the heritability), and the overlap between genetic and environmental factors that influence the two traits. Threshold models were fitted to the raw ordinal data using the software package Mx (Neale et al., 2003). The threshold model assumes that a categorical variable has an underlying liability with a continuous and standard normal distribution. Thresholds divide the liability distribution into discrete categories (e.g. 'Less than good health' and 'Good/Excellent health'). The thresholds are based on the prevalence of the different categories in the population. The resemblances between relatives for the liabilities of exercise and self-rated health are estimated with polychoric correlations.

The comparison of monozygotic (MZ) and dizygotic (DZ) twin correlations is a first step in evaluating the relative influences of genetic and environmental factors on variation in a trait. MZ twin pairs have all, or nearly all, genes in common. DZ twin pairs and siblings, share on average only half of their segregating genes. If the MZ correlation is higher than the DZ /sib correlation, genetic influences are implicated. If the DZ correlation is higher than half the MZ correlation, there is evidence for common environmental effects (C) shared by family members. If MZ correlations are lower than 1, this implies that unique environmental effects (E) play a role. Unique environmental effects are those environmental effects that are not shared between family members. Genetic effects can be additive (A); the summation of effects of multiple alleles at different loci) or can act in a non-additive manner. Non-additive genetic effects, or dominance (D), occur when there is interaction between alleles within a locus or across loci (epistasis). MZ twin pairs share all additive genetic, non-additive genetic and shared environmental effects, whereas DZ twin pairs and full siblings share on average half of their additive genetic effects, a quarter of their non-additive effects and all shared environmental effects.

The effects of C and D cannot be estimated simultaneously from data from twins and siblings and a choice for an ADE or ACE model has to be made based on the pattern of MZ – DZ correlations (C will make these correlations more alike, D will make them more different from each other).

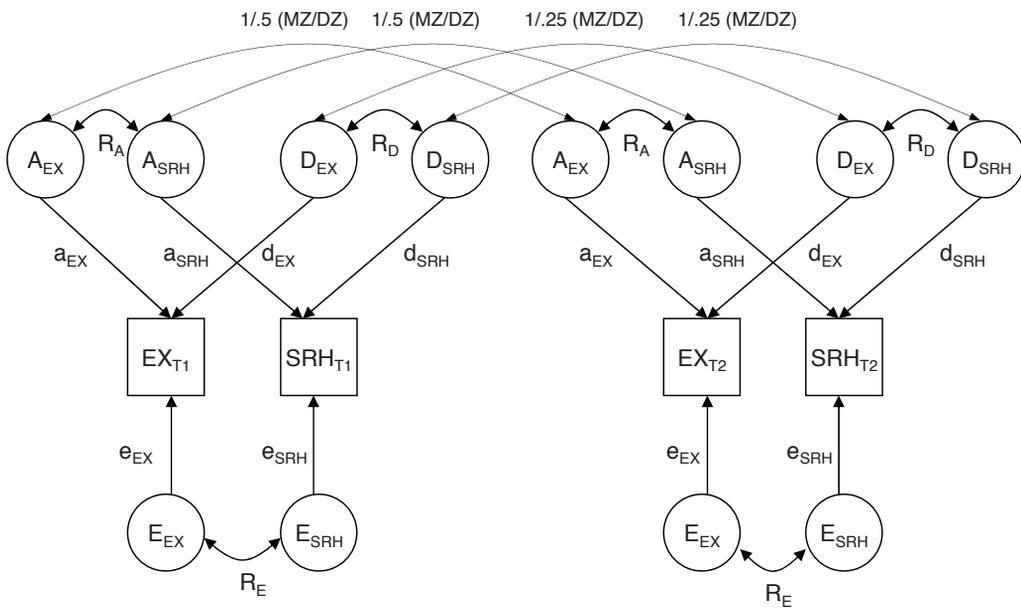
The comparison of cross-twin cross-trait correlations provides information on the etiology of the *association* between exercise participation and self-rated health. If the cross-twin cross-trait correlations are larger than zero, this suggests that the etiology of the correlation between exercise and self-rated health is familial. If the MZ cross-twin cross-trait correlation is larger than the DZ cross-correlation, this suggests that common genetic factors explain the correlation. If the MZ cross-twin cross-trait correlation equals the DZ cross-correlation, common shared environmental factors account for the correlation. Data from siblings were added to the design to increase the power to detect possible additive and non-additive genetic and shared environmental influences (Posthuma & Boomsma, 2000). Correlations for DZ twins, twin-sibling pairs and sibling-sibling pairs have the same expectations, because all those pairs share on average 50% of their genes and are brought up in the same family. Therefore, in subsequent models, non-twin siblings are treated as DZ twins.

A number of different models were fitted to the data and their fit was compared by means of the log-likelihood ratio test (LRT). The difference in minus two times the log-likelihood (-2LL) between two nested models has a χ^2 distribution and the degrees of freedom (df) equals the difference in df between the two models. If the χ^2 -test yielded a p-value higher than 0.01 the fit of the constrained model was not significantly worse than the fit of the more complex model and the constrained model was kept as the most parsimonious and best fitting model.

Firstly, in a bivariate saturated model, the thresholds, the correlation between traits, the twin correlations and the cross-twin cross-trait correlations were estimated. We tested for sex and zygosity effects on the thresholds and the correlations. Secondly, we fitted a series of genetic models. We started with an ADE model or an ACE model. The bivariate ADE model for exercise participation and self-rated health is depicted in Figure 3.1. In the ADE model, there are 7 free parameters (the proportions of variance due to A of both traits, the proportions of variance due to D of both traits and the additive genetic, non-additive genetic and unique environmental correlations). The proportions of variance due to E for both traits are not free parameters, since the total variance of the liability distribution is fixed at 1. Next, an AE model was fitted to the data (with 4 free parameters: two heritabilities, an additive genetic and a unique environmental

correlation) to test the significance of D (or C). Finally, we fitted a model with the genetic correlation fixed at zero (3 free parameters), a model with the unique environmental correlation fixed at zero (3 free parameters) and a model with both correlations at zero (2 free parameters) to test their significance. In all these bivariate genetic models, three thresholds were estimated (2 for exercise participation and 1 for self-rated health).

Figure 3.1. Bivariate ADE model for exercise participation and self-rated health



EX_{T1}: Exercise participation of twin 1, SRH_{T1}: Self-rated health of twin 1, EX_{T2}: Exercise participation of twin 2, SRH_{T2}: Self-rated health of twin 2, A_{EX}: Additive genetic factor for exercise participation, A_{SRH}: Additive genetic factor specific for self-rated health, D_{EX}: Non-additive genetic factor for exercise participation, D_{SRH}: Non-additive genetic factor for self-rated health, E_{EX}: Unique environmental factor for exercise participation, E_{SRH}: Unique environmental factor for self-rated health, a_{EX}: Square root of the proportion of variance due to A of exercise participation, a_{SRH}: Square root of the proportion of variance due to A of self-rated health, d_{EX}: Square root of the proportion of variance due to D of exercise participation, d_{SRH}: Square root of the proportion of variance due to D of self-rated health, e_{EX}: Square root of the proportion of variance due to E of exercise participation, e_{SRH}: Square root of the proportion of variance due to E of self-rated health, R_A: Additive genetic correlation between exercise participation and self-rated health, R_D: Non-additive genetic correlation between exercise participation and self-rated health, R_E: Unique environmental correlation between exercise participation and self-rated health.

Results

The prevalence of no, light/moderate or vigorous exercise participation (represented by the thresholds in the saturated model) was different in males and females ($\chi^2=65.08$, $\Delta df=4$, $p<0.001$), but not different between MZ and DZ twins ($\chi^2=9.02$, $\Delta df=4$, $p=0.06$). As much as 39% of the sample is not involved in any kind of exercise, 44% exercises at a light to moderate level and only 17% is regularly participating in vigorous exercise activities. The sex difference in exercise participation is found in the percentage of vigorous exercisers: 23% of the males are involved in regular vigorous exercise activities, compared with 14% of the females. The prevalence of self-rated health was not significantly different between MZ and DZ pairs ($\chi^2=8.10$, $\Delta df=2$, $p=0.02$), but females reported more often less than good health (12%) than males (7%) ($\chi^2=37.55$, $\Delta df=2$, $p<0.001$).

The correlation between exercise participation and self-rated health was 0.20 (95% CI: 0.15-0.22) and did not significantly differ for males and females ($\chi^2=1.42$, $\Delta df=1$, $p=0.23$). The estimates of the twin correlations for exercise participation and self-rated health and the cross-twin cross-trait correlations are presented in Table 3.1. The twin correlations of exercise participation were equal in males and females ($\chi^2=2.02$, $\Delta df=3$, $p=0.57$). MZ twin correlations were significantly larger than DZ correlations ($\chi^2=33.84$, $\Delta df=2$, $p<0.001$). The twin correlations of self-rated health were equal in males and females ($\chi^2=1.21$, $\Delta df=3$, $p=0.75$). MZ correlations were significantly larger than the DZ correlations ($\chi^2=12.72$, $\Delta df=2$, $p=0.002$). There were no significant differences across sex in the cross-correlations ($\chi^2=3.64$, $\Delta df=3$, $p=0.30$). The twin correlations for exercise and self-rated health suggest that additive and non-additive genetic factors play a role. Therefore, we next fitted an ADE model.

Table 3.1. Twin and cross-twin cross-trait correlations for exercise participation and self-rated health

	MZM	DZM*	MZF	DZF*	DOS*	MZ	DZ*
Twin correlations for exercise participation							
r	0.59	0.25	0.55	0.28	0.21	0.56	0.24
CI	0.45;0.69	0.11;0.38	0.46;0.62	0.19;0.36	0.13;0.28	0.49;0.62	0.19;0.29
N	203	352	552	911	1,043	755	2,306
Twin correlations for self-rated health							
r	0.64	0.03	0.58	0.19	0.24	0.58	0.19
CI	0.02;0.91	-0.28;0.38	0.42;0.72	-0.00;0.38	0.05;0.41	0.43;0.72	0.07;0.31
N	201	346	540	898	1,021	741	2,265
Cross-twin cross-trait correlations for exercise participation and self-rated health							
r	0.08	0.13	0.15	0.13	0.03	0.15	0.08
CI	-0.22;0.32	-0.04;0.29	0.05;0.24	0.04;0.23	-0.07;0.11	0.05;0.23	0.02;0.14
N	404	699	1,092	1,808	2,064	1,496	4,571

r: Polychoric correlations, CI: 95% confidence interval, N: Number of twin pairs, MZM: Monozygotic male twin pairs, DZM: Dizygotic male twin pairs, MZF: Monozygotic female twin pairs, DZF: Dizygotic female twin pairs, DOS: Dizygotic opposite sex twin pairs, MZ: Monozygotic twin pairs pooled over sex, DZ: Dizygotic twin pairs pooled over sex, * Note that DZ pairs also include siblings

The results of the tests for the significance D and the genetic and environmental correlations are given in Table 3.2. The most parsimonious and best fitting model is an AE model with a genetic correlation but no environmental correlation. In this model, 54.4% (95% CI: 47.9 - 58.3) of the variance in liability of exercise participation is explained by additive genetic factors. The remaining 45.6% (95% CI: 40.0 - 52.0) is entirely explained by unique environmental factors. For self-rated health, 53.8% (95% CI: 39.4 - 66.1) of the variance in liability of self-rated health is explained by additive genetic factors and 46.2% by unique environmental factors (95% CI: 43.0 - 60.6). The genetic correlation is 0.36, which means that the additive genetic factors influencing exercise participation and self-rated health show moderate overlap. These overlapping additive genetic factors explain the entire correlation between exercise participation and self-rated health (0.20); no contribution of the unique environment to this correlation is found.

Table 3.2. Bivariate model fitting results and parameter estimates for exercise participation and self-rated health

Model fitting results						
Model	vs.	-2LL	df	χ^2	Δ df	p
1. Saturated	–	13462.72	10172	–	–	–
2. ADE	1	13472.13	10184	9.31	12	0.68
3. AE, r_A , r_E	1	13475.41	10187	12.59	15	0.63
4. AE, r_E	3	13490.26	10188	14.85	1	<0.01
5. AE, r_A	3	13476.75	10188	1.34	1	0.25
6. AE	3	13529.09	10189	53.68	2	<0.01

Parameter estimates									
Model	Exercise participation					Self-rated health			
	a^2	d^2	e^2	a^2	d^2	e^2	r_A	r_D	r_E
1. Saturated	–	–	–	–	–	–	–	–	–
2. ADE	39.3	16.9	43.8	26.9	31.0	42.2	0.67	-0.31	0.13
3. AE, r_A , r_E	54.1	–	–	53.5	–	–	0.30	–	0.10
4. AE, r_E	52.8	–	–	50.8	–	–	–	–	0.00
5. AE, r_A	54.4	–	45.6	53.8	–	46.2	0.36	–	–
6. AE	54.1	–	–	54.8	–	–	–	–	–

Bold: most parsimonious and best fitting model, vs.: versus, -2LL: -2 log-likelihood, df: degrees of freedom, χ^2 : chi-square test statistic, Δ df: degrees of freedom of chi-square test, p: p-value, a^2 : proportion of variance due to additive genetic factors of exercise participation/self-rated health, d^2 : proportion of variance due to non-additive genetic factors of exercise participation/self-rated health, e^2 : proportion of variance due to unique environmental factors of exercise participation/self-rated health, r_A : additive genetic correlation, r_D : non-additive genetic correlation, r_E : unique environmental correlation

Discussion

The main finding in this study is that exercise participation and self-rated health are modestly related ($r=0.20$), and this association can be explained by a set of common genes predisposing to both exercise participation and self-rated health. Individual differences in both exercise participation (heritability 54.4%) and self-rated health (heritability 53.8%) are explained by genetic influences and unique environmental influences. The association between exercise participation and self-rated health can be fully explained by overlapping genetic factors.

The finding of a modest correlation between exercise and self-rated health is consistent with two other studies reporting a correlation of 0.20 in adults aged 18 years and older (Rutten et al., 2001) and of 0.36 in elderly females (Leinonen et al., 2005). Our finding that common genes account for the association between exercise participation and self-rated health is in line with the single previous study that investigated the nature of this association in terms of common genetic and/or environmental influences (Simonen et al., 2004). They found that common genes accounted for the association between lifetime exercise and self-rated health in a small sample of 300 male twin pairs aging 35-70 years. The present study extends these findings by showing that common genes also explain the association between exercise participation and self-rated health in young adulthood and females.

The importance of common genetic influences on exercise participation and self-rated health in adulthood can be interpreted in different ways. A first scenario is that there is a set of genes that is expressed in different parts of the body (i.e., in muscle tissues, the brain or different organs) or that has different functions (i.e., the same genes coding for different proteins, or post-expression protein processing), which might independently cause individual differences in exercise behavior and individual differences in self-rated health. For example, a recent study (Moran et al., 2006) showed that genetic variation in the Angiotensin I-Converting Enzyme (ACE) gene, a gene known to have different functions, such as the regulation of blood pressure and cell growth, is also associated with physical activity. It is not unthinkable that a gene with such a diversity of body functions might also influence self-rated health, through its influence on physical and possibly also mental health.

On the other hand, our finding of a common genetic predisposition does *not* exclude the existence of causality between exercise and self-rated health. It could be, for example, that genes that directly influence exercise

participation indirectly influence self-rated health, because exercise exerts a causal influence on self-rated health. In this second, causal, scenario, an individual's genetic predisposition to be sedentary might for instance cause a poor self-rated health. This may come about through an actual decrease in physical health and well-being caused by a sedentary lifestyle, or because people weigh the absence of any exercise behavior in the evaluation of their health. In a third, reverse causal, scenario, genes influencing mental or physical health may also influence exercise behavior. If a person judges his or her health to be low, for instance, this is known to constitute a barrier for (continuation of) exercise behavior (King et al., 1992; Sallis et al., 1986).

A good future strategy to discriminate between these scenarios would be to identify the actual genetic variation influencing exercise participation. If the genetic variants associated with exercise behavior also influence self-rated health and do so independently of exercise behavior, that is the association is of comparable strength in sedentaries, moderate and vigorous exercisers, this argues in favor of genetic pleiotropy. If a causal mechanism is driving the association, for example when exercise causes improved self-rated health, the genetic variants associated with exercise behavior will no longer be associated with self-rated health when tested within the separate groups of sedentary, moderate and vigorous exercisers.



4

Testing causality in the association between regular exercise and symptoms of anxiety and depression

This chapter is published as:

De Moor, M.H.M., Boomsma, D.I., Stubbe, J.H., Willemsen, G., & De Geus, E.J.C. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of General Psychiatry*, 65 (8), 897-905.

Abstract

Context In the population at large, regular exercise is associated with reduced anxious and depressive symptoms. Results of experimental studies in clinical populations suggest a causal effect of exercise on anxiety and depression, but it is unclear whether such a causal effect also drives the population association. We cannot exclude the major contribution of a third underlying factor influencing exercise behavior and symptoms of anxiety and depression.

Objective To test causal effects of exercise on anxious and depressive symptoms in a population-based sample.

Design Population-based, longitudinal study (1991-2002) in a genetically informative sample of twin families.

Setting Causal effects of exercise were tested by bivariate genetic modeling of the association between exercise and symptoms of anxiety and depression, correlation of intrapair differences in these traits among genetically identical twins, and longitudinal modeling of changes in exercise and anxious and depressive symptoms.

Participants A total of 5,952 twins from the Netherlands Twin Register, 1,357 additional siblings, and 1,249 parents. All participants were aged 18 to 50 years.

Main outcome measurements Survey data about leisure-time exercise (metabolic equivalent task hours per week based on type, frequency and duration of exercise) and four scales of anxious and depressive symptoms (depression, anxiety, somatic anxiety and neuroticism, plus a composite score).

Results Cross-sectional and longitudinal associations were small and were best explained by common genetic factors with opposite effects on exercise behavior and symptoms of anxiety and depression. In genetically identical twin pairs, the twin who exercised more did not display fewer anxious and depressive symptoms than the co-twin who exercised less. Longitudinal analyses showed that increases in exercise participation did not predict decreases in anxious and depressive symptoms.

Conclusion Regular exercise is associated with reduced anxious and depressive symptoms in the population at large, but the association is not because of causal effects of exercise.

Epidemiological studies consistently report a small to moderate cross-sectional association between lack of exercise and depressive symptoms in population-based samples (Camacho et al., 1991; De Moor et al., 2006; Farmer et al., 1988; Kritz-Silverstein et al., 2001; Mobily et al., 1996; Norris et al., 1992; Stephens, 1988; Steptoe et al., 1997; Strawbridge et al., 2002; van Gool et al., 2003). The association holds using various measures of exercise behavior and using either the Beck's Depression Inventory (BDI) (De Moor et al., 2006; Kritz-Silverstein et al., 2001; Steptoe et al., 1997) or the Center for Epidemiological Studies Depression Scale (CES-D) (Farmer et al., 1988; Mobily et al., 1996; Stephens, 1988; van Gool et al., 2003) as a measure of depressive symptoms. Lack of regular exercise is also associated with high levels of anxiety and neuroticism measured using various instruments (De Moor et al., 2006; Rhodes & Smith, 2006; Stephens, 1988). Longitudinal studies further show that exercise at baseline predicts fewer depressive symptoms at follow-up. Of nine studies (Brown et al., 2005; Camacho et al., 1991; Cooper-Patrick et al., 1997; Farmer et al., 1988; Kritz-Silverstein et al., 2001; Mobily et al., 1996; Strawbridge et al., 2002; van Gool et al., 2003; Wise et al., 2006) six studies (Brown et al., 2005; Camacho et al., 1991; Mobily et al., 1996; Strawbridge et al., 2002; van Gool et al., 2003; Wise et al., 2006) found a prospective association. One study (Farmer et al., 1988) observed an effect only in white women, and another study (Weyerer, 1992) that examined the association between clinical diagnoses of depressive disorders and exercise behavior in the population found evidence for a cross-sectional but not for a prospective association.

Experimental studies investigating the effect of exercise training on symptoms of clinical depression and anxiety are numerous, and many reviews (Brosse et al., 2002; Byrne & Byrne, 1993; Folkins & Sime, 1981; Gauvin & Spence, 1996; Salmon, 2001; Scully et al., 1998; Steptoe, 2006) and meta-analyses (Craft & Landers, 1998; Lawlor & Hopker, 2001; Long & Vanstavel, 1995; North et al., 1990) have been published. These reviews and meta-analyses often differ in their inclusion criteria with regard to measurement of depression or anxiety (self-reported symptoms or clinical diagnosis), sample characteristics (healthy participants or clinical population) and research design (differences in control groups, randomization and period of training and follow-up). Nevertheless, most authors of these reviews conclude that exercise training seems to relieve symptoms of anxiety and depression. Furthermore, some randomized controlled trials suggest that exercise training reduces depressive symptoms, with effect sizes comparable to those of antidepressant use (Blumenthal et al., 2007).

The evidence from these prospective and experimental studies makes it tempting to interpret the association at the population level as reflecting a causal effect of exercise on the symptoms of anxiety and depression. This explanation fits folk wisdom. However, results from experimental studies may not always generalize to the population at large. There may be ascertainment in these studies such that only subjects attracted to exercise may enroll and persist. Treatment effects in psychiatric patients may not generalize to the extent that exercise relieves mild forms of symptoms of anxiety and depression in healthy participants. Finally, although findings of prospective studies are compelling, they are based on correlations between traits that are time-lagged. We cannot rule out that some underlying factors that influence exercise behavior at one time point also influence symptoms of anxiety and depression at a later time point. Genetic variation among individuals is such a potential underlying factor. Exercise behavior and symptoms of anxiety and depression are heritable traits, with genetic factors explaining about 40 to 50% of the variation in anxious and depressive symptoms (Boomsma et al., 2000) and about 50 to 60% of the variation in exercise behavior (De Moor et al., 2007a; Stubbe et al., 2006a). Some genetic factors influencing exercise behavior may overlap with genetic factors influencing anxious and depressive symptoms.

The objective of the present study was to test the causal effects of exercise on reducing anxious and depressive symptoms, taking this potential genetic confounding into account. To this end, we used longitudinal data from genetically informative individuals (twins and their family members) that allows a strong nonexperimental design to test the causal

hypothesis in a population-based sample. Specifically, we tested four predictions generated by the causal hypothesis.

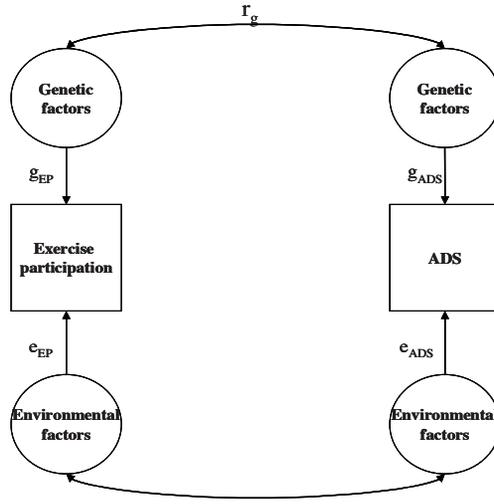
A first prediction is that, if exercise causally influences symptoms of anxiety and depression, all genetic and environmental factors that influence exercise behavior will also, through the causal chain, influence these symptoms. We can test this in a bivariate model using identical (i.e., monozygotic [MZ]) and fraternal (i.e., dizygotic [DZ]) twins by computing the genetic and environmental correlations between two traits (Neale & Cardon, 1992). Under the causal hypothesis, genetic and environmental correlations should be significant (Figure 4.1a), whereas a significant genetic correlation in the absence of a significant environmental correlation falsifies the hypothesized causal effect of exercise.

A second prediction made by the causal hypothesis is that in genetically identical (MZ) twins the within-twin pair differences in levels of exercise participation should be associated with within-twin pair differences in anxious and depressive symptoms. This means that the twin who participates in more exercise should display fewer anxious and depressive symptoms than the co-twin who exercises less (Figure 4.1b). The absence of this relationship between the within-pair differences in exercise participation and symptoms of anxiety and depression falsifies the hypothesized causal effect of exercise, whereas the presence of this relationship would argue in favor of the causal hypothesis because it excludes confounding by genetic factors (the twins are genetically identical).

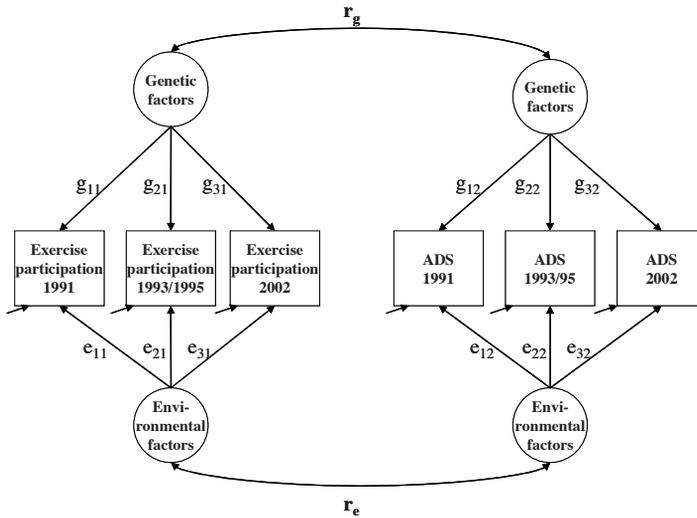
A third prediction made by the causal hypothesis is that, if exercise reduces symptoms of anxiety and depression, exercise at baseline should predict anxious and depressive symptoms longitudinally, and this should be independent of genetic factors influencing exercise behavior and symptoms of anxiety and depression. This prediction can be tested in a twin sample by computing the genetic and environmental correlations between exercise and symptoms of anxiety and depression at successive time points (Figure 4.1c). Under the causal hypothesis, longitudinal genetic and environmental correlations should be significant, whereas a significant longitudinal genetic correlation in the absence of a significant environmental correlation falsifies the hypothesized causal effect of exercise.

Figure 4.1. Graphical representation of four models used to test the hypothesis that exercise causes reduced symptoms of anxiety and depression.

a. Bivariate genetic model



c. Longitudinal correlation between MET hours and symptoms of anxiety and depression

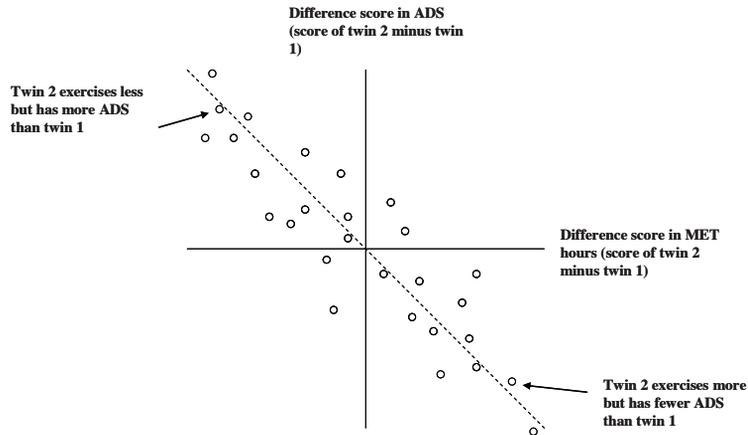


A: A causal effect of exercise predicts significant overlap (i.e., correlation) in both the genetic and environmental factors influencing exercise behavior and anxious depressive symptoms (ADS). Cross-sectional correlation between exercise and anxious depressive symptoms = $(g_{EP} * r_g * g_{ADS}) + (e_{EP} * r_e * e_{ADS})$.

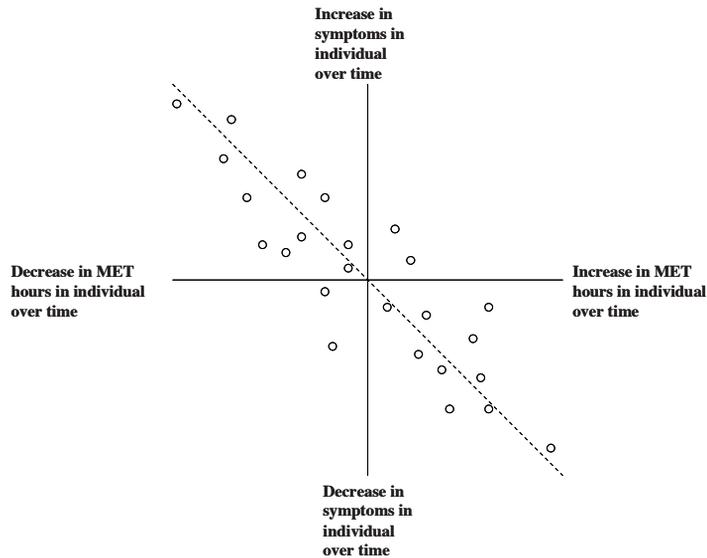
C: A causal effect of exercise predicts significant overlap (i.e., correlation) between genetic and environmental factors influencing exercise behavior at an early time point and ADS at a later time point. For example, longitudinal correlation between exercise in 1991 and ADS in 2002 = $(g_{11} * r_g * g_{32}) + (e_{11} * r_e * e_{32})$.

Figure 4.1. Graphical representation of four models used to test the hypothesis that exercise causes reduced symptoms of anxiety and depression.

b. MZ intra-pair differences model



d. Longitudinal correlation of difference scores



B: A causal effect of exercise predicts that in genetically identical twins, the twin who exercises more has fewer ADS.

D: A causal effect of exercise predicts that an increase in exercise leads to a decrease in ADS in an individual over time. E indicates environmental factor loading; EP, exercise participation; g, genetic factor loading; MET, metabolic equivalent task; MZ, monozygotic; r_e , environmental correlation; and r_g , genetic correlation.

A fourth prediction by the causal hypothesis is that within-subject changes in exercise behavior should predict parallel changes in anxious and depressive symptoms such that increases in frequency and intensity of exercise behavior over time would reduce symptoms of anxiety and depression and that, vice versa, decreases in frequency and intensity of exercise behavior over time would increase symptoms of anxiety and depression (Figure 4.1d). Such a pattern would argue in favor of the causal hypothesis because it excludes confounding by a genetic factor (the subjects' genotypes do not change over time). In contrast, the absence of a correlation between changes in exercise and symptoms over time would falsify the hypothesized causal effect of exercise.

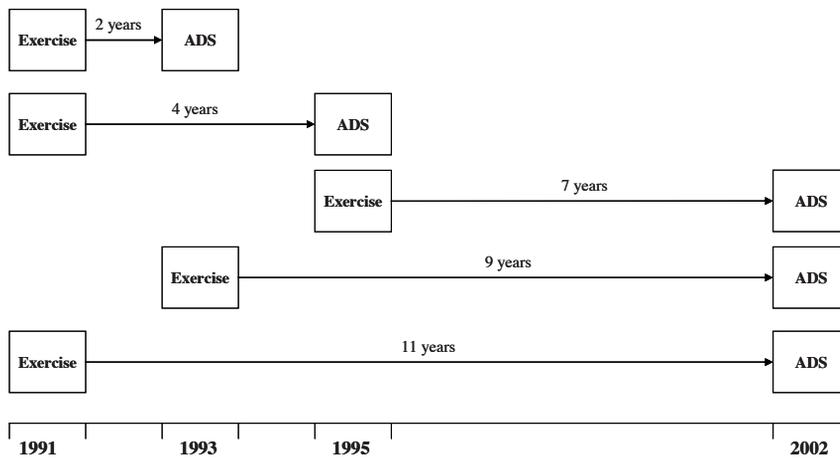
Methods

Participants

This study was part of an ongoing study on health, lifestyle and personality in adult twins and their relatives who are voluntarily registered at the Netherlands Twin Register (Boomsma et al., 2002; Boomsma et al., 2006b). Since 1991, every two to three years the participants receive a mailed questionnaire with questions about their health, lifestyle, and personality. Data about exercise behavior (type, frequency and duration) and symptoms of anxiety and depression were available for 1991, 1993, 1995 and 2002 (Figure 4.2). We included individuals within the age range of 18 to 50 years (i.e., at no time point the age of the participants was below 18 or above 50). The mean age (SD) was 27.9 (8.0) years. We excluded 37 twins with unknown zygosity and 47 genetically unrelated siblings and half siblings. In 1991 and 1993, twins and their parents were invited to participate, and data on exercise participation and three measures of anxious and depressive symptoms (anxiety, somatic anxiety and neuroticism scales, described herein) were collected. In 1991, twins also completed the Young Adult Self Report (YASR) (Achenbach, 1990) depressive symptoms scale. In 1995, twins, siblings and parents were invited, but only twins filled in questions about depressive symptoms (YASR scale). In 2002, twins, siblings, parents and spouses provided data on exercise participation, and four measures of anxious and depressive symptoms were collected. For the longitudinal analyses, only data from twins and parents were used. There were 15,961 twins and parents from 4,496 families who participated at least once in 1991, 1993, 1995 or 2002. Of these, 8,662 individuals from 2,288 families participated at two or more time points. Individuals who participated in

more than one survey were, on average, less anxious and depressed and more physically active than individuals who participated only once, but the differences are small (data not shown). Similar findings were reported in a study (Vink et al., 2004) in which it was shown that individuals from highly cooperative families are slightly less anxious and depressed but not more frequent exercisers compared with individuals from less cooperative families.

Figure 4.2. Overview of the longitudinal data on exercise participation and symptoms of anxiety and depression in the Netherlands Twin Registry



Note: ADS indicates anxious depressive symptoms.

Zygoty was determined by DNA typing for 33.2% of the same-sex twin pairs. For the other same-sex twins, zygoty was based on eight items regarding physical similarity and the frequency of confusion between the twins by parents, other family members, and strangers. Agreement between zygoty based on these eight items and zygoty based on DNA typing was 97% (Willemsen et al., 2005).

For the cross-sectional analyses, the most recent data from complete twin pairs and a maximum of one additional brother and one sister were selected. For all measures, data from 2002 were selected first. For all measures except the YASR scale, we next selected data from 1993 and then from 1991. For the YASR scale, we next selected data from 1995 and then from 1991. After this selection of complete pairs, the most recent data of incomplete twin pairs were selected. This ensured a maximum of complete twin pairs in the dataset and resulted in a cross-sectional dataset of 941 MZ male, 656 DZ male, 1,802 MZ female, 1,067 DZ female, 1,486 DZ opposite-sex twins, 586 brothers and 771 sisters (from 3,619 families). Of these, there were 2,547 complete twin pairs (406 MZ male, 281 DZ male, 789 MZ female, 455 DZ female and 616 DZ opposite-sex pairs).

For the longitudinal analyses, there were 1,878 twins and parents with valid data about exercise behaviour and symptoms of anxiety and depression in both 1991 and 1993, 584 twins and parents with valid data in both 1991 and 2002, and 1,037 with data in both 1993 and 2002. There were 534 twins with data in both 1991 and 1995 and 1,166 with data in both in 1995 and 2002.

Measurements

Leisure-time exercise was measured by several questions. The first question asked whether the respondent participated in exercise regularly and could be answered with yes or no. If the participants responded in the affirmative, further information about type, frequency and duration of exercise was gathered. Reported non-leisure-time activities such as walking or biking to work, were not counted as exercise. All remaining exercise activities were assigned a metabolic equivalent task (MET) value, using the Compendium of physical activity by Ainsworth et al. (2000). A MET score of 1 corresponds to the rate of energy expenditure when at rest (1 kcal/kg/h). MET hours were computed as MET multiplied by hours per week. Scores of nonexercising individuals were coded as zero. The distribution of MET hours was highly skewed (2.67) and demonstrated kurtosis (12.18). Therefore, for the computation of correlations and in the bivariate genetic models, log transformation was applied to the MET hours, which significantly reduced the skewness and kurtosis to 0.47 and -1.06, respectively. For computation of the differences scores, we used the untransformed MET hour scores because the difference scores were normally distributed. The test-retest reliability of MET hours was 0.82 as computed among 200 individuals who completed the questions about exercise in December 2004 and again 6 months later.

Anxious and depressive symptoms were measured using four continuous scales. Middeldorp and colleagues (2006) demonstrated that these measures are good predictors for the DSM-IV-based disorders of major depressive disorder, dysthymia, generalized anxiety disorder, panic disorders and social phobia. Depressive symptoms were measured using the YASR scale (16 items in 3 categories). Anxious symptoms were measured using the Spielberger State-Trait Anxiety Inventory trait version (20 items on a 4-point Likert scale) (Spielberger et al., 1970). Neuroticism and somatic anxiety were measured using the Amsterdamse Biografische Vragenlijst (Wilde, 1970), a Dutch questionnaire that is similar in content to the Eysenck Personality Questionnaire by Eysenck and Eysenck (1964). The neuroticism scale consists of 30 items and the somatic anxiety scale of 17 items (3 answering categories). The somatic anxious symptom scale contains items about somatic complaints that are related to anxiety and depression such as headaches and sleep problems. Summed scores of each of these measures were computed. For correlations and bivariate genetic modeling, the depression, anxiety, somatic anxiety, and neuroticism scales were transformed before analysis to improve normality according to the recommendations by Boomsma et al. (2000) using the following formulas: 5 times the natural logarithm (neuroticism), 12 times the natural logarithm (depression), 10 times the natural logarithm (anxiety) and 9 times the natural logarithm (somatic anxiety). The difference scores of the untransformed data regarding anxious and depressive symptoms were normally distributed. We further included a common factor score variable of all measures of anxious and depressive symptoms in the analyses. This common factor score variable was derived in a previous study (Boomsma et al., 2000) showing that the genetic variation in these measures derives from common genetic factors.

Statistical analysis

The first two tests of the causal hypothesis (the bivariate genetic model and the MZ twin intrapair differences model) were based on the cross-sectional data only. The bivariate genetic model tested whether an association between exercise and symptoms could be explained by an overlap in latent genetic or environmental factors that influence these traits. Similar to decomposition of the trait variance in a univariate genetic model in which MZ and DZ cross-twin or sibling correlations are compared (Neale & Cardon, 1992), MZ and DZ cross-twin or sibling cross-trait correlations are compared to evaluate whether covariance between two traits can be explained by overlapping additive genetic factors or by overlapping

unique environmental factors. If the MZ cross-twin cross-trait correlation is approximately twice the DZ cross-twin or sibling cross-trait correlation, overlapping additive genetic factors are implied. If the MZ cross-twin cross-trait correlation is smaller than the within-person cross-trait correlation, overlapping environmental factors that are unique to the individual are indicated.

Maximum likelihood estimation (Mplus; Muthén & Muthén, 2007) was used to obtain genetic and environmental correlations (Neale & Cardon, 1992). The hypothesis of a causal effect was tested by constraining either the genetic or the environmental correlations to zero. If this constraint is allowed (i.e., the genetic or the environmental correlation between exercise and symptoms of anxiety and depression is zero), a causal effect cannot be the source of their association. This genetic model also allowed us to test whether only the genetic correlation was significant. This would be compatible with the hypothesis of a common genetic factor underlying the cross-sectional association.

Using the MZ twin intrapair differences model, we computed the differences of all measures of anxious and depressive symptoms of an MZ twin and his or her co-twin. These intrapair differences in symptoms were then regressed on the difference in MET hours. Significant (and negative) regression coefficients would be compatible with the causal hypothesis, whereas nonsignificant regression analysis would falsify this hypothesis.

The third and fourth tests of the causal hypothesis were based on the longitudinal dataset. We computed the longitudinal correlation between MET hours and symptoms of anxiety and depression for 2-, 4-, 7-, 9- and 11-year intervals. Because individuals in the longitudinal dataset were not independent observations but were clustered in families, we used the COMPLEX option in combination with the robust maximum likelihood estimator in Mplus (Muthen & Satorra, 1995), which successfully corrects for bias in standard errors and chi-square test statistics due to this dependency (Rebollo et al., 2006). If exercise causes decreased anxious and depressive symptoms in the population, there should be a significant longitudinal association between MET hours at baseline and subsequent symptoms of anxiety and depression. If longitudinal correlations are nonsignificant, the hypothesis of a causal effect is falsified. However, the reverse is not true. A significant longitudinal correlation does not necessarily provide evidence for causality if common genetic factors drive the correlation between MET hours at baseline and symptoms of anxiety and depression at follow-up. We again used the genetic information in our sample to more robustly test causality using the multivariate model shown in Figure

4.1.c. Because of the complexity of the model, this model was fitted to the data in Mx (Neale et al., 2006). As with the cross-sectional data, we tested for a causal effect by constraining the genetic or environmental correlation between exercise and symptoms of anxiety and depression to zero. In this case, the genetic and environmental factors were not based on a single cross-sectional measurement but rather on all available successive measurements of exercise and symptoms of anxiety and depression. If the genetic or the environmental correlation between exercise and symptoms can be constrained to zero, a causal effect of exercise cannot be the source of the longitudinal or the cross-sectional association. As in the cross-sectional analysis, this genetic model also allowed us to test whether only the genetic correlation was significant. This would be compatible with the hypothesis of a common genetic factor underlying the longitudinal association.

The fourth test of causality was based on the longitudinal correlation of difference scores. If commencement of an exercise regimen causes reductions in anxious and depressive symptoms among the population, an increase in weekly MET hours should predict a decrease in anxious and depressive symptoms over time. Likewise, cessation of regular exercise should be associated with an increase in anxious and depressive symptoms. Regression analyses were performed in Mplus that predicted within-subject changes in symptoms by changes in MET hours. The COMPLEX option and robust maximum likelihood estimator were used to correct for bias due to family clustering of data. To account for a dependency of changes in symptoms on MET hours and symptoms at baseline, we included these as additional predictors in the model.

Results

Cross-sectional associations

The correlations between MET hours and symptoms of anxiety and depression are given in Table 4.1 together with the cross-sibling cross-trait correlations. The within-person cross-trait correlations between MET hours and anxious and depressive symptoms were small but significant, ranging from -0.06 to -0.14 , demonstrating that increased exercise behaviour is cross-sectionally associated with decreased anxious and depressive symptoms. The associations were somewhat stronger in women.

Table 4.1. Within-person cross-trait and cross-sibling cross-trait correlation between metabolic equivalent (MET) hours and symptoms of anxiety and depression

MET hours with:	Depressive symptoms (YASR)			Anxious symptoms (STAI)		
	N	r	95% CI	N	r	95% CI
Within-person cross-trait correlations						
All	6,187	-0.09*	-0.12 ; -0.07	6,391	-0.10*	-0.12 ; -0.07
Men	2,377	-0.09*	-0.13 ; -0.05	2,414	-0.06*	-0.11 ; -0.02
Women	3,810	-0.10*	-0.13 ; -0.06	3,977	-0.12*	-0.15 ; -0.09
Cross-sib cross-trait correlations						
MZM	479	-0.10*	-0.16 ; -0.04	511	-0.07*	-0.13 ; -0.01
DZM	2,525	-0.08*	-0.14 ; -0.01	2,532	-0.03	-0.10 ; 0.04
MZF	943	0.11*	-0.15 ; -0.07	983	-0.08*	-0.13 ; -0.04
DZF	4,149	-0.05*	-0.11 ; -0.01	4,237	-0.04	-0.09 ; 0.01
DOS	4,940	-0.06*	-0.11 ; -0.02	5,104	-0.04	-0.09 ; -0.00

N=number of pairs, r =Pearson correlation, 95% CI = 95% confidence interval, MZM=Monozygotic male twin pairs; DZM=Dizygotic male twin and male sibling pairs; MZF=Monozygotic female twin pairs; DZF=Dizygotic female twin and female sibling pairs; DOS=Dizygotic opposite-sex twin and opposite-sex sibling pairs, *=significant based on 95% confidence interval

Continuation of Table 4.1.

MET hours with:	Somatic anxious symptoms (ABV)			Neuroticism (ABV)			Anxious depressive symptoms (FS)		
	N	r	95% CI	N	r	95% CI	N	r	95% CI
Within-person cross-trait correlations									
All	6,320	-0.11*	-0.14 ; -0.09	6,307	-0.09*	-0.12 ; -0.07	6,226	-0.12*	-0.14 ; -0.09
Men	2,393	-0.07*	-0.11 ; -0.03	2,388	-0.06*	-0.10 ; -0.02	2,346	-0.08*	-0.12 ; -0.04
Women	3,927	-0.14*	-0.17 ; -0.11	3,919	-0.11*	-0.14 ; -0.08	3,880	-0.14*	-0.17 ; -0.10
Cross-sib cross-trait correlations									
MZM	510	-0.07*	-0.13 ; -0.01	510	-0.06*	-0.12 ; -0.00	508	-0.09*	-0.14 ; -0.03
DZM	2,518	-0.03	-0.10 ; 0.04	2,513	-0.02	-0.08 ; 0.05	2,478	-0.04	-0.11 ; 0.03
MZF	982	-0.12*	-0.16 ; -0.08	982	-0.09*	-0.13 ; -0.05	981	-0.11*	-0.15 ; -0.07
DZF	4,315	-0.08*	-0.13 ; -0.03	4,208	-0.07*	-0.11 ; -0.02	4,172	-0.07*	-0.12 ; -0.02
DOS	5,076	-0.03	-0.08 ; 0.02	5,068	-0.05	-0.09 ; 0.00	5,021	-0.05	-0.10 ; -0.00

N=number of pairs, r=Pearson correlation, 95% CI = 95% confidence interval, MZM=Monozygotic male twin pairs; DZM=Dizygotic male twin and male sibling pairs; MZF=Monozygotic female twin pairs; DZF=Dizygotic female twin and female sibling pairs; DOS=Dizygotic opposite-sex twin and opposite-sex sibling pairs, *=significant based on 95% confidence interval

The cross-sibling cross-trait correlations in MZ twin pairs were significant and similar in size to the within-person cross-trait correlations, indicating that exercise behaviour in one twin is predictive of anxious and depressive symptoms in the genetically identical co-twin. In contrast, in DZ cross-twin or sibling pairs, who share only part of their genetic makeup, exercise behavior of one sibling is not predictive of anxious and depressive symptoms in the other sibling.

Bivariate genetic model

Estimates of the genetic correlations obtained from cross-sectional bivariate genetic models of MET hours and anxious and depressive symptoms in men and women are given in Table 4.2. Genetic correlations were significant and ranged from -0.16 and -0.24. No environmental correlations, ranging from -0.07 to 0.05, were significantly different from zero, suggesting that the association between exercise behaviour and symptoms of anxiety and depression is not explained by a causal effect. Instead, common (i.e., overlapping) genetic factors explain the cross-sectional association between MET hours and symptoms of anxiety and depression.

MZ twin intrapair difference method

Regression of the intrapair difference scores in anxious and depressive symptoms on the intrapair difference scores in MET hours ranged between -0.04 and 0.03 and were nonsignificant. Therefore, in genetically identical twin pairs, a twin who exercises more is not less depressed than his or her co-twin who exercises less, which does not support the hypothesis that exercises causes relief in anxious and depressive symptoms.

Table 4.2. Genetic correlations of MET hours with symptoms of anxiety and depression, estimated in the cross-sectional bivariate genetic models (Figure 4.1a)

MET hrs with:	Depressive symptoms (YASR)		Anxious symptoms (STAI)		Somatic anxious symptoms (ABV)		Neuroticism (ABV)		Anxious depressive sympt. (FS)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
$r_{g,M}$	-0.24*	-0.36 ; -0.12	-0.16*	-0.28 ; -0.04	-0.20*	-0.35 ; -0.05	-0.16*	-0.28 ; -0.07	-0.22*	-0.34 ; -0.10
$r_{g,F}$	-0.22*	-0.30 ; -0.13	-0.17*	-0.25 ; -0.08	-0.25*	-0.33 ; -0.16	-0.17*	-0.24 ; -0.09	-0.20*	-0.28 ; -0.12

95% CI=95% confidence interval, $r_{g,M}$ = additive genetic correlation in men, $r_{g,F}$ = additive genetic correlation in women, * = significant based on 95% confidence interval

Longitudinal associations

An overview of the longitudinal within-person correlations of MET hours at baseline and anxious and depressive symptoms at follow-up is given in Table 4.3. The longitudinal correlations are significant and range from -0.07 to -0.14. Therefore, they are comparable in magnitude to the cross-sectional within-person correlations given in Table 4.1.

Longitudinal genetic correlation between MET hours and symptoms of anxiety and depression

The genetic correlations between MET hours and symptoms of anxiety and depression, obtained from the longitudinal multivariate genetic model, are given in Table 4.4. They were significant and ranged from -0.21 to -0.40. None of the environmental correlations, ranging from -0.12 to 0.22, were significantly different from zero, suggesting that the longitudinal associations between exercise and symptoms of anxiety and depression are not explained by a causal effect. Instead, common (i.e., overlapping) genetic factors explain the longitudinal correlations between MET hours and symptoms of anxiety and depression.

Longitudinal correlation of difference scores

Regression analyses were used to test whether a change in MET hours in an individual over time predicted a parallel change in anxious and depressive symptoms. An increase (or, vice versa, a decrease) in MET hours did not predict a decrease (or increase) in anxious and depressive symptoms during intervals of 2, 4, 7, 9 and 11 years. The estimated correlations between the change scores ranged from -0.05 to 0.05 and were nonsignificant. These results suggest that there is no evidence for a causal effect of exercise on relief of anxious and depressive symptoms in the population.

Table 4.3. Longitudinal within-person correlation between MET hours (baseline) and measures of symptoms of anxiety and depression (follow-up)

MET hours with:	1991 – 1993/95** (two/four years)			1993/95**– 2002 (nine/seven years)			1991 – 2002 (eleven years)		
	N	r	95% CI	N	r	95% CI	N	r	95% CI
Depressive symptoms (YASR)	804	-0.10*	-0.11 ; -0.06	1166	-0.10*	-0.12 ; -0.06	534	-0.14	-0.06 ; 0.16
Anxious symptoms (STAI)	1864	-0.11*	-0.11 ; -0.07	1033	-0.10*	-0.12 ; -0.05	581	-0.12*	-0.14 ; -0.06
Neuroticism (ABV)	1861	-0.09*	-0.10 ; -0.06	1011	-0.08*	-0.11 ; -0.05	571	-0.13*	-0.15 ; -0.07
Somatic anxious symptoms (ABV)	1871	-0.07*	-0.10 ; -0.06	1017	-0.11*	-0.12 ; -0.06	571	-0.13*	-0.17 ; -0.08
Anxious depressive symptoms (factor score)	725***	-0.08*	-0.10 ; -0.05	991	-0.08*	-0.11 ; -0.04	520	-0.12*	-0.14 ; -0.05

N=number of individuals with data on both time points, r=Pearson correlation, 95% CI=95% confidence intervals, *=significant based on 95% confidence interval, **Depression was measured in 1995, all other measures of anxious and depressive symptoms in 1993, ***Factor scores were not available in parents

Table 4.4. Genetic correlations of MET hours with symptoms of anxiety and depression, estimated in the longitudinal bivariate genetic models (Figure 4.1c)

MET hours with:	Depressive symptoms (YASR)		Anxious symptoms (STAI)		Somatic anxious symptoms (ABV)		Neuroticism (ABV)		Anxious depressive symptoms (FS)	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
$r_{g,M}$	-0.20*	-0.35 ; -0.06	-0.26*	-0.39 ; -0.13	-0.25	-0.43 ; 0.02	-0.21*	-0.36 ; -0.06	-0.24*	-0.36 ; -0.12
$r_{g,F}$	-0.34*	-0.43 ; -0.24	-0.28*	-0.38 ; -0.17	-0.40*	-0.50 ; -0.30	-0.27*	-0.38 ; -0.14	-0.32*	-0.41 ; -0.22

$r_{g,M}$ = genetic correlation in men, $r_{g,F}$ = genetic correlation in women, 95% CI = 95% confidence interval, *=significant based on 95% confidence interval

Discussion

This study corroborates previous findings in studies (Brown et al., 2005; Camacho et al., 1991; Cooper-Patrick et al., 1997; De Moor et al., 2006; Farmer et al., 1988; Kritz-Silverstein et al., 2001; Mobily et al., 1996; Norris et al., 1992; Stephens, 1988; Steptoe et al., 1997; Strawbridge et al., 2002; van Gool et al., 2003; Weyerer, 1992; Wise et al., 2006) with non-genetically informative designs that showed modest but significant cross-sectional and prospective associations of regular exercise with reduced anxious and depressive symptoms. In addition to these population-based findings, evidence from experimental studies (Brosse et al., 2002; Byrne & Byrne, 1993; Craft & Landers, 1998; Folkins & Sime, 1981; Gauvin & Spence, 1996; Lawlor & Hopker, 2001; Long & Vanstavel, 1995; North et al., 1990; Salmon, 2001; Scully et al., 1998; Steptoe, 2006) in healthy and clinical samples indicating decreased symptoms after standardized training programs leads to the dominant hypothesis that the association between exercise behavior and reduced symptoms of anxiety and depression derives from a causal effect of exercise. In the present study, a series of rigorous tests of this hypothesis was performed using a genetically informative design. None of these tests supported the causal hypothesis. Instead, the association of exercise with reduced anxious and depressive symptoms was explained by common genetic factors; there is a common genetic vulnerability to both lack of regular exercise and the risk for anxiety and depression in the population. It is unknown which genes might be involved in both voluntary exercise behaviour and the risk for anxiety and depression, but genes involved in the dopaminergic, norepinephrenergic, opioidergic or serotonergic pathways of the brain are likely candidates to simultaneously affect the regulation of exercise drive and mood (Chaouloff, 1997; Dishman et al., 2006; Dishman, 1997; Goldfarb & Jamurtas, 1997).

This outcome seems at odds with findings from randomized controlled trials in clinical and population samples indicating that regular exercise can relieve symptoms in subclinical individuals and in patients diagnosed as having an anxiety or depressive disorder (Brosse et al., 2002; Craft & Landers, 1998; Lawlor & Hopker, 2001; Long & Vanstavel, 1995; North et al., 1990). To understand the different outcomes of both types of study, it is crucial to make a distinction between the effects of prescribed and externally monitored exercise treatment in selected subgroups and the effects of voluntary leisure-time exercise at the population level. Only voluntary leisure-time exercise is influenced by genetic factors, whereas the other type of exercise is environment-driven. The absence of causal effects

of voluntary exercise on symptoms of anxiety and depression does not imply that manipulation of exercise cannot be used to change such symptoms. It means that the population association, either cross-sectional or longitudinal, cannot be used to justify exercise as a treatment without an actual randomized controlled trial. The possible difference in antidepressant effects of prescribed vs voluntary exercise is consistent with findings from a recent study (Blumenthal et al., 2007) suggesting that the therapeutic effects of exercise are nonspecific to exercise. The antidepressant effects of exercise may only occur if the exercise is monitored and part of a therapeutic program.

Several limitations of this study should be noted. First, some selection bias may have been present in the sample. In general, ascertainment by twinning is a good way to obtain a random population-based sample of families, and our Dutch sample was shown to be representative of the general Dutch population in the prevalence of exercise participation and anxiety and depressive disorders, as well as with regard to socio-economic status, smoking behaviour and religion (Boomsma et al., 2002; De Moor et al., 2006; Middeldorp et al., 2005). However, individuals from highly cooperative families (i.e., families in which most individuals participate) are slightly less anxious and depressed than individuals from less cooperative families (i.e., families in which only some of the individuals participate) (Vink et al., 2004), although they were not more frequent exercisers. In our study, individuals who participated in more than one survey were, on average, less anxious and depressed than individuals who participated only once. However, this effect was small, and the associations between exercise participation and symptoms of anxiety and depression did not decrease significantly when they were recomputed in individuals who participated more than once.

A second limitation concerns the power to detect genetic and environmental correlations in the bivariate genetic model. The correlations of exercise with anxious and depressive symptoms are small. In a bivariate genetic model with a within-person correlation of 0.10, power analyses showed that necessary numbers of complete twin pairs are 3,210 to detect the genetic correlation with a power of 0.80 and 3,864 to detect the environmental correlation. With a within-person cross-trait correlation of 0.15, the numbers of pairs drop to 1,427 for the genetic correlation and 1,703 for the environmental correlation. Therefore, given the phenotypic correlations and sample sizes in this study, there was sufficient power to detect both the genetic and environmental correlations in women (>0.80) but only moderate power in men (>0.60).

A third limitation is that the present study could not explicitly test more complex mechanisms of causality (e.g., a combination of common genetic factors and a causal effect of exercise or reciprocal causality). These mechanisms can in principle be studied using direction-of-causation models developed by Heath et al. (1993) and extended by Duffy et al. (1994). However, we could not apply these models to our data because they require that heritability estimates for exercise behavior and symptoms of anxiety and depression be substantially different, which they are not (Boomsma et al., 2000; De Moor et al., 2007a; Stubbe et al., 2006a).

A fourth limitation may be the generalizability of the results in this Dutch sample to different age groups such as adolescents and older persons and to other countries. It has been shown that the prevalence of exercise participation decreases with age (De Moor et al., 2006). Although the relationship with anxious and depressive symptoms does not depend on age (De Moor et al., 2006), it remains possible that causal effects of exercise are specific to certain age ranges. The prevalence of exercise participation can also differ markedly across countries, and we do not know whether the correlation between exercise behaviour and symptoms of anxiety and depression, phenotypic and genetic, differs across countries and races/ethnicities as well (Haase et al., 2004; Steptoe et al., 1997).

Conclusions

To summarize, this study is the first (to our knowledge) to test the causal effect of exercise on symptoms of anxiety and depression in the population using longitudinal and genetically informative designs. The findings confirm that lower levels of regular exercise are associated with higher levels of anxious and depressive symptoms, but they falsify the causal hypothesis. These findings do not detract from the beneficial effects of regular exercise on numerous aspects of physical health, such as cardiovascular disease and type II diabetes mellitus (Albright et al., 2000; Kaplan et al., 1996). Our results signal psychiatrists and epidemiologists that the small but robust cross-sectional and longitudinal correlations between voluntary exercise behaviour and mental health should be interpreted with caution.



PART II
Genetics of exercise behavior



5

Genome-wide linkage scan for exercise participation in Dutch sibling pairs

This chapter is published as:

De Moor, M.H.M., Posthuma, D., Hottenga, J.J., Willemsen, G., Boomsma, D.I., & De Geus, E.J.C. (2007). Genome-wide linkage scan for exercise participation in Dutch sibling pairs. *European Journal of Human Genetics*, 15 (12), 1252–1259.

Abstract

This study was aimed at identifying the genomic loci linked to exercise participation in males and females. Cross-sectional exercise data of twins and siblings (18-50 years) were used from the Netherlands Twin Registry. The sample consisted of 1,432 genotyped sibling pairs from 622 families (1 120 sibling pairs were genotyped on all chromosomes). Exercise participation (No/Yes, based on a cut-off criterion of 4 metabolic equivalents and 60 minutes weekly) was assessed by survey. Genotyping was based on 361 markers and an average marker density of 10.6 cM. IBD status was estimated for a 1 cM grid. A variance components based sex-limited linkage scan was carried out for exercise participation. The heritability of exercise participation in males was 68.5% and in females 46.3%. The genetic overlap was estimated at 0.32, indicating that partly different genes affect exercise in the two sexes. Suggestive linkage was found in all subjects on chromosome 19p13.3 (LOD=2.18). Although sex differences in linkage effect were not significant, mainly females contributed to the suggestive linkage. The 19p13.3-13.2 region harbors a number of genes related to muscle performance and muscle blood flow, which might affect exercise behavior through exercise ability. Most likely, a large number of genes with each small effects affect exercise participation in males and females. Large collaborative samples are needed to detect these effects.

Numerous epidemiological and experimental studies have demonstrated the beneficial effects of regular exercise participation on physical and mental health (Berlin & Colditz, 1990; Farmer et al., 1988; Salmon, 2001). Despite these well-known effects, about 30% of European and North-American populations remains sedentary (Caspersen et al., 2000; Martinez-Gonzalez et al., 1999). It is well known that individual differences in exercise behavior can be explained by a combination of both environmental and genetic factors. Reviews of twin and family studies have shown a significant contribution of genetic effects to variation in adolescent and adult exercise participation and (leisure-time) physical activity. Heritability estimates range from 25 to 75%, with the lower estimates found in early adolescence, peak heritability at late adolescence/young adulthood and heritabilities of around 50% in adults (Beunen & Thomis, 1999; Stubbe et al., 2005; Stubbe & De Geus, 2006b). At all ages, there is evidence that the genetic factors influencing exercise behavior in males and females are different. Two studies found a higher heritability in male than in female adolescents (Beunen & Thomis, 1999; Maia et al., 2002), and a recent study in adults found significant lower correlations in opposite-sex than in same-sex twin pairs in four out of five large datasets from different countries (Stubbe et al., 2006a). Such a pattern of correlations in first-degree relatives suggests that either different expression patterns of the same genes or different genes play a role in exercise behavior in men and women.

The influence of specific environmental factors on exercise behavior is well researched in the epidemiological literature on determinants

of exercise behavior, although causality often still needs to be established (King et al., 1992). In contrast, there is only a handful of molecular genetic studies that identified the actual genetic variants related to exercise behavior. In one study, the dopamine 2 receptor (DRD2) gene was associated with physical activity, sports participation and occupational physical activity in adult females (Simonen et al., 2003a). In a study of adolescent females, the calcium sensing receptor (CASR) gene was associated with hours spent on physical activities per week (Lorentzon et al., 2001). In a sample of post-menopausal women, the aromatase (CYP19) gene was associated with physical activity (Salmen et al., 2003). In yet another study, the melanocortin-4 receptor (MC4R) gene was associated with daily physical activity levels in a combined sample of adult men and women (Loos et al., 2005). Finally, in a study of mild male and female hypertensives, the angiotensin-converting enzyme (ACE) gene was associated with leisure-time physical activity (Winnicki et al., 2004).

There are two genome-wide linkage studies on physical activity and none on exercise participation (Cai et al., 2006; Simonen et al., 2003b). In the first study (Simonen et al., 2003b), 172 male and 223 female adults and their parents from 207 families were genotyped and four physical activity phenotypes measured. Genotyping was based on 432 markers (average map density 7.06 cM). Three physical activity phenotypes (inactivity, moderate to strenuous physical activity and total daily activity level) were derived from a three day activity diary. The fourth physical activity phenotype (time spent on most common physical activity during the past year) was survey-based. For time spent on physical activity, suggestive linkage was found on chromosomes 11p15 and 15q13.3. For the three day diary-based physical activity phenotypes promising evidence was found on chromosome 2p22-p16 (for inactivity) and suggestive linkages were found for different loci on chromosomes 4q28.2, 7p11.2, 9q31.1, 13q22-q31 and 20q13.1.

In a second study (Cai et al., 2006), 1,030 children (both boys and girls) and 631 parents from 319 Hispanic-American families were genotyped and phenotyped. Genotyping was based on markers with an average spacing of 10 cM. Daily physical activity was measured using accelerometers. For percentage of awake time spent in sedentary activity, significant linkage was found on chromosome 18q12-q21, where the MC4R gene is located.

There was no overlap in the findings of these linkage studies and, with the exception of MC4R, the genes identified in previous association studies are not located on or nearby the identified regions in the linkage

studies. If different genes cause variation in exercise behavior in males and females, as suggested by heritability studies, then ignoring these sex differences might result in a failure to detect the separate genetic effects in males and females. In this paper, we present a sex-limited autosomal linkage scan, carried out in 1,570 individuals from 622 families using on average 361 markers.

Methods

Subjects

This study was part of an on-going study on lifestyle and health in twin families that are voluntarily registered at the Netherlands Twin Register (NTR) (Boomsma et al., 2002; Boomsma et al., 2006b). Since 1991, every two to three years participants receive questionnaires on health, lifestyle and personality. Data on exercise participation were collected in each survey in 1991, 1993, 1995, 1997, 2000 and 2002. A cross-sectional data-set was created using the most recent data on exercise participation from each family that participated one or more times in the longitudinal study.

Twins and their siblings aging between 18 and 50 years were selected. We excluded twins with unknown zygosity (N=67, note that these twins were not genotyped) and genetically unrelated siblings and half siblings (N=47). The total sample consisted of 4,230 families (9,408 twins and siblings). A subsample was genotyped and used in the linkage analyses. Genotyping procedures are described below. A detailed overview of the sample characteristics is given in the Tables 5.1 and 5.2. Zygosity of the same-sex twins was determined by DNA typing for 26.1% of the same-sex twin pairs. For the other same-sex twins zygosity was based on eight items on physical similarity and the frequency of confusion of the twins by parents, other family members and strangers. Agreement between zygosity based on these items and zygosity based on DNA was 97% (Willemsen et al., 2005).

Table 5.1. Number of families, individuals, and sibling pairs in the non-genotyped, genotyped and total sample

	Non-genotyped sample	Genotyped sample	Total sample
Number of families	3,608	622	4,230
Number of individuals	7,670	1,738	9,408
MZM pairs	445 (125)	63 (4)	508 (129)
DZM pairs	267 (90)	79 (13)	346 (103)
MZF pairs	906 (201)	101 (4)	1,007 (205)
DZF pairs	401 (152)	157 (13)	558 (165)
DOS pairs	597 (276)	146 (30)	743 (306)
Brothers	711	234	945
Sisters	883	348	1,231

Number of incomplete twin pairs in brackets (arise when data on exercise participation are missing in the co-twin), note that for the linkage analysis one MZ twin was randomly selected from each complete MZ twin pair, because MZ twin pairs share the same genotype.

Table 5.2. Prevalence of exercise participation and distribution of sex and age in the non-genotyped, genotyped and total sample

	Non-genotyped sample	Genotyped sample	Total sample
Prevalence of exercise (no/yes)	47% / 53%	50% / 50%	47% / 53%
Proportion of males/females	39% / 61%	40% / 60%	39% / 61%
Mean age (standard deviation)	27.8 (7.4)	32.3 (8.5)	28.6 (7.8)

Phenotype

Exercise participation was measured with a number of questions. The first question was “Do you participate in exercise regularly?”. This question could be answered with Yes or No. If the participants responded affirmative, further information on type, frequency and duration of exercise was gathered. All exercise activities were assigned a metabolic equivalent (MET) value, using Ainsworth’s Compendium of physical activity (Ainsworth et al., 2000). A MET score of 1 corresponds to the rate of

energy expenditure when at rest (1 kcal/kg/h). In the group of exercisers, the data of frequency and duration of the exercise activities and their MET scores did not follow a normal distribution but was highly skewed. Non-exercisers did not have values on frequency, duration and intensity level, since they are not involved in any exercise activities. Therefore, in keeping with existing epidemiological studies (Haase et al., 2004), exercise participation was defined as a dichotomous variable, classifying participants as either regular exerciser or non-exerciser. A cut-off criterion of exercising at 4 METs or more for at least 60 minutes a week in the recent year was used to classify participants as regular exercisers.

Genotyping

DNA was extracted for a sample of twins, non-twin siblings and their parents from either whole blood or buccal swabs following standard protocols (Meulenbelt et al., 1995; Miller et al., 1988). One part of the sample (N=2,399 subjects) was genotyped by the Mammalian Genotyping Service in Marshfield for a 400 marker 10 cM genome scan (2 batches, Screening sets 10 and 16). The other part of the sample (N=985 subjects) was genotyped by the Molecular Epidemiology Section, Leiden University Medical Centre, using the 10 cM Applied Biosystems Human Linkage Set v2.5 MD10 with some additional markers (419 in total). The genotype data from these screens were then combined. Allele calling and binning was equalized between 111 markers that were present in both datasets with the use of 302 overlapping samples. In case there were inconsistencies, the data was set to unknown for the tested markers (binning and allele calling inconsistencies) and persons (genotyping errors). Sex and zygosity were checked with the marker data. Pedigree relations in the entire samples were checked with the GRR program (Abecasis et al., 2001). Errors of Mendelian inheritance were detected with Pedstats (Abecasis et al., 2002). Markers and samples were removed if their total error rate was more than 1%; in all other cases the specific erroneous genotypes were set as unknown. Unlikely recombinants were detected using Merlin and erroneous genotypes were removed with Pedwipe (Abecasis et al., 2002). IBD estimation was carried out in this full genotype dataset.

There were 1,738 monozygotic (MZ) and dizygotic (DZ) twins and non-twin siblings from 622 families who were both genotyped and phenotyped (see also Table 5.1). For the present linkage analysis, we randomly selected 1 MZ twin from each MZ twin pair. Further, siblings were selected per chromosome if they had markers with an average spacing of 18 cM or less. This corresponds to the average spacing of markers if subjects with

more than 200 markers genotyped along the genome are selected (i.e. more than 50%). This resulted in a total sample of 1,570 genotyped individuals from 622 families (1,432 all-possible sibling pairs). The number of sibling pairs per chromosome ranged from 1,196 to 1,432 (1,120 sibling pairs were successfully genotyped on all chromosomes). There were 266 male, 525 female and 641 opposite sex sibling pairs. The average heterozygosity of autosomal markers was 76%. In the 1,120 sibling pairs, the average number of markers genotyped was 361 (201-761) with an average spacing of 10.6 cM. For 1,046 pairs, the genotype data came from the Marshfield marker sets, for 171 pairs the data came from the Leiden marker set. 119 pairs were genotyped in both data sets, and in 22 pairs the siblings were genotyped in the two different marker sets. For the statistical analyses the Haldane mapping function was used. All reported values are in Haldane cM. The marker positions were interpolated via locally weighted linear regression from the National Center for Biotechnology Information (NCBI) build 35.1 physical map positions and the Rutgers genetic map (Duffy, 2006; Kong et al., 2004).

IBD estimation

A sibling pair shares an allele at a specific locus identical by descent (IBD) if the allele is inherited from the same ancestor. A sibling pair can share 0, 1 or 2 alleles IBD at a specific locus. Because IBD status is not always known (depending on the availability of genotypic data in the parents, the map density of the markers and the informativeness of the markers), the probability of sharing 0, 1 or 2 alleles IBD needs to be estimated. This was done for a 1 cM grid multipoint scan using the Lander-Green algorithm implemented in Merlin (Abecasis et al., 2002). The proportion of alleles shared IBD at a specific position along the genome was estimated as:

$$\hat{\pi} = 0.5 p(\text{IBD} = 1) + p(\text{IBD} = 2)$$

where $p(\text{IBD} = 1)$ is the probability that IBD status is 1 and $p(\text{IBD} = 2)$ is the probability that IBD status is 2.

Linkage analysis

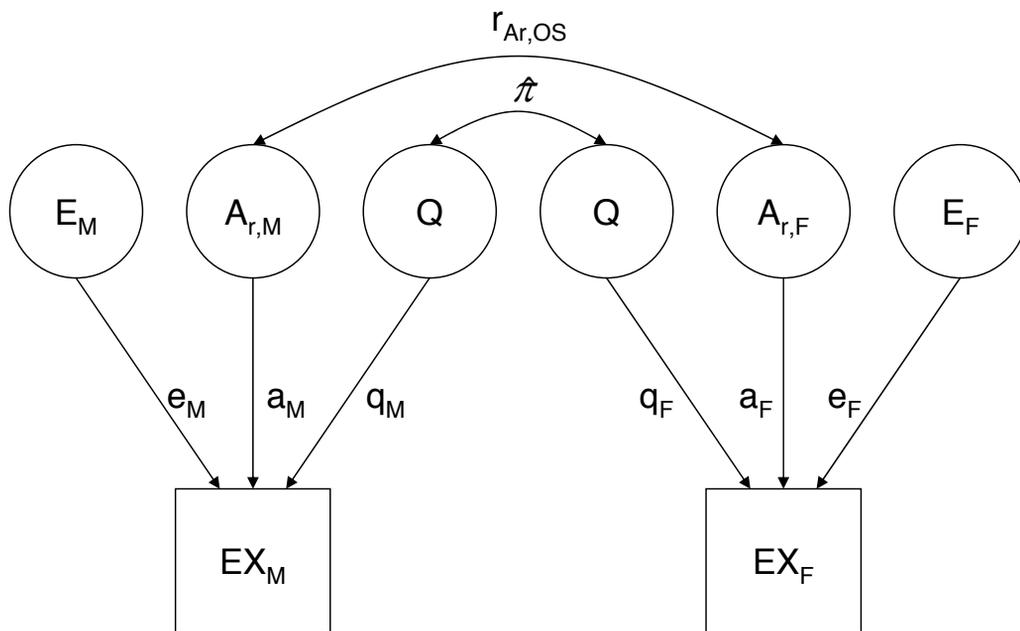
Sibling pair analysis using variance components modeling in Mx (Neale, 2000; Neale et al., 2003) was used to estimate the heritability and linkage of a putative quantitative trait locus (QTL) to exercise participation in the sample of genotyped sibling pairs. Because exercise participation is a dichotomous variable, we used a threshold liability model, in which the estimated threshold divides the latent standard normal liability distribution

into an “affected” and “unaffected” part (Neale & Cardon, 1992). Estimates of twin correlations and heritabilities in the total sample have been reported in a previous study (Stubbe et al., 2006a). Before carrying out the linkage analysis, we evaluated whether the prevalences and heritabilities in the genotyped sample were the same as in the total sample. We fitted an AE threshold model with non-scalar sex limitation (Medland, 2004) to the data in both the genotyped and non-genotyped sample. In this model, the variance in liability for exercise participation in each sample (which is set at 1) was decomposed into additive genetic (A) and unique environmental (E) components, allowing for both quantitative differences in variance decomposition (i.e., different heritabilities in males and females) and qualitative differences (i.e., a lower genetic correlation in opposite-sex pairs). We tested for differences in prevalences and variance decomposition in the two samples by constraining the parameters to be equal across samples. The fit of the models was evaluated by use of the log-likelihood ratio test (LRT), computing the difference in minus twice the log-likelihood (-2LL) between two nested models. This difference is χ^2 distributed. The degrees of freedom (df) equals the difference in df between the two models. An alpha level of 0.01 was used for this test. A significant χ^2 means that the constrained model fits significantly worse than the full model.

We next carried out the linkage scan for all positions along the genome, allowing not only for sex limitation in the heritability, but also in the linkage effect (Medland, 2005). The estimate of the proportion of alleles shared IBD in a specific region along the genome ($\hat{\pi}$) was used to model the covariance in a sibling pair that is due to the putative locus. If the QTL effect is significant, the predicted (model-based) phenotypic covariance will be largest for sibling pairs sharing all alleles IBD in a specific region along the genome, the covariance will be smallest for sibling pairs sharing no alleles IBD in this region. Figure 5.1 shows the path model that was used (drawn for OS sibling pairs). All path loadings, including the QTL effect on the trait, are allowed to be different in males and females. In full sibling pairs, for reasons of identification, the part of the phenotypic covariance that is not explained by the QTL is usually modeled as a shared background factor, consisting of both common environmental and genetic factors (Neale, 2000). However, because it was known that covariance between sibling pairs is explained by additive genetic factors only (Stubbe et al., 2006a), we modeled this background factor as an additive genetic factor. The square of the estimate of the QTL effect in the model reflects the proportion of the total variance in liability to exercise participation (fixed at one) that is explained by the QTL. The genetic correlation

in the OS sibling pairs was freely estimated (but restricted to vary between 0 and 0.5, based on biological plausibility), while the genetic correlation in same-sex sibling pairs was fixed at 0.5. The correlation between QTL's was given by $\hat{\pi}$, a value that is specific to each sibling pair (specified as a definition variable in Mx). The threshold was modeled according to the best fitting model in the heritability analysis, which included an age effect on the threshold.

Figure 5.1. Path diagram used to model non-scalar sex-limited linkage (shown for an opposite sex sibling pair)



EX_M : Exercise participation in males, EX_F : Exercise participation in females, e_M : unique environmental path males, a_M : additive genetic path males, q_M : QTL effect males, e_F : unique environmental path females, a_F : additive genetic path females, q_F : QTL effect females, E_M : male unique environmental factor, E_F : female unique environmental factor, $A_{r,M}$: male additive genetic background factor, $A_{r,F}$: female additive genetic background factor, Q: Quantitative trait locus (QTL), $r_{Ar,OS}$: genetic correlation between additive genetic background factor of an opposite sex sibling pair, $\hat{\pi}$: estimated proportion of alleles shared IBD at the QTL

We separately tested the significance of the QTL effects in males and females. To test the significance of the male QTL effect, we compared the fit of the full model with both sex-specific QTL effects with the fit of the model in which the effect of the QTL in males was fixed at 0. Similarly, we evaluated the significance of the female QTL effect. We also evaluated

whether the sex differences in QTL effects in males and females were significant, by comparing the model with two QTL effects with the model in which these two effects were equated. We further tested the significance of the QTL effect equated across sex, by comparing the fit of this model with the model in which the QTL parameter was dropped.

Significance of effects was evaluated by use of the likelihood ratio test, from which the LOD score can be computed by dividing the obtained chi-square test statistic by $2 \ln 10$ (~ 4.6). The linkage is considered significant if a LOD larger than 3.6 is observed, corresponding to a genome-wide 5% chance that a significant LOD score due to random fluctuations is found somewhere on the genome, thereby correcting for multiple testing. The linkage is considered suggestive if a LOD score larger than 2.2 is observed, corresponding to the expectation that a LOD score of this magnitude as a result of random fluctuations is found once in a genome scan (Lander & Kruglyak, 1995).

We also computed the empirical thresholds for suggestive and significant linkage in males and females, by randomly permuting the datasets a 1,000 times. Permutations were carried out by randomly assigning the IBD estimates to the sibling pairs, keeping the sibling pairs and IBD structure along the genome intact. Each permuted dataset was then analyzed. The empirical thresholds for suggestive linkage in males and in females were computed by obtaining the maximum sex-specific LOD scores for each chromosome out of the 1,000 analyses, and determining what sex-specific LOD score occurs a 1,000 times out of 22,000. The empirical threshold for suggestive linkage in males was 1.89 and in females 1.91. The thresholds for significant linkage in males and females were computed by recording the maximum sex-specific LOD scores in each linkage scan in one of the permuted datasets, and then determining which sex-specific LOD scores occur 50 out of 1,000 times. The empirical threshold for significant linkage in males was 3.22 and in females 3.21.

Results

The prevalence of exercise participation in the genotyped individuals was not significantly different from the prevalence of exercise in the non-genotyped individuals ($\chi^2=1.44$, $\Delta df=1$, $p=0.23$, see Table 5.2 for prevalence estimates). There were also no significant differences in heritability of exercise participation in both males and females between the genotyped and non-genotyped sample ($\chi^2=1.67$, $\Delta df=1$, $p=0.64$). In the genotyped

sample, the heritability in males was estimated at 68.5% (95% confidence interval (CI): 56.7% - 82.9%) and in females at 46.3% (95% CI: 26.0% - 63.6%). The proportion of variance explained by E in males is estimated at 31.5% (95% CI: 15.7% - 61.8%) and in females at 53.7% (95% CI: 39.2% - 73.9%). The genetic correlation in opposite sex pairs is 0.32 (95% CI: 0.08 - 0.50). These results suggest that no selection occurred with regard to both phenotype and genotype in the genotyped sample, and the results from the linkage analysis can be generalized to the total sample under study. Qualitative sex differences in genetic effects influencing exercise participation were also found in the genotyped sample, which justifies modeling sex-specific QTL effects in the linkage analysis.

Figure 5.2 displays the LOD scores for males, females and the combined sample plotted for each chromosome. Tests of sex heterogeneity showed that nowhere along the genome are the sex differences in QTL effect significant, when correcting for multiple testing. Suggestive linkage is found in all subjects on chromosome 19 (maximum LOD=2.18 at 13 cM nearby marker D19S247). It becomes clear from figure 2 that females contribute more strongly to this LOD score than males (maximum LOD in females=2.87 at 11 and 12 cM, versus 0.83 in males at 9-12 cM). The proportion of variance explained at this QTL is 38.0% (95% CI: 16.8% - 55.6%). The estimate of the genetic correlation of the additive genetic residual factors in opposite sex pairs at this locus is zero. Dropping 1 LOD at both sides of the peak, the confidence interval around the peak is 0-28 cM. This region is flanked by markers D19S591 at 19p13.3 and D19S865 at 19p13.2.

Figure 5.2. LOD scores across the autosomal genome for males (thin dotted line), females (thin solid line) and males and females combined (thick solid line)

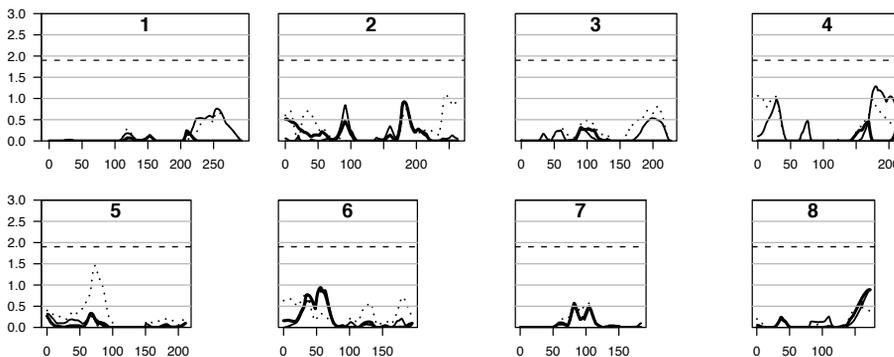
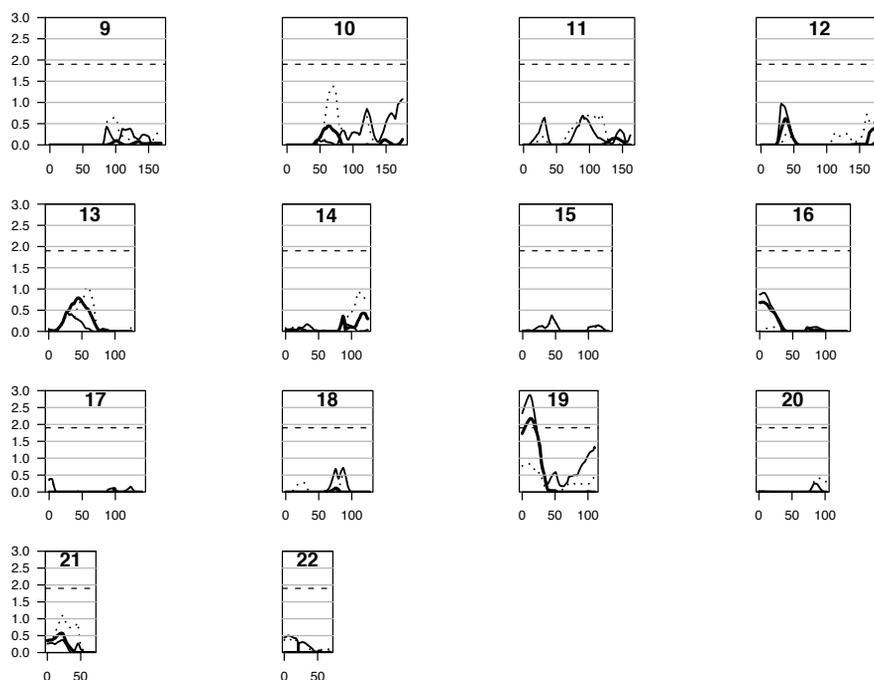


Figure 5.2. LOD scores across the autosomal genome for males (thin dotted line), females (thin solid line) and males and females combined (thick solid line)



Discussion

This study shows suggestive linkage on chromosome 19p13.3 near marker D19S247 (LOD=2.18), explaining 38.0% of the total variance in liability for exercise participation. The maximum LOD score in females in this region was 2.87 and in males 0.83. The region on 19p does not coincide with the regions that were found for physical activity levels in the previous linkage studies (Cai et al., 2006; Simonen et al., 2003b), which could be partly explained by the different definitions of exercise behavior that were used. The region further does not harbor genes that have been related to exercise or physical activity phenotypes in previous association studies (Loos et al., 2005; Lorentzon et al., 2001; Salmen et al., 2003; Simonen et al., 2003a; Winnicki et al., 2004). Also, no convergence was found to genes on the latest version of the human gene map for performance and health-related fitness phenotypes (Wolfarth et al., 2005) that provides an overview of all genes and QTLs identified through association and linkage studies that have been related to physical performance, physical activity, or health-related fitness phenotypes.

A tentative search for genes located under the peak on 19p that are possibly related to exercise participation was made using the Ensembl database (Birney et al., 2006). We hypothesize that three biological pathways might explain how genes influence exercise behavior. First, genes that influence exercise *ability* might indirectly influence voluntary engagement in exercise activities. A person's genetic make-up determines whether this person is good at exercise or not, and this innate exercise ability might in turn influence whether a person actually engages in exercise behavior. Second, the engagement in exercise activities might depend on personality traits such as neuroticism, extraversion or sensation seeking, which are all also under genetic control (Johnson & Krueger, 2004; Rettew et al., 2006; Stoel et al., 2006; Viken et al., 1994), or clinical end-points like depression. For example, a linkage study on depressive disorders reported significant linkage of region 19p13.2-13.1 to depressive spectrum disorder (Zubenko et al., 2003). Depression, which is two times more prevalent in women, may prevent people to take part in regular exercise. Genes influencing depression might therefore also influence exercise behavior.

A third biological pathway by which the influence of genes on exercise behavior might be mediated is through the acute (rewarding) effects of exercise. A person's genetic predisposition to experience more rewarding than aversive acute effects of exercise (for example large increase in performance or physical fitness, enhanced feelings of well-being) may determine whether a person engages and continues to engage in exercise activities. A number of possibly interesting genes are located at region 19p13.3-13.2, all concerning exercise ability through their influence on either muscle performance or muscle blood flow: the muscle integrin-binding protein gene (MIBP), the thyroid receptor-interacting protein 10 gene (TRIP-10), the myosin IE gene (MYO1F), the endothelial differentiation G-protein coupled receptors 5 and 6 genes (EDG5 and EDG6), the thromboxane A2 receptor gene (TBXA2R) and the calponin-1 gene.

The main limitation of this study is the use of a dichotomy to quantify exercise. It is well-known that large samples are needed in order to detect linkage signals of small effect. The power to detect variance components such as additive genetic QTL variance with ordinal data is even lower than with continuous data (Neale et al., 1994a). We carried out a number of analyses to investigate the power to detect non sex-specific and sex-specific QTL effects and the power to detect sex differences in QTL effects in the non-scalar sex-limitation model for ordinal data. In the simulations, we assumed heritabilities and a genetic correlation in the opposite-sex pairs that correspond to the values of our real data (heritability males

68%, heritability females 46%, genetic correlation opposite-sex pairs 0.32). With 1 440 sibling pairs in total and proportions of male-male, female-female and opposite-sex sibling pairs that correspond to the real data proportions, the power to detect a QTL effect of 25% explained variance is 0.84 for the male-specific, 0.91 for the female-specific and 0.96 for the non sex-specific QTL effect. For a QTL explaining 10% of the variance these values are 0.22, 0.25 and 0.31, respectively. The power to detect a difference of 15% explained variance between males and females is 0.14 (assuming the variance in males is 0.05 and in females 0.20). Similar power results are obtained when different variances are assumed (range between 0 and 20%) but with the same difference of 15%. Thus, the power to detect linkage signals of small effects is rather low and the power to detect sex differences in QTL effects is very low. However, we stress that our sample size is large compared with previous linkage studies on physical activity phenotypes and other linkage studies on complex phenotypes.

Taken together, this study suggests that the substantial heritability of exercise behavior in both males and females cannot be attributed to a few major genes with large effects. Rather, exercise behavior should be considered among the complex, polygenic traits, with in part different genes affecting exercise behavior in males and females. Considering the diversity of the hypothesized biological pathways through which genes might affect exercise behavior, it is likely that a large number of genes with all minor effects account for the heritability of exercise participation. For gene finding efforts for exercise behavior to be successful, large collaborative samples will be needed to detect and replicate the linkage signals.



6

Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs

This chapter is published as:

De Moor, M.H.M., Spector, T.D., Cherkas, L.F., Falchi, M., Hottenga, J.J., Boomsma, D.I., & De Geus, E.J.C. (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics*, 10 (6), 812-820.

Abstract

Association studies, comparing elite athletes with sedentary controls, have reported a number of genes that may be related to athlete status. The present study reports the first genome-wide linkage scan for athlete status. Subjects were 4,488 adult female twins from the TwinsUK Adult Twin Registry (793 monozygotic (MZ) and 1,000 dizygotic (DZ) complete twin pairs, and single twins). Athlete status was measured by asking the twins whether they had ever competed in sports and what was the highest level obtained. Twins who had competed at the county or national level were considered elite athletes. Using structural equation modeling in Mx, the heritability of athlete status was estimated at 66%. Seven-hundred DZ twin pairs that were successfully genotyped for 1946 markers (736 microsatellites and 1,210 SNPs) were included in the linkage analysis. Identical-by-descent probabilities were estimated in Merlin for a 1 cM grid, taking into account the linkage disequilibrium of correlated SNPs. The linkage scan was carried out in Mx using the $\hat{\pi}$ -approach. Suggestive linkages were found on chromosomes 3q22-q24 and 4q31-q34. The peak on 3q22-q24 was found at the SLC9A9 gene. The region 4q31-q34 overlaps with the region for which suggestive linkages were found in two previous linkage studies for physical fitness (FABP2 gene) (Bouchard et al., 2000) and physical activity (UCP1 gene) (Simonen et al., 2003b). Future association studies should further clarify the possible role of these genes in athlete status.

Elite athletic performance is thought to be the result of the combination of a high genetic potential and the optimal environmental conditions, such as training and nutrition (MacArthur & North, 2005). It has long been recognized that physical performance is genetically determined (Bouchard & Malina, 1998), as evidenced by the significant heritabilities ranging from 31 to 85% of physical performance traits involving different aspects of cardiorespiratory fitness (Bouchard et al., 1998; Bouchard et al., 1999; Perusse et al., 2001) and skeletomuscular strength and performance (De Mars et al., 2007; Thomis et al., 1998).

The number of studies aiming to identify the actual genetic variants that account for the heritability of physical performance phenotypes is increasing and are summarized in the 'human gene map for performance and health-related exercise phenotypes', which is regularly updated (Perusse et al., 2003; Rankinen et al., 2001; Rankinen et al., 2002; Rankinen et al., 2004; Wolfarth et al., 2005). The human gene map includes, besides linkage and association studies on different physiological parameters related to exercise, also a number of association studies of candidate genes in which elite athletes are compared with sedentary controls. Genes that have been related to elite athlete status are for example the angiotensin-converting enzyme (ACE) gene (Woods et al., 2000) and the bradykinin receptor (B2R) gene (Williams et al., 2004). These genes are thought to have an impact on physical performance through increasing cardiorespiratory fitness as a result of training, but possibly also through enhanced muscle efficiency (Williams et al., 2000). Other genes that have been related to elite athlete status are the actin-binding protein α -actinin-3 (ACTN3)

gene (MacArthur & North, 2007; Paparini et al., 2007; Yang et al., 2003) and two adrenergic receptor genes (ADRA2A and ADRB2) (Moore et al., 2001; Wolfarth et al., 2000).

To date, no linkage studies have been published that identified the genomic locations that may be related to athlete status. In this study, the results of a heritability analysis and genome-wide linkage scan are presented for data on athlete status in a sample of British female twin pairs. Athlete status was measured by asking the twins whether they ever participated in sports and what was the highest level they ever competed at. Twins who competed at the national or county level were considered elite athletes.

Methods

Sample

Phenotype information came from a mailed survey sent out in 2,000 to female twins registered at the TwinsUK Adult Twin Registry (Spector & Williams, 2006). There were 4,527 twins with valid data on the level of participation in competitive sports. Twins with unknown zygosity were excluded (39 individuals), resulting in a total sample of 4,488 individuals, from which 1,793 complete twin pairs could be formed (i.e., both individuals had data on sports level). Table 6.1 gives an overview of the number of complete and incomplete twin pairs as a function of zygosity. Zygosity of the twin pairs was determined using a standardized questionnaire and DNA fingerprinting (Spector et al., 1996). The mean age of the individuals was 51.9 (standard deviation: 12.8), the minimum age was 20 and the maximum 83 years old.

Table 6.1. Number of complete and incomplete female MZ and DZ twin pairs and their polychoric correlations (95% Confidence Interval)

	Complete twin pairs	Incomplete twin pairs	Total pairs (individuals)	Polychoric twin correlation (95% CI)
MZ	793	348	1,141 (1,934)	0.66 (0.59 ; 0.71)
DZ	1,000	554	1,554 (2,554)	0.32 (0.24 ; 0.40)
Total	1,793	902	2,695 (4,488)	

Phenotyping

Athlete status was measured by asking the participants to indicate for a list of sports whether they had ever participated in each of these sports and what was the highest level at which they had ever competed in these

sports. The levels were social only, in a school, club or university team, at the county level or at the national level. Table 6.2 lists the sports that were included in the survey. Yoga and walking were not included in the analyses, since these were low-impact sports (at less than 4 metabolic equivalents). Participants were classified into three categories: 1) never participated in sports in any organized form (no exercise and exercise at the social level), 2) ever participated in sports at the school, club or university level and 3) ever participated in sports at the county or national level. Women who ever participated at sports at the county or national level are thought to have been elite athletes (MacArthur & North, 2005).

Table 6.2. Number of women who never or ever participated in sports at the school/club/university or county/national level for a list of sports

	Never participated in organized sport	Ever participated in sport at school/club/university level	Ever participated in sport at county/national level
1. Swimming	3,849 (85.8)	567 (12.6)	72 (1.6)
2. Cycling	4,463 (99.4)	23 (0.5)	2 (0.0)
3. Running	3,656 (81.5)	742 (16.5)	90 (2.0)
4. Keep fit / Aerobics	4,382 (97.6)	101 (2.3)	5 (0.1)
5. Gymnastics	4,092 (91.2)	377 (8.4)	19 (0.4)
6. Tennis	3,983 (88.7)	487 (10.9)	18 (0.4)
7. Badminton	4,195 (93.5)	280 (6.2)	13 (0.3)
8. Squash	4,413 (98.3)	68 (1.5)	7 (0.2)
9. Golf	4,437 (98.9)	46 (1.0)	5 (0.1)
10. Skiing	4,468 (99.6)	18 (0.4)	2 (0.0)
11. Ice skating	4,480 (99.8)	8 (0.2)	0 (0.0)
12. Yoga	4,472 (99.6)	13 (0.3)	3 (0.1)
13. Dancing	4,351 (96.9)	122 (2.7)	15 (0.3)
14. Rugby / Football	4,423 (98.6)	61 (1.4)	4 (0.1)
15. Cricket	4,433 (98.8)	52 (1.2)	3 (0.1)
16. Martial arts	4,457 (99.3)	26 (0.6)	5 (0.1)
17. Boxing	4,485 (99.9)	2 (0.0)	1 (0.0)
18. Hill walking	4,463 (99.4)	24 (0.5)	1 (0.0)
19. Walking	4,444 (99.0)	40 (0.9)	4 (0.1)
20. Other	3,765 (83.9)	597 (13.3)	126 (2.8)
Any sport	2,498 (55.7)	697 (37.4)	311 (6.9)

Note: Because some women have ever participated in more than one sports at the school/club/university or county/national level, the numbers given for each sport do not add up to number given for any sport.

Genotyping

DNA of 2,600 individuals was extracted from venous blood samples following standard protocols. Two different marker sets were used: a microsatellite marker set using standard ABI Prism genotyping methodologies (Applied Biosystems, Foster City, CA, USA) and a single nucleotide polymorphism (SNP) marker set (HuSNP GeneChips, Affymetrix), which are more fully described elsewhere (Wilson et al., 2003). Basic checks for sex and zygosity were carried out. Errors of Mendelian inheritance were detected using Pedstats (Abecasis et al., 2002). Unlikely recombinants were detected using Merlin (Abecasis et al., 2002). Unlikely or erroneous genotypes due to Mendelian or genotyping errors were set as unknown. SNPs with minor allele frequency lower than 0.05 were excluded. Genetic map positions of the markers were taken from the interpolated genetic map developed by Duffy (2006). For markers that were not present on this map, their physical position was taken from the National Center for Biotechnology Information (NCBI) build 35.1. Locally weighted linear regression of the genetic map positions on the physical positions was used to impute the missing genetic positions. Markers that could not be reliably located on the physical map (which was the case for some SNPs) were excluded from the dataset. This resulted in 1,946 markers, of which 736 were microsatellites and 1,210 SNPs. All genotyped dizygotic (DZ) twin pairs with complete data on athlete status were selected (705 pairs). Twin pairs for whom one of the individuals had less than 200 successfully genotyped markers were excluded (5 pairs). These were pairs with only microsatellites genotyped. There were 99 pairs with less than 200 successfully genotyped microsatellites, but with a large number of SNPs. These pairs were retained in the analyses. Thus, the final sample used for the linkage analysis consisted of 700 pairs. The mean age of these individuals was 54.6 (standard deviation: 11.5), the minimum age was 21 and the maximum 83 years old. Average spacing of all 1,946 markers was 6.22 cM and the average heterozygosity 61.0%. Average spacing of the 736 microsatellites was 15.9 cM and the average heterozygosity 74.1%. Excluding the 99 pairs with mainly SNPs genotyped, the average spacing was 7.35 cM. For the statistical analyses the Haldane mapping function was used. All reported values are in Haldane cM.

Markers that were close together on the genetic map, as was the case for some of the SNPs in the dataset, may be in linkage disequilibrium. If linkage disequilibrium among markers is ignored in the linkage analysis, this might lead to an upward bias in linkage signals (Abecasis & Wigginton, 2005). Merlin accommodates a feature that identifies clusters

of markers that are in linkage disequilibrium and estimates the haplotype frequencies for the clustered markers. Using an r^2 of 0.30 as a cut-off value, clusters of markers and their haplotype frequencies were estimated. This resulted in 73 clusters, of which 62 contained two markers, 10 contained 3 markers and 1 contained 4 markers. Three clusters contained both SNPs and microsatellites. For markers that were not clustered, the allele frequencies were determined by counting the alleles that occurred in the dataset. The obtained allele and haplotype frequencies were used to estimate the identical-by-descent (IBD) probabilities, thereby assuming that within clusters the recombination fraction is zero and across clusters there is linkage equilibrium (Abecasis & Wigginton, 2005).

IBD estimation

A sibling pair shares an allele at a specific genomic locus identical by descent (IBD) if the allele is inherited from the same parent. Thus, a sibling pair can share 0, 1 or 2 alleles IBD at a specific locus. If parental genotypes are not available, the probabilities of IBD status 0, 1 or 2 were estimated based on the allele frequencies as observed in the population. The IBD probabilities for each sibling pair were estimated using the Lander-Green algorithm in Merlin for a 1 cM grid (Abecasis et al., 2002). The IBD probabilities were used to estimate the proportion of alleles shared IBD for each sibling pair, using the formula (Sham, 1998):

$$\hat{\pi} = 0.5p(\text{IBD} = 1) + p(\text{IBD} = 2)$$

where $p(\text{IBD} = 1)$ is the probability that IBD status is 1 and $p(\text{IBD} = 2)$ is the probability that IBD status is 2, and $\hat{\pi}$ can take values ranging between 0 and 1. IBD estimates for all genomic positions for all sibling pairs were saved to files and later used for the linkage analysis in Mx.

Statistical analyses

Standard structural equation modeling methods were used in Mx (Neale et al., 2003; Neale & Cardon, 1992). Threshold models were fitted to the raw ordinal data on athlete status, including data from single twins, using full information maximum likelihood estimation. The threshold model assumes that a categorical variable has an underlying liability with a continuous and standard normal distribution. For athlete status, two thresholds divided the liability distribution into the three observed categories. In a saturated model, which was fitted to the data of MZ and DZ twin pairs, two thresholds were estimated for each twin in each zygosity group (8 thresholds in total). A cohort effect was allowed on each threshold, by

modeling age as a definition variable on the thresholds. It was assumed that the effect was the same on all first thresholds and on all second thresholds (2 regression coefficients were estimated). Polychoric MZ and DZ twin correlations were estimated. Thus, the saturated model contained 12 free parameters. We tested for the significance of the cohort effects on the first and second thresholds and we tested for birth order and zygosity effects in the thresholds. Based on the pattern of MZ-DZ correlations, an ADE model was fitted to the data and the proportions of variance in liability for athlete status due to A, D and E were estimated. We tested for the significance of D and of A and D. The significance of these parameters was evaluated by the log-likelihood ratio test (LRT). The difference in minus two times the log-likelihood (-2LL) between the full ADE model and the submodel in which a parameter was dropped is χ^2 distributed. The degrees of freedom of the χ^2 -test (Δdf) equal the difference in degrees of freedom between the two models. If the χ^2 -test yielded a p-value higher than 0.05 the fit of the submodel was not significantly worse than the fit of the full model and the submodel was kept as the most parsimonious and best fitting model.

In the sample of genotyped DZ twin pairs, the DZ twin correlation and the thresholds were computed to examine whether the genotyped sample deviated from the total sample with respect to the DZ twin correlation and the prevalence of sports participation at the various levels. Next, an AQE model was fitted to the data to test, at each locus, whether the proportion of alleles shared IBD in each DZ twin pair explained their resemblance in athlete status (Neale et al., 2003). Using the previously given formula, $\hat{\pi}$ was computed in Mx from the estimated IBD probabilities. At each locus, the significance of the effect of Q was evaluated by use of the likelihood ratio test. From this test, the LOD score can be computed by dividing the obtained chi-square test statistic by $2\ln 10$ (~ 4.6).

Empirical thresholds for suggestive and significant linkage were computed by randomly permuting the datasets a 1,000 times. Permutations were carried out by randomly assigning the IBD estimates to the sibling pairs, keeping the sibling pairs and IBD structure of the whole genome intact. A linkage scan was then performed on each permuted dataset. The empirical threshold for suggestive linkage was computed by obtaining the maximum LOD score for each chromosome out of the 1,000 analyses, and determining what LOD score occurred a 1,000 times out of 22,000. The threshold for significant linkage was computed by recording the maximum LOD score in each linkage scan on each permuted dataset, and then determining which LOD score occurred 50 out of 1,000 times (see also

Boomsma et al., 2006a). The empirical threshold for suggestive linkage was 1.63 and for significant linkage 2.96.

Results

Prevalence of sports participation at the various levels

From 4,488 women, there were 311 women who ever participated in sports at the county or national level (6.9%). Sixteen hundred and seventy-nine women ever participated in sports at the school, club or university level (37.4%). The remaining 2,498 women never participated in any organized form of sports (55.7%). There were no differences in these prevalences between first and second-born twins ($\chi^2=3.53$, $\Delta df=4$, $p=0.47$) nor between MZ and DZ twins ($\chi^2=2.50$, $\Delta df=4$, $p=0.64$). Women from older birth cohorts less frequently participated in sports at the school, club or university level and the county or national level, as indicated by the significant and positive regression coefficients of age on both the first ($\chi^2=100.64$, $\Delta df=1$, $p<0.001$) and the second threshold ($\chi^2=33.63$, $\Delta df=1$, $p<0.001$).

Heritability of athlete status

The polychoric MZ twin correlation was 0.66 (95% Confidence Interval (CI): 0.59-0.71) and the polychoric DZ twin correlation 0.32 (95% CI: 0.24-0.40). We fitted an ADE model to the data. The thresholds were modeled according to the best fitting constrained saturated model, which contained two thresholds and a cohort effect on each threshold. The model fitting results of the heritability analyses of athlete status are given in Table 6.3. The AE-model described the data adequately. Of the total variance in liability for athlete status, 65.5% was explained by A (95% CI: 59.6% - 70.6%) and 34.5% by E (95% CI: 29.3% - 40.4%). Thus, a combination of additive genetic and non-shared environmental factors influences athlete status.

Table 6.3. Model fitting results for heritability analyses of athlete status

Fit statistics						
Model:	-2LL	df	vs.	χ^2	Δ df	p
1. Saturated	7479.07	4,476	–	–	–	–
2. ADE	7483.82	4,482	1	4.751	6	0.576
3. AE	7483.85	4,483	1	4.783	7	0.686
4. E	7767.05	4,484	3	283.196	1	<0.001

Parameter estimates							
Model:	A	D	E	τ 1	τ 2	β 1	β 2
1. Saturated	–	–	–	–	–	–	–
2. ADE	62.5	0.03	34.4	-0.720	0.797	0.017	0.014
3. AE	65.5	–	34.5	-0.719	0.797	0.017	0.014
4. E	–	–	100	-0.673	0.799	0.016	0.014

Note: -2LL=-2 log-likelihood; df=degrees of freedom; vs.=model to which the fitted model is compared; χ^2 =chi-square value; Δ df=difference in degrees of freedom between fitted and comparison model; p=p-value; A=proportion of variance explained by additive genetic factors; D=proportion of variance explained by non-additive genetic factors; E=proportion of variance explained by non-shared environmental factors; τ 1=first thresholds; τ 2=second threshold; β 1=regression coefficient of age on the first threshold; β 2=regression coefficient of age on the second threshold; Positive age regression coefficients indicate that athlete status was less frequent in older women; The most parsimonious and best fitting model is shown in bold.

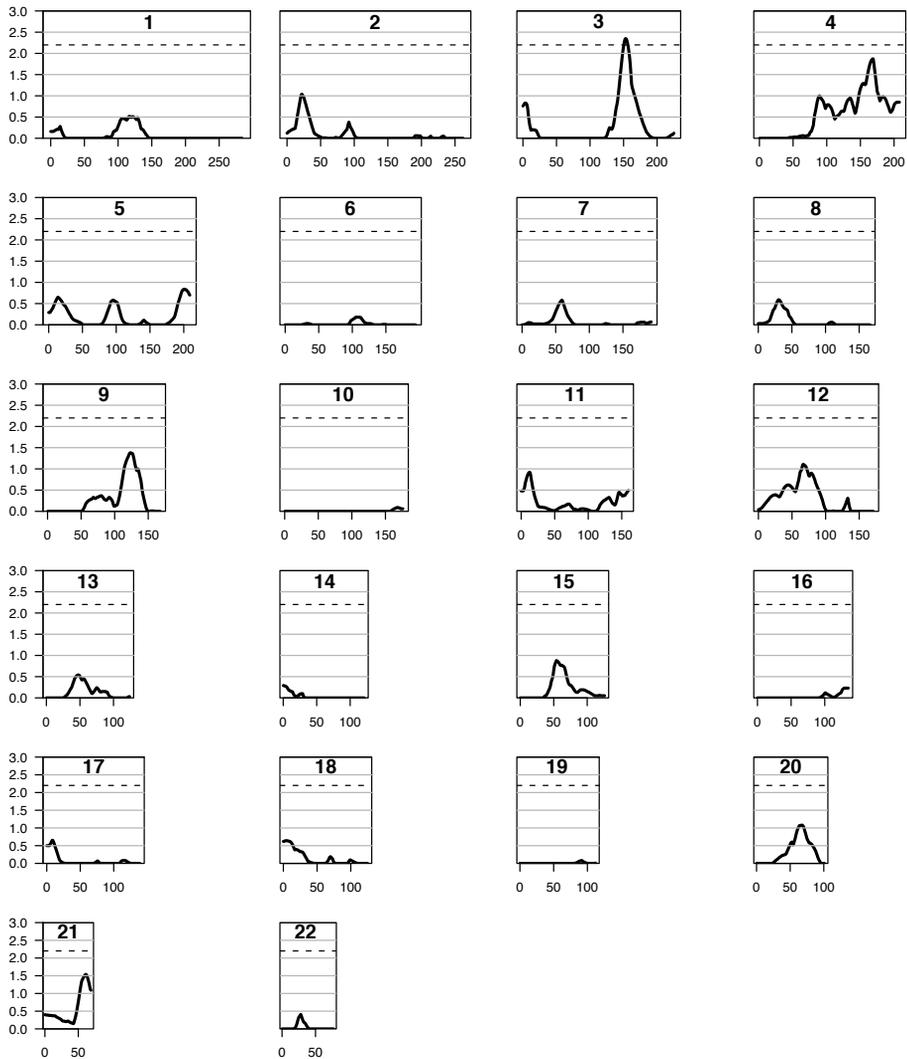
Genome-wide linkage scan for athlete status

The prevalence of sports participation at the various levels in the genotyped DZ twins was highly similar to the percentages in the total sample: 6.8% ever participated in sports at the county or national level, 36.9% ever participated at the school, club or university level and 56.3% did not ever participate in sports at the organized level. The twin correlation in the genotyped DZ twin pairs was 0.33. Thus, the genotyped sample was representative for the total sample in terms of both prevalence and the DZ twin correlation.

The results of the genome-wide linkage scan are depicted in Figure 6.1. Suggestive linkages were found on chromosomes 3q24 and 4q32.3. The peak on 3q24 has a maximum LOD score of 2.35 at 153 cM. The confidence interval of this peak, using the drop 1 LOD rule, is 3q22-q24

(145-160 cM). The markers at the peak are D3S1569 and the clustered markers rs2801 and rs2800, all residing within the sodium/hydrogen exchanger 9 (SLC9A9) gene. The second peak is on 4q32.3 has a maximum LOD score of 1.87 and is found at 168/169 cM near marker D4S1597. The confidence interval of the peak is 4q31-q34 (147-188 cM).

Figure 6.1. LOD scores of the genome-wide linkage scan for athlete status in the sample of genotyped DZ twin pairs



Note: dashed line indicates the empirical threshold for suggestive linkage

Discussion

Seven percent of the British adult women have ever competed in sports at the county or national level. Women from older cohorts have less frequently participated in sports in general and also at the county or national level. Athlete status in women is a heritable trait: around 66% of the variance in athlete status is explained by additive genetic factors. The remaining variance is due to non-shared environmental factors. Non-additive genetic factors and shared environmental factors do not play a role in explaining individual differences in athlete status.

The heritability of athlete status is consistent with the significant heritabilities that have been reported for a range of measures of athletic abilities. For example, the heritabilities of different measures of oxygen uptake, such as maximal oxygen uptake in the sedentary state and in response to training and submaximal oxygen uptake at different power outputs, which are all indicators of aerobic capacity, ranged from 47 to 74% (Bouchard et al., 1998; Bouchard et al., 1999). Heritabilities of muscle strength phenotypes (muscle cross-sectional area and isometric, concentric and eccentric strength) ranged from 31 to 85% (Thomis et al., 1998), with the majority of the genetic variance shared among the strength phenotypes (De Mars et al., 2007).

Suggestive linkage for athlete status was found on chromosome 3q24 (LOD=2.35) at the SLC9A9 gene. This gene encodes a cell membrane transport protein that is part of the sodium/hydrogen exchange family and is expressed at high levels in heart and skeletal muscle. SLC9A9 has been related to attention deficit hyperactivity disorder (de Silva et al., 2003) but not to sports ability. We also examined whether there were genes listed in the 'human gene map' that were located at or nearby the peak on 3q24 (Wolfarth et al., 2005). The calcium-sensing receptor (CASR) gene is located just outside the confidence interval of the peak on 3q21. The CASR gene has been found to be associated with physical activity (hours per week during the last year) in adolescent girls, and is involved in calcium homeostasis (Lorentzon et al., 2001). It is well-known that calcium is involved in muscle activity, bone formation, blood clotting and nerve activity (Wilmore & Costill, 2004).

Suggestive linkage was also found on chromosome 4q32 (LOD=1.87). There are two genes that are located in the region 4q28-31 that are listed in the 'human gene map'. The first is the fatty-acid binding protein 2 (FABP2) gene. In a linkage study carried out in 412 sibling pairs using 289 markers (Bouchard et al., 2000), suggestive linkage ($p < 0.01$) was found on

4q26 at marker FABP2 for maximal oxygen uptake in response to a 20 week exercise training program. Fatty-acid binding proteins are involved in the metabolism of long-chain fatty-acids. Through its involvement in metabolic processes, one could hypothesize that the FABP2 gene is related to sports performance and participation. The second gene that is located in the 4q28-31 region that is listed in the 'human gene map' is the uncoupling protein 1 (UCP1) gene. The UCP1 gene is involved in heat generation. In a sample of 395 offspring and 372 parents using 432 markers, suggestive linkage ($p < 0.01$) was found at the UCP1 gene for moderate to strenuous daily physical activity (Simonen et al., 2003b). To conclude, the peaks on 3q and 4q reported in the present study are of potential interest because they are close to genes that have been previously related to aerobic fitness and physical activity.

A first limitation of this study is that we used a very general measure of athlete status. Because the grouping also reflected the contrast between regular exercisers and sedentary subjects, athlete status was partly confounded with voluntary exercise participation. Substantial heritability for leisure-time exercise participation (Simonen et al., 2002; Stubbe et al., 2006a), has been found, ranging from 16 to 71%. Genes for athletic status may well constitute part of the heredity of voluntary regular exercise behaviour. Individuals with the genetic advantage to perform well at sports are more likely to continue participating in sports. They find that they are good at sports and they possibly also experience the physical need to be active, which together drives them to regularly participate in exercise. A GxE scenario may unfold where regular participation in exercise in those individuals 'who are good at it' provides the additional training that is needed to maximize actual performance, ultimately leading to elite athlete status (Brutsaert & Parra, 2006).

A second limitation is that the sports for which the twins could indicate their maximum level of competition were biased in favor of endurance sports requiring mainly aerobic fitness rather than strength or sprint capacity. The genes affecting endurance or power athlete status might be different or different alleles of the same gene might be associated with endurance or power athlete status, the latter having been suggested for the ACE gene (Williams et al., 2000), and more recently also for the ACTN3 gene (MacArthur et al., 2007). It is increasingly acknowledged however that the distinction between endurance and power sports is not that sharp; many endurance sports also contain power elements and vice versa (Wilmore & Costill, 2004).

A third limitation of the study involves the inclusion of women from different ages in the sample. One could argue that the younger women (less than 30 years) might not have reached their maximum level of competition yet, which could have affected the linkage results. Still, the prevalence of competition at the county or national level is highest in the younger adults, also in women aging between 20 and 25 years. The prevalence of competition at the county or national level is smallest in the oldest cohorts. The significant cohort effect on the prevalence of sporting ability can be explained by the fact that in the past it was less common for women to participate in sports at a competitive level. This cohort effect is unlikely to have biased the results of the genetic analyses however, since we regressed out the cohort effect in the analyses. It is possible, though, that the heritability of athlete status is higher in older cohorts since it was rarer, but we do not have the power to detect this.

A final limitation of this study relates to the sample size for the linkage scan. It is well-known that in order to detect linkage signals large sample sizes are needed, especially when categorical phenotypes are analysed (Neale et al., 1994a). We therefore investigated the power to detect QTL effects of different magnitudes, given a sample size of 700 sibling pairs and a total heritability of 65%. The power to detect a QTL effect of 30% of the total variance is 0.91 and the power to detect a QTL effect of 25% of the variance is 0.79. For QTL effects explaining, respectively, 20%, 15% and 10% of the variance the power is 0.60, 0.38 and 0.20. Thus, the power to detect linkage signals of 25% or larger is good and the power to detect smaller effects is moderate to low. The limited power of the present study underlines the need of replication studies.

To summarize, it was shown that athlete status in women is a heritable trait. A genome-wide linkage scan revealed suggestive linkages on chromosomes 3q22-q24 and 4q31-q34. The peak at 3q22-q24 is found at the SLC9A9 gene. The region 4q31-q34 overlaps with the region for which suggestive linkages were found in two previous linkage studies for physical fitness (FABP2 gene) (Bouchard et al., 2000) and physical activity (UCP1 gene) (Simonen et al., 2003b). Future association studies should further clarify the possible role of these genes in athlete status.





7

Genome-wide association study of exercise behavior in Dutch and American adults

This chapter is submitted as:

De Moor, M.H.M., Liu, Y-J., Boomsma, D.I., Li, J., Hamilton, J.J., Hottenga, J.J., Levy, S., Liu, X-G., Pei, J-F., Posthuma, D., Recker, R.R., Sullivan, P.F., Wang, L., Willemsen, G., Yan, H., De Geus, E.J.C. & Deng, H-W. Genome-wide association study of exercise behavior in Dutch and American adults. *Medicine and Science in Sports and Exercise*, under review.

Abstract

Objective The aim of this study was to identify genetic variants that are associated with adult leisure-time exercise behavior via the method of genome-wide association in two independent samples.

Methods Exercise behavior was measured in 1,772 unrelated Dutch adults from the Netherlands Twin Register and 978 unrelated American adults with detailed questions about type, frequency and duration of exercise. Individuals were classified into regular exercisers or non-exercisers using a threshold of 4 METhours (metabolic equivalents*hours per week). Regular exercisers were further divided into 5 categories of METhours, ranging from moderate (≥ 4 METhours) to highly vigorous (≥ 40 METhours) exercisers. After rigorous quality control, a total of 470,719 SNPs was available in the combined sample. Genome-wide association analyses were conducted using SNPtest using regression-based techniques, including sex, age and BMI as covariates.

Results Several novel SNPs located in the genes SGIP1, DNASE2B, PRSS16, ERCC2 and PAPSS2 were associated with exercise participation (combined p-value between 0.0004 and 4.5×10^{-6}). Associations of candidate genes based on existing literature were replicated for the LEPR gene in the American sample (rs12405556, $p=0.0005$) and for the CYP19A1 gene in the Dutch sample (rs2470158, 0.0098). The associations were independent of sex, age and BMI.

Conclusion Two of the genes found (SGIP1 and LEPR) are expressed in the hypothalamus and involved in the regulation of energy homeostasis. Their effects were independent of BMI, suggesting a direct role of hypothalamic factors in the drive to exercise.

A sedentary lifestyle is an important risk factor for a variety of physical health problems, such as obesity, cardiovascular disease, type II diabetes and osteoporosis (Albright et al., 2000; Berlin & Colditz, 1990; Martinez-Gonzalez et al., 1999; Todd & Robinson, 2003). Although lack of exercise participation is generally considered as a modifiable risk factor, twin studies have found that genetic factors play a substantial role in adult exercise participation (Beunen & Thomis, 1999; Stubbe et al., 2005; Stubbe et al., 2006a), suggesting that not all individuals have the same intrinsic drive to participate and persist in exercise. In a large study conducted in over 85,000 adult twins from seven different countries (Stubbe et al., 2006a) it was shown that between 48 to 71% of the variance in adult exercise behavior is explained by genetic factors. The remaining variance is accounted for by environmental factors that are not shared within a twin pair.

The heritability of exercise behavior in adults has been well established but not much is known about the actual genetic variants that are associated with this trait. So far, three genome-wide linkage studies and six candidate gene association studies have been conducted for exercise behavior or related physical activity phenotypes. The first linkage study for exercise behavior was conducted in 395 Caucasian adults plus their parents from 207 families (Simonen et al., 2003b). Four physical activity phenotypes were measured, of which three reflected daily physical activity and one past year physical activity. For past year physical activity, suggestive linkage ($p < 0.01$) was found on chromosomes 11p15 and 15q13.3. For daily physical activity promising linkage ($p < 0.0023$) was found on chromosome

2p22-p16 and suggestive linkages were found for different loci on chromosomes 4q28.2, 7p11.2, 9q31.1, 13q22-q31 and 20q13.1. The second study consisted of 1,030 children and 631 parents from 319 Hispanic-American families (Cai et al., 2006). Significant linkage was found on chromosome 18q12-q21 (LOD=4.09) for daily physical activity. The third study was conducted in 1,432 adult Dutch sibling pairs from 622 families (De Moor et al., 2007a). Suggestive linkage was found for regular exercise participation on chromosome 19p13.3 (LOD=2.18).

Six studies tested for association of genetic variation in a number of candidate genes with exercise or physical activity phenotypes. In a study of adolescent girls (Lorentzon et al., 2001), the calcium sensing receptor (CASR) gene was associated with weekly hours spent on physical activities. In a study of Pima Indians (Stefan et al., 2002), the leptin receptor (LEPR) gene was associated with 24 hour energy expenditure and physical activity levels. In another study (Simonen et al., 2003a), the dopamine 2 receptor (DRD2) gene was associated with past year physical activity in women reporting European ancestry but not in subjects reporting African ancestry. In a sample of postmenopausal women (Salmen et al., 2003), the aromatase (CYP19) gene was associated with physical activity. In a study of mild hypertensives (Winnicki et al., 2004), the angiotensin-converting enzyme (ACE) gene was associated with leisure-time physical activity. Finally, a study conducted in adults (Loos et al., 2005) showed that the melanocortin-4 receptor (MC4R) gene was associated with daily physical activity levels, independent of sex, age and BMI. The MC4R gene is located on chromosome 18, in the same region for which a significant linkage with child physical activity has been found (Cai et al., 2006).

This study is the first to report the results of a genome-wide association study for leisure-time exercise behavior, conducted in two independent samples comprising 1,772 Dutch and 978 American subjects genotyped on 470,719 SNP markers that passed quality controls. The aims of the study are 1) to identify new genetic variants that are associated with leisure-time exercise behavior, 2) to replicate the associations to previously reported candidate genes and linkage regions.

Methods

Subjects

The Netherlands

Dutch data on leisure-time exercise behavior were obtained from an ongoing longitudinal study (1991-2004) on health, lifestyle and personality in twins and their family members registered at the Netherlands Twin Register (NTR) (Boomsma et al., 2002; Boomsma et al., 2006b). A total of 1,860 unrelated individuals registered at the NTR were selected to be genotyped as part of the Genetic Association Information Network (GAIN) initiative (Manolio et al., 2007), of which 1,703 served as controls and 160 as cases in a genome-wide association study for major depressive disorder (GAIN-MDD) (Boomsma et al., 2008; Sullivan et al., 2008). The study was approved by the Central Ethics Committee on Research Involving Human Subjects. All subjects provided written informed consent. After quality control of the genotype data (Sullivan et al., 2008), data from 1,777 individuals were left for analysis, of whom 5 did not have valid data on exercise. For the remaining 1,772 individuals, we used their most recent exercise data. For 1,204 individuals (67.9%), data came from a survey sent out in 2004, for 325 individuals (18.3%) data came from a survey from 2002 and for the remaining individuals data came from earlier surveys (1991-2000). Mean age of the participants was 43.5 (sd=14.6, range=14.5-79.8) at the time of the survey collection. Mean BMI (defined as weight (kg) / height (m)²) was 24.3 (sd=3.6, range=14.3-42.0). There were 649 men (36.5%) and 1,128 women (63.5%).

United States of America

American data on leisure-time exercise behavior were collected as part of a larger study into the genetics of common human complex diseases/traits (e.g., osteoporosis, obesity, and height) in normal healthy subjects (Deng et al., 2002). The study was approved by the necessary Institutional Review Boards of involved institutions. Signed informed-consent was obtained from all study subjects before they entered the study. A random sample containing 978 unrelated Caucasian subjects was identified from our ongoing study currently containing more than 6,000 individuals. All of the chosen subjects were of US Caucasians of Northern European origin living in Omaha, Nebraska and its surrounding areas. The inclusion and exclusion criteria were well defined (Deng et al., 2002). Briefly, subjects with chronic diseases and conditions involving vital organs (heart, lung, liver, kidney, and brain) and severe endocrinological, metabolic, and nutritional

diseases were excluded from this study. Mean age of the participants was 50.0 (sd=18.3, range=19.1-87.2) at the time of the survey collection. Mean BMI was 27.3 (sd=5.2, range=14.2-49.4). There were 494 men (50.5%) and 484 women (49.5%).

Phenotypes

Leisure-time exercise behavior was measured in a comparable way in the Dutch and American samples, except for a minor difference in the first question asked. In the Dutch sample, the first question was “Do you participate in exercise regularly?” and in the American question this was “Do you take exercise for 60 minutes per week?”. These questions could be answered with ‘Yes’ or ‘No’. If the participants responded affirmative, they were asked to list all exercise activities, and to indicate on type, frequency and duration of each activity. All exercise activities were assigned a metabolic equivalent (MET) value according to the widely accepted Ainsworth’s Compendium of physical activity (Ainsworth et al., 2000). A MET score of 1 corresponds to the rate of energy expenditure when at rest (1 kcal/kg/h). Reported non-leisure time activities (biking to work, gardening, and work-related physical activity) were not counted as leisure time exercise. Over the remaining exercise activities, the total METhours were computed as MET*hours/week. Scores of non-exercising individuals were coded as zero. To keep consistent with existing epidemiological studies (Haase et al., 2004), we classified individuals as regular versus non-exercisers, based on a minimal threshold of at least 4 METhours weekly. For ancillary analysis, we further classified regular exercisers into five categories: 4-12 METhours, 13-21 METhours, 22-30 METhours, 31-39 METhours and ≥ 40 METhours).

In the Dutch sample, all 1,772 individuals could be classified as regular (878 individuals, 49.5%) or non-exerciser (894 individuals, 50.5%). From the 878 regular exercisers, 817 individuals had valid data on type, frequency and duration of exercise and could be further classified into five categories of METhours: 4-12 METhours (385 individuals, 47.1%), 13-21 METhours (218 individuals, 26.7%), 22-30 METhours (98 individuals, 12.0%), 31-39 METhours (44 individuals, 5.4%) and ≥ 40 METhours (72 individuals, 8.8%).

In the American sample, the 978 individuals could be classified as regular (612 individuals, 62.6%) or non-exerciser (366 individuals, 37.4%). All the 612 regular exercisers had valid data on type, frequency and duration of exercise and could be further classified into five categories of METhours: 4-12 METhours (57 individuals, 9.3%), 13-21 METhours (329 individuals,

53.6%), 22-30 METhours (194 individuals, 31.7%), 31-39 METhours (19 individuals, 3.1%) and ≥ 40 METhours (14 individuals, 2.3%).

Genotypes

The Netherlands

DNA was extracted from frozen whole blood samples using the Puregene DNA Isolation kit (Gentra, Minneapolis, MN, USA). All procedures were performed according to the manufacturer's protocols. Genotyping was conducted by Perlegen Sciences, using a high-density oligonucleotide array-based platform (Sullivan et al., 2008). A total of 599,156 genotyped SNPs from 98.5% of all individuals participating in the GAIN-MDD study were returned. After careful quality control, 435,291 SNPs were left for analyses, of which 427,024 were autosomal SNPs. SNPs were excluded because of gross mapping errors (1,487 SNPs), duplicate errors (1,143), Mendelian inconsistencies (536 SNPs), minor allele frequency < 0.01 (41,495 SNPs) and missing genotypes > 0.05 (156,673 SNPs), or a combination of these reasons. More details on genotyping procedures, genotype calling, and quality control checks in the GAIN-MDD sample can be found in Sullivan et al. (2008).

United States of America

Genomic DNA was extracted from whole human blood using a commercial isolation kit (Gentra systems, Minneapolis, MN, USA) following the protocols detailed in the kit. Genotyping with the Affymetrix Mapping 250K Nsp and 250K Sty arrays was performed using the standard protocol recommended by the manufacturer. Out of the initial full-set of 500,568 SNPs, we discarded 32,961 SNPs with sample call rate $< 95\%$, another 36,965 SNPs with allele frequencies deviating from Hardy-Weinberg equilibrium (HWE) and 51,323 SNPs with minor allele frequencies (MAF) $< 1\%$. Therefore, the final SNP set maintained in the subsequent analyses contained 381,100 SNPs.

Genotype imputation

From the 427,024 successfully genotyped autosomal SNPs in the Dutch sample and the 381 100 successfully genotyped autosomal SNPs in the American sample, there were 101,067 overlapping SNPs (14.3%) out of a total of 707,057 SNPs on both platforms. In order to be able to compare results at the SNP level, we imputed all SNPs that were observed in the Dutch sample but unobserved in the American sample and vice versa. Thus, 325,957 SNPs were imputed in the American sample (46.1%) and

280,033 SNPs were imputed in the Dutch sample (39.6%). Imputation was carried out in IMPUTE (Marchini et al., 2007) using the HAPMAP phase II data available on the IMPUTE website <http://www.stats.ox.ac.uk/~marchini/software/gwas/impute.html#>. IMPUTE computes the probabilities of each of the three possible genotypes for each unobserved SNP for each individual in the sample using the information of the surrounding observed genotypes for that individual and the LD information available from the HAPMAP data. IMPUTE also gives information about how well each SNP is imputed, quantified as the maximum posterior call averaged over all individuals for each SNP. The maximum posterior call averaged over all individuals over all SNPs that were imputed in the Dutch sample was 0.98 (median=0.99, minimum=0.44, maximum=1). 98.1% of the imputed SNPs had an average maximum posterior call of 0.90 or larger and 99.5% of 0.80 or larger. The mean, median, minimum and maximum average posterior call of all imputed SNPs in the American sample was, respectively, 0.90, 0.94, 0.46 and 1. 64.5% of the imputed SNPs had an average maximum posterior call of 0.90 or larger and 85.2% of 0.80 or larger. Thus, the quality of the imputation was very good in the American and Dutch samples.

Statistical analysis

Genome-wide association analyses were conducted using logistic regression for the exercise participation phenotype in SNPtest software (Marchini et al., 2007). We used WGAviewer to create Manhattan and QQ plots and to annotate SNPs (Ge et al., 2008). Analyses were performed in both samples independently while taking the uncertainty of the imputed genotypes into account and including sex and age as covariates. Association of each SNP with the exercise participation phenotype was tested using the 2 degrees of freedom genotypic test. SNPtest outputs a p-value for all SNPs regardless of the number of individuals observed for each genotype. However, it is well-known that the genotypic test only performs well if there are a sufficient number of individuals within each genotype group. Therefore, we excluded within each sample all SNPs with less or equal than 5 individuals in at least one of the genotype groups, while taking a genotype calling threshold of 0.90 for the imputed SNPs. Given the size of our datasets, this roughly corresponded to excluding all SNPs with minor allele frequency smaller than 0.05 in the Dutch sample (9.3%) and 0.07 in the American sample (26.1%). After this, SNPs that were not well imputed (average maximum posterior probability < 0.80) in one of the samples were excluded in that sample (1.3% for the NTR and 2.9% for

the American sample). This resulted in 632 044 SNPs in the Dutch sample (224,229 imputed and 407,815 observed) and 501 859 SNPs in the American sample (190,253 imputed and 311,606 observed). The overlap between the two samples was 470,719 SNPs that survived QC in both samples.

Novel gene finding

On the remaining 470,719 SNPs we computed the combined p-values across the two samples for each phenotype using Fisher's method in Haploview (Barrett et al., 2005). To identify SNPs that might contribute to the heritability of exercise behavior we selected SNPs that 1) reached the genome-wide level for suggestive association ($p < 1/500,000 = 2.0 \times 10^{-6}$) based on the combined p-value, or 2) that were nominally significant ($p < 0.01$) in both samples. To protect against false positives, an additional criterion for both types of SNPs was that the direction of the effects was similar in the Netherlands and the USA.

Candidate gene replication

We inspected the association of all SNPs located in or in close vicinity (<10kb) of all candidate genes for physical activity phenotypes from previous candidate gene association studies, including the associated SNP reported in those studies if available. We used a replication strategy at the gene level, i.e. any SNP (not just the original SNP reported on) reaching a combined p-value smaller than 0.01 was considered to constitute a replication.

Linkage region replication

We inspected association of all SNPs located in the 95% confidence interval of all linkage peaks previously reported for exercise behavior and physical activity phenotypes. Because this involved thousands of SNPs with low average pairwise LD, only SNPs reaching a Bonferroni-corrected p-value of $0.01 / \text{number of SNPs tested} \approx 1.0 \times 10^{-5}$ in the combined sample was considered to be a replication.

For both novel identified SNPs as well as replicated SNPs, we performed checks for Hardy-Weinberg equilibrium (HWE tested at $p < 1.0 \times 10^{-5}$), inspection of genotype calling cluster plots, and (if imputed) average maximum posterior call. Finally, to bolster our confidence in the relevance of the SNP for exercise behavior we performed an additional association analyses in which the weekly METhours phenotype was regressed on the SNP genotype, using a combined $p < 0.01$ as a criterion for significance.

Results

Novel gene finding

The Manhattan plot for exercise participation in the combined sample is given in Figure 7.1. The QQ plot is given in Figure 7.2. A list of the 20 most significant SNPs for exercise participation in the NTR and in the American sample are given, respectively, in Tables 7.1 and 7.2. In the Dutch sample, the lowest p-value was $2.8 \cdot 10^{-6}$. In the American sample, the lowest p-value was $7.9 \cdot 10^{-7}$. Table 7.3 displays the results of all SNPs that reach the threshold for genome-wide suggestive association (combined $p < 2.0 \cdot 10^{-6}$) or that were nominally significant in both samples ($p < 0.01$). The lowest combined p-value was $5.9 \cdot 10^{-7}$. Only the 12 highlighted SNPs in table 3 reached our a priori criteria of genome-wide significance in the combined sample or a nominal significance in both samples with the same direction of the allelic effects. Four of these SNPs are located in introns of genes; rs9633417 is located in the SH3-domain GRB2-like (endophilin) interacting protein 1 (SGIP1) gene, rs667923 in the deoxyribonuclease II beta (DNASE2B) gene, rs10946904 in the protease serine 16 (PRSS16) gene and rs238404 in the excision repair cross-complementing rodent repair deficiency, complementation group 2 (ERCC2) gene. SNP rs2762527 is located upstream of the 3'-phosphoadenosine 5'-phosphosulfate synthase 2 (PAPSS2) gene. The other seven SNPs are intergenic.

Figure 7.1. Manhattan plot of exercise participation for the combined p-values

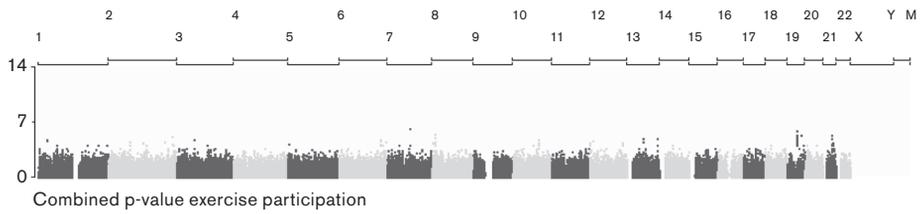


Figure 7.2. QQ plot of exercise participation for the combined p-values

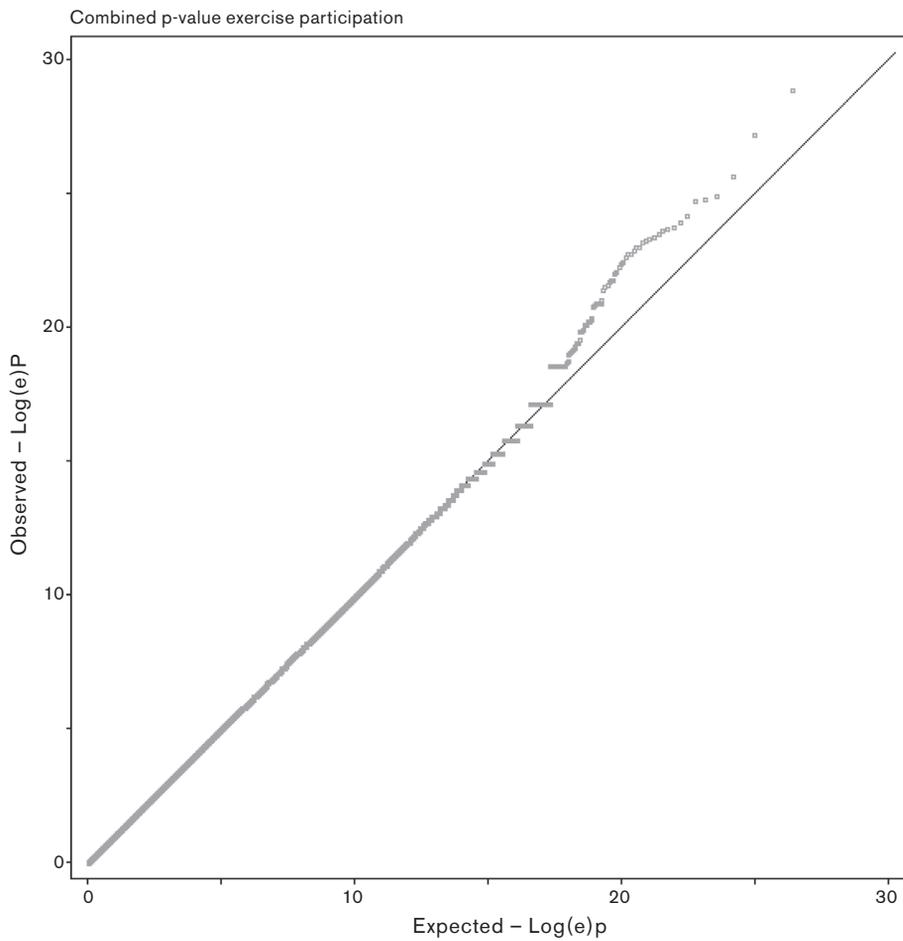


Table 7.1. List of top 20 most significant SNPs for exercise participation in the Netherlands Twin Registry sample

SNP	Chrom	Bp	Alleles	P-value	β -1	β -2	Maf	Associated gene	Location	P-value American sample
rs10803261	1	241596482	C/G	3.2×10^{-5}	-0.72	-0.89	0.27	EFCAB2	Intronic	0.37
rs11678571	2	47140345	C/G	3.8×10^{-5}	-0.79	-0.69	0.31	TTC7A	Intronic	0.65
rs10193648	2	134462763	A/G	1.6×10^{-5}	-1.12	-0.89	0.22	-	Intergenic	0.61
rs10197646	2	135596388	A/G	2.7×10^{-5}	-1.66	-1.45	0.14	YSK4	Intronic	0.97
rs2159724	2	229817726	A/G	2.9×10^{-5}	-0.17	0.47	0.41	PID1	Intronic	0.29
rs6550329	3	35211525	G/T	2.7×10^{-5}	-0.48	-0.23	0.25	-	Intergenic	0.50
rs3843917	6	154732923	C/G	2.1×10^{-5}	0.40	0.59	0.38	NP_056968.1	Intronic	0.42
rs2891117	9	17977408	C/T	2.8×10^{-6}	0.59	0.58	0.48	SH3GL2	Intergenic	-
rs7919331	10	5541181	A/G	1.5×10^{-5}	-0.03	-0.64	0.41	O6ZR35_HUMAN	Intergenic	0.07
rs8181407	10	72194444	C/T	1.5×10^{-5}	-0.35	0.73	0.19	ADAMTS14	Downstream	0.67
rs17311303	11	119882018	A/G	2.5×10^{-5}	0.82	0.43	0.34	Q3C1X2_HUMAN	Intergenic	0.48
rs631750	12	3946116	C/T	3.0×10^{-5}	0.19	-0.52	0.35	PARP11	Intergenic	0.82
rs11829154	12	3947492	A/G	3.4×10^{-5}	0.65	0.33	0.34	PARP11	Intergenic	-
rs7971136	12	12051742	A/G	2.8×10^{-6}	0.01	0.66	0.41	BCL2L14	Intergenic	0.57
rs7982922	13	42771336	A/G	2.5×10^{-5}	1.85	1.58	0.11	ENOX1	Intronic	0.80
rs10500278	19	43186344	A/G	2.3×10^{-5}	0.69	0.81	0.30	SIPA1L3	Intronic	0.68
rs1433089	19	57198797	C/T	2.2×10^{-5}	-0.57	-0.08	0.16	ZNF615	5prime_utr	0.49
rs2070513	21	34275744	G/T	1.6×10^{-5}	0.46	-0.03	0.28	O6ZRX0_HUMAN	Intergenic	0.20
rs2834315	21	34277280	G/T	2.4×10^{-5}	0.45	-0.08	0.28	O6ZRX0_HUMAN	Intergenic	0.31
rs11088265	21	34277488	A/G	2.5×10^{-5}	0.53	0.09	0.28	O6ZRX0_HUMAN	Intergenic	0.24

Chrom=Chromosome, Bp=Basepair position, β -1 = the log-odds ratio to be an exerciser of the heterozygous genotype group compared with the first homozygous genotype group (for example CG versus CC for the first SNP in the table), β -2 = the log-odds ratio to be an exerciser of second homozygous genotype group compared with the heterozygous genotype group (for example GG versus CG for the first SNP in the table), Maf=Minor allele frequency. Allele in bold is the minor allele. Note that sex and age were included as covariates in these analyses.

Table 7.2. List of top 20 most significant SNPs for exercise participation in the American sample

SNP	Chrom	Bp	Alleles	P-value	β -1	β -2	Maf	Associated gene	Location	P-value Dutch sample
rs7521315	1	5195141	A/G	1.9*10 ⁻⁵	-0.33	-0.91	0.36	-	Intergenic	0.95
rs878465	1	30916121	A/G	1.9*10 ⁻⁵	-0.60	0.07	0.49	LAPTM5	Intergenic	0.07
rs912216	1	30916182	A/G	1.9*10 ⁻⁵	-0.67	-0.07	0.49	LAPTM5	Intergenic	0.08
rs10737353	1	30920144	A/C	1.4*10 ⁻⁵	-0.66	-0.07	0.49	LAPTM5	Intergenic	0.08
rs17065829	3	62018255	C/G	1.3*10 ⁻⁵	1.82	1.02	0.11	PTPRG	Intronic	0.78
rs32732	5	3509763	A/G	2.0*10 ⁻⁵	-0.68	-1.25	0.12	IRX1	Intergenic	0.27
rs7743055	6	145718922	C/T	1.3*10 ⁻⁵	0.73	1.20	0.11	EPM2A	Intergenic	0.63
rs4896792	6	145727562	A/T	2.1*10 ⁻⁶	-0.92	-1.64	0.12	EPM2A	Intergenic	0.67
rs4896794	6	145730006	C/G	2.0*10 ⁻⁶	0.72	1.63	0.12	EPM2A	Intergenic	0.70
rs6926799	6	145741712	A/G	7.1*10 ⁻⁶	-1.22	-1.91	0.11	EPM2A	Intergenic	0.65
rs6561985	13	58290774	C/G	4.4*10 ⁻⁶	-0.45	1.10	0.20	-	Intergenic	0.53
rs2321884	13	58349733	C/T	2.1*10 ⁻⁶	-0.41	1.30	0.19	-	Intergenic	0.36
rs7989735	13	109160955	C/T	2.4*10 ⁻⁵	-0.42	0.55	0.35	IRS2	Intergenic	0.03
rs17773922	19	34522469	C/T	2.8*10 ⁻⁵	0.90	0.48	0.38	UQCRL1	Intergenic	0.003
rs13051865	21	30531338	C/T	7.9*10 ⁻⁷	0.69	0.78	0.40	CLDN8	Intergenic	0.34
rs2832691	21	30537456	C/T	1.8*10 ⁻⁶	-0.09	-0.78	0.41	CLDN8	Intergenic	0.46
rs720901	21	30537905	A/G	2.2*10 ⁻⁵	-0.02	-0.64	0.36	CLDN8	Intergenic	0.72
rs2832700	21	30542823	A/T	1.6*10 ⁻⁵	0.63	0.66	0.36	CLDN8	Intergenic	0.83
rs2832741	21	305564220	A/C	1.9*10 ⁻⁶	-0.10	-0.78	0.41	KR241_HUMAN	Intergenic	0.52
rs2835881	21	37955001	A/G	1.7*10 ⁻⁵	-0.82	-0.96	0.46	KCNJ6	Intronic	0.83

Chrom=Chromosome, Bp=Basepair position, β -1 = the log-odds ratio to be an exerciser of the heterozygous genotype group compared with the first homozygous genotype group, β -2 = the log-odds ratio to be an exerciser of second homozygous genotype group compared with the heterozygous genotype group, Maf=Minor allele frequency. Allele in bold is the minor allele. Note that sex and age were included as covariates in these analyses.

Table 7.3. List of SNPs for exercise participation that reach threshold for genome-wide suggestive association (combined $p < 2.0 \times 10^{-6}$) or that are nominally significant in both samples ($p < 0.01$)

SNP	Chr	Bp	Alleles	P-value combined	P-value Dutch sample	P-value American sample	β -1 NTR	β -2 NTR	β -1 USA	β -2 USA	Maf NTR	Maf USA	Associated gene	Location
rs9633417	1	66883433	A/C	0.0001	0.003	0.003	-1.65	-1.86	-2.41	-2.11	0.09	0.10	SGIP1	Intronic
rs667923	1	84583998	A/G	0.0003	0.006	0.005	0.72	0.46	0.18	0.61	0.18	0.19	DNASE2B	Intronic
rs1766581	1	228754028	C/T	0.0008	0.009	0.009	0.25	0.40	0.06	0.58	0.44	0.44	SIPA1L2	Intergenic
rs4355145	2	75867660	A/G	0.0003	0.009	0.003	-0.18	-0.46	-1.09	-0.83	0.22	0.20	C2orf3	Intergenic
rs9789774	2	75905098	A/G	2.9×10^{-5}	0.005	0.0004	-0.26	-0.53	-1.23	-0.95	0.22	0.20	C2orf3	Intergenic
rs3935887	2	75993336	A/G	0.0001	0.009	0.001	0.05	-0.26	-0.75	-0.53	0.29	0.28	C2orf3	Intergenic
rs1829618	2	212258831	A/G	0.0002	0.004	0.004	-0.16	0.19	-1.27	-1.01	0.20	0.17	ERBB4	Intronic
rs10203022	2	212284296	C/T	0.0004	0.004	0.009	0.25	0.59	-1.43	-1.06	0.13	0.13	ERBB4	Intronic
rs7601664	2	212318270	A/G	0.0005	0.006	0.007	-0.36	-0.42	-0.39	1.05	0.13	0.12	ERBB4	Intronic
rs13013897	2	228973701	A/G	6.1×10^{-6}	0.002	0.0002	0.28	-0.07	0.10	-0.49	0.38	0.38	SKIP	Intergenic
rs1283098	3	108450567	A/G	7.5×10^{-5}	0.001	0.006	-0.21	0.26	0.20	0.61	0.45	0.45	CCDC54	Intergenic
rs2593943	3	110172911	A/G	0.0007	0.009	0.007	-0.05	0.28	0.39	-0.04	0.42	0.42	MORC1	Nonsynon
rs10946904	6	27325215	C/T	0.0004	0.008	0.004	0.19	0.59	0.41	0.50	0.25	0.24	PRSS16	Intronic
rs7793555*	7	80191396	A/G	5.9×10^{-7}	0.0004	8.8×10^{-5}	-0.05	0.35	0.60	0.02	0.26	0.28	SEMA3C	Intronic
rs17837697	7	156120121	C/T	0.0004	0.004	0.009	-0.06	0.28	-0.81	-0.73	0.26	0.27	LMBR1	Intronic
rs1869075	8	530949	A/G	5.2×10^{-5}	0.008	0.0005	0.30	-0.01	-0.90	-0.57	0.32	0.30	Q8NF75_HUMAN	Intergenic
rs330075	8	9195294	A/G	8.0×10^{-6}	0.001	0.0004	0.39	0.19	-0.42	0.35	0.40	0.38	Q6XYA8_HUMAN	Intergenic
rs330069	8	9199283	A/G	7.9×10^{-5}	0.004	0.002	0.23	-0.13	-0.64	-0.27	0.41	0.38	Q6XYA8_HUMAN	Intergenic
rs330068	8	9199458	A/G	3.0×10^{-6}	0.0005	0.0004	0.28	-0.13	-0.84	-0.50	0.37	0.33	Q6XYA8_HUMAN	Intergenic

Continuation of Table 7.3.

rs6988046	8	9985843	A/C	3.2*10 ⁻⁵	0.008	0.0003	-0.16	-0.67	0.04	1.33	0.22	0.23	MSRA	Intronic
rs1149996	10	9395673	A/G	0.0004	0.006	0.005	0.28	0.40	0.06	-0.53	0.43	0.42	Q6ZMVO_HUMAN	Intergenic
rs1149997	10	9395973	C/T	0.0004	0.005	0.008	-0.10	-0.40	0.56	0.52	0.43	0.42	Q6ZMVO_HUMAN	Intergenic
rs1150001	10	9396637	A/G	0.0004	0.006	0.007	0.28	0.40	0.05	-0.53	0.43	0.42	Q6ZMVO_HUMAN	Intergenic
rs1150002	10	9396735	A/G	0.0005	0.006	0.008	0.28	0.40	0.03	-0.53	0.43	0.42	Q6ZMVO_HUMAN	Intergenic
rs10827786	10	37847539	A/G	0.0002	0.002	0.008	1.44	1.08	0.26	0.71	0.09	0.11	ZNF248	Intergenic
rs2762527	10	89408147	G/T	0.0001	0.005	0.002	-0.33	-0.30	-0.25	-0.72	0.37	0.39	PAPSS2	Upstream
rs10509410	10	89429981	A/G	5.1*10 ⁻⁵	0.003	0.001	-0.06	0.32	0.10	0.61	0.46	0.47	PAPSS2	Intronic
rs7908056	10	89434630	A/G	1.8*10 ⁻⁵	0.003	0.0004	-0.06	0.32	0.10	0.64	0.46	0.46	PAPSS2	Intronic
rs1626521	11	73391987	A/G	0.0007	0.007	0.009	-0.47	-0.19	0.16	-0.23	0.27	0.24	UCP3	Intronic
rs12101846	15	59358620	G/T	0.0003	0.003	0.008	-0.35	0.04	-0.11	0.67	0.30	0.29	RORA	Intergenic
rs12905612	15	91581352	A/T	0.0002	0.004	0.004	0.12	0.55	0.45	0.07	0.32	0.34	Q6UXP9_HUMAN	Intergenic
rs9944863	18	1590972	C/T	0.0002	0.002	0.0096	-0.35	-0.62	0.22	-0.22	0.23	0.22	C18orf2	Intergenic
rs2166030	18	61340512	C/T	0.0003	0.002	0.0096	-0.22	0.17	0.18	0.48	0.45	0.44	CDH7	Intergenic
rs8108111	19	34501691	A/G	3.2*10 ⁻⁵	0.002	0.001	-0.19	0.27	0.32	-0.36	0.43	0.40	UQCRFSL1	Intergenic
rs10415631	19	34502402	C/G	9.1*10 ⁻⁵	0.002	0.004	-0.47	-0.28	0.62	0.35	0.43	0.40	UQCRFSL1	Intergenic
rs10423930	19	34502527	G/T	4.2*10 ⁻⁵	0.002	0.002	-0.47	-0.28	0.65	0.34	0.43	0.41	UQCRFSL1	Intergenic
rs10425007	19	34504237	C/T	4.6*10 ⁻⁵	0.002	0.002	-0.47	-0.28	0.64	0.33	0.43	0.40	UQCRFSL1	Intergenic
rs10424492	19	34505271	A/G	5.2*10 ⁻⁵	0.002	0.002	-0.47	-0.28	0.64	0.34	0.43	0.40	UQCRFSL1	Intergenic
rs7253662	19	34507562	C/T	0.0001	0.002	0.006	-0.19	0.28	0.32	-0.22	0.43	0.42	UQCRFSL1	Intergenic
rs10423259	19	34511945	A/G	4.6*10 ⁻⁵	0.002	0.002	-0.45	-0.28	0.63	0.28	0.44	0.41	UQCRFSL1	Intergenic
rs8113496	19	34512369	A/G	7.0*10 ⁻⁵	0.002	0.002	-0.17	0.28	0.35	-0.25	0.44	0.41	UQCRFSL1	Intergenic
rs8102393	19	34512400	C/T	3.3*10 ⁻⁵	0.002	0.001	-0.45	-0.28	0.65	0.29	0.44	0.41	UQCRFSL1	Intergenic
rs8107784	19	34516147	C/T	9.7*10 ⁻⁶	0.003	0.0002	-0.17	0.30	0.43	-0.28	0.42	0.41	UQCRFSL1	Intergenic

Continuation of Table 7.3.

SNP	Chr	Bp	Alleles	P-value combined	P-value Dutch sample	P-value American sample	β -1 NTR	β -2 NTR	β -1 USA	β -2 USA	Maf NTR	Maf USA	Associated gene	Location
rs1122022	19	34517054	A/G	9.1×10^{-6}	0.002	0.0003	-0.45	-0.28	0.73	0.35	0.44	0.41	UQCRFSL1	Intergenic
rs17773922*	19	34522469	C/T	1.4×10^{-6}	0.003	2.8×10^{-5}	-0.44	-0.26	0.90	0.48	0.44	0.38	UQCRFSL1	Intergenic
rs10415713	19	34523381	C/T	4.7×10^{-6}	0.002	0.0001	-0.17	0.28	0.41	-0.33	0.44	0.41	UQCRFSL1	Intergenic
rs10423829	19	34524171	C/G	7.9×10^{-6}	0.002	0.0002	-0.45	-0.28	0.71	0.31	0.44	0.41	UQCRFSL1	Intergenic
rs10425546	19	34524238	C/T	9.9×10^{-6}	0.002	0.0003	-0.17	0.28	0.38	-0.32	0.44	0.41	UQCRFSL1	Intergenic
rs10406567	19	34524595	C/T	9.5×10^{-6}	0.002	0.0003	-0.45	-0.28	0.70	0.31	0.44	0.41	UQCRFSL1	Intergenic
rs13346397	19	34542618	A/T	7.0×10^{-6}	0.002	0.0002	-0.17	0.28	0.40	-0.33	0.44	0.41	UQCRFSL1	Intergenic
rs7508455	19	34542918	A/C	1.1×10^{-5}	0.002	0.0003	-0.17	0.28	0.38	-0.36	0.44	0.40	UQCRFSL1	Intergenic
rs7508469	19	34543002	A/C	1.3×10^{-5}	0.002	0.0004	-0.17	0.28	0.39	-0.30	0.44	0.41	UQCRFSL1	Intergenic
rs8107741	19	34550442	G/T	7.7×10^{-5}	0.002	0.003	-0.46	-0.28	0.58	0.22	0.44	0.42	UQCRFSL1	Intergenic
rs11880983	19	34551705	C/T	1.3×10^{-5}	0.002	0.0004	-0.46	-0.28	0.69	0.28	0.44	0.41	UQCRFSL1	Intergenic
rs6509387	19	34553547	A/T	2.4×10^{-5}	0.002	0.0007	-0.46	-0.28	0.68	0.33	0.44	0.41	UQCRFSL1	Intergenic
rs6509388	19	34553722	A/G	8.6×10^{-6}	0.004	0.0002	-0.17	0.28	0.41	-0.32	0.45	0.41	UQCRFSL1	Intergenic
rs1558364	19	34557185	A/G	7.6×10^{-6}	0.004	0.0001	-0.45	-0.28	0.75	0.34	0.45	0.40	UQCRFSL1	Intergenic
rs17624933	19	34559723	C/T	1.8×10^{-5}	0.004	0.0003	-0.18	0.27	0.40	-0.29	0.45	0.41	UQCRFSL1	Intergenic
rs238404	19	50557530	C/T**	4.5×10^{-6}	8.8×10^{-5}	0.003	0.26	-0.41	0.21	-0.32	0.49	0.45	ERCC2	Intronic
rs4816397	21	31766413	A/C	0.0001	0.009	0.010	0.63	0.67	-0.97	-1.15	0.22	0.22	TIAM1	Intronic

* = genome-wide suggestive association, ** = minor allele in Dutch sample is T and in American sample C. In grey: SNPs have same direction of effect in both samples. Allele in bold is the minor allele. Note that sex and age were included as covariates in these analyses. Chr=Chromosome, Bp=basepair position, Maf=Minor allele frequency. β -1 is the log-odds ratio to be an exerciser of the heterozygous genotype group compared with the first homozygous genotype group. β -2 is the log-odds ratio to be an exerciser of second homozygous genotype group compared with the heterozygous genotype group

Candidate gene replication

Table 7.4 presents the association of SNPs in the six candidate genes with exercise participation in the two samples. Two genes (the CYP19A1 and LEPR gene) reached our a priori criterion for replication. The lowest combined p-value in the CYP19A1 gene is 9.8×10^{-3} for SNP rs2470158. This SNP is nominally significant in the Dutch sample ($p=7.2 \times 10^{-3}$), but not in the American sample ($p=0.18$). There are no other SNPs in the CYP19A1 gene that are nominally significant. Inspecting the LD structure in this gene, it is apparent that SNP rs2470158 is not in high LD with any surrounding SNPs. The lowest combined p-value in the LEPR gene is 1.8×10^{-3} for SNP rs12405556, which is only 4604 basepair away from the significant SNP that was reported in the previous candidate gene study (rs1137101) (Stefan et al., 2002). SNP rs12405556 is in moderate LD with SNP rs1137101 ($r^2=0.44$). SNP rs12405556 was significant in the American sample ($p=5.0 \times 10^{-4}$) but not in the Dutch sample ($p=0.39$). In total, 14 of the tested SNPs in the LEPR gene have a p-value smaller than 0.01 in the American sample, but none of these SNPs were nominally significant in the Dutch sample.

Linkage region replication

Table 7.5 presents an overview of the most significant SNPs in the linkage regions that have been reported for exercise and physical activity phenotype in previous studies (Cai et al., 2006; De Moor et al., 2007a; Simonen et al., 2003b). None of the SNPs located in the linkage regions reached our a priori criterion for replication. Moreover, most SNPs with the lowest combined p-values were often not located near the peak markers.

Table 7.4. Testing association with 6 candidate genes for exercise participation

Candidate gene	Study	Chrom	Lowest p-value combined sample	SNP	#SNPs in gene
ACE	Winnicki et al (2004)	7q23.3	0.38	rs8066114	7
CASR	Lorentzon et al. (2001)	3q21.1	0.06	rs9831894	42
CYP19A1	Salmen et al. (2003)	15q21.2	0.0098	rs2470158	43
DRD2	Simonen et al. (2003a)	11q23	0.21	rs2234689	22
LEPR	Stefan et al. (2002)	1p31.3	0.002	rs12405556	59
MC4R	Loos et al. (2005)	18q21.32	0.10	rs9965495	5

Candidate gene	Coverage	Lowest p Dutch sample	SNP	Lowest p American sample	SNP
ACE	0.50	0.39	rs8066114	0.21	rs4611524
CASR	0.68	0.05	rs9831894	0.08	rs10934578
CYP19A1	0.70	0.007	rs2470158	0.18	rs2470158
DRD2	0.82	0.14	rs2470158	0.07	rs2471857
LEPR	0.88	0.01	rs11585329	0.0005	rs12405556
MC4R	0.69	0.10	rs17066829	0.04	rs9965495

Note: lowest p-value in gene or 10kb around gene. Chrom=Chromosome, Coverage = Number of observed SNPs in gene + Number of tagged common SNPs ($r^2 > 0.80$) divided by the total number of common SNPs in HAPMAP. In grey: $p < 0.01$, Note that sex and age were included as covariates in these analyses.

Table 7.5. Testing association in 10 linkage regions for exercise participation

Linkage region	Study	Marker at peak	Flanking markers*	Lowest p-value combined sample	SNP
19p13.3	De Moor et al. (2007)	D19S247	D19S591 - D19S865	0.005	rs2118706
18q12.2	Cai et al. (2006)	D18S1102	D18S1102 - D18S474	0.0009	rs10853477
18q21.2	Cai et al. (2006)	D18S474	D18S474 - D18S61	0.0003	rs2166030
2p22.3	Simonen et al. (2003b)	D2S2347	-	0.002	rs17010549
4q31.21	Simonen et al. (2003b)	UCP1	-	0.008	rs10440457
7p13-p12	Simonen et al. (2003b)	IGFBP1	-	0.0001	rs11763891
9q31	Simonen et al. (2003b)	D9S938	-	0.0004	rs11792745
13q22	Simonen et al. (2003b)	D13S317	-	0.002	rs7938327
15q13	Simonen et al. (2003b)	D15S165	-	0.0009	rs11161249
20q12	Simonen et al. (2003b)	PLCG1	-	0.0001	rs2425456

Linkage region	Distance of SNP to peak marker (bp)	#SNPs in region	Gene associated with SNP	Lowest p Dutch sample	Lowest p American sample
19p13.3	5.524.705	447	ADAMTS10	0.004	0.002
18q12.2	3.728.030	2702	-	0.0004	0.001
18q21.2	14.391.594	3814	CDH7	0.0004	0.0007
2p22.3	2.081.010	1269	GALNT14	0.34	0.0007
4q31.21	359.319	1021	SCOC	0.0009	0.004
7p13-p12	2.572.126	1230	AC004455.2	0.0003	7.8 *10 ⁻⁵
9q31	1.280.865	1313	SMC2	4.2*10 ⁻⁵	0.003
13q22	1.331.016	2602	-	0.0008	0.002
15q13	5.367.880	2018	ATP10A	0.0009	0.001
20q12	186.970	2023	TOP1	0.0009	0.0001

- Not available, * In the study by Simonen et al. (2003b) no confidence intervals or flanking regions are given. For regions reported in this study we used a 6 Mb interval around the peak. Note that the peak at the marker C11P15_3 reported by Simonen et al. (2003b) is omitted from this table, since this marker could not be identified. Sex and age were included as covariates in these analyses.

Additional checks and association analyses for the surviving SNPs

The 12 surviving novel SNPs and the 2 remaining SNPs in candidate genes were all in Hardy Weinberg equilibrium in both the group of non-exercisers and exercisers. The imputation quality was very good for the 13 SNPs imputed in either sample (average maximum posterior probability > 0.96). Table 7.6 presents additional association analyses on exercise participation for the 14 surviving SNPs using BMI as an additional covariate. Notably, all associations remained significant after correction for BMI. Table 7.6 also presents the association of these SNPs to the 6-category METHours phenotype. Nine of the surviving SNPs were associated with exercise intensity expressed as weekly METHours at a nominal significant level.

Table 7.6. Additional association analyses for exercise participation (also correcting for BMI) and for METHours for the remaining SNPs

		Association with exercise participation (covariates sex, age and BMI)			Association with METHours (covariates sex and age)		
		P value Dutch sample	P value American sample	P value Combined sample	P value Dutch sample	P value American sample	P value Combined sample
rs9633417	SGIP1	0.006	0.003	0.0002	0.06	0.007	0.003
rs667923	DNASE2B	0.01	0.005	0.0005	0.06	0.02	0.009
rs1766581	Intergenic	0.005	0.007	0.0004	0.0004	0.001	8.00*10 ⁻⁶
rs4355145	Intergenic	0.006	0.0009	6.60*10 ⁻⁵	0.64	0.0008	0.004
rs9789774	Intergenic	0.004	0.002	7.34*10 ⁻⁵	0.76	0.0005	0.004
rs13013897	SKIP	0.003	0.004	0.0001	0.08	0.001	0.001
rs10946904	PRSS16	0.005	0.01	0.0005	0.25	0.02	0.03
rs10827786	Intergenic	0.003	0.02	0.0006	0.04	0.01	0.03
rs2762527	PAPSS2	0.01	0.002	0.0002	0.02	0.01	0.003
rs12101846	Intergenic	0.004	0.0001	7.08*10 ⁻⁶	0.04	0.08	0.02
rs12905612	Intergenic	0.003	0.01	0.0004	0.40	0.004	0.01
rs238404	ERCC2	0.0007	0.003	2.82*10 ⁻⁵	0.0003	0.02	0.0001
rs2470158	CYP19A1	0.006	0.18	0.008	0.0062	0.23	0.01
rs12405556	LEPR	0.37	0.003	0.009	0.76	3.10*10 ⁻⁵	0.0003

Discussion

This study is the first to report the results from a genome-wide association study for leisure-time exercise behavior, using data from 1,772 unrelated Dutch adults and 978 unrelated American adults with valid genotypes on 470,719 SNPs (observed and imputed). The genome-wide association analyses revealed several novel SNPs in the genes SGIP1, DNASE2B, PRSS16, ERCC2 and PAPSS2 that are associated with exercise participation. Inspection of for the DNASE2B, PRSS16 and ERCC2 genes on the Ensembl and Entrez Gene websites revealed no immediate relevance for exercise behavior. The PAPSS2 gene encodes a protein that is involved in the sulfation of compounds such as lipids, carbohydrates and exogenous drugs and has been related to skeletal development and arthrosis (Ikeda et al., 2001; Ul Haque et al., 1998).

The SGIP1 gene is primarily expressed in the hypothalamus of the brain and is implicated in the regulation of energy homeostasis. A study in rodents (Trevaskis et al., 2005) suggests that the SGIP1 protein encoded by this gene has a physiological role in the hypothalamic neuronal systems that promotes positive energy balance and weight gain. Suppression of the mRNA product of the SGIP1 gene in rodents was shown to result in decreased food intake and increased metabolic rate. The SGIP1 gene is highly conserved between species and may therefore also be of importance for energy homeostasis in humans. In our combined samples the SGIP1 gene was associated with exercise behavior independent of BMI. This suggests that the gene is involved also in energy expenditure and the drive to exercise.

Of the candidate genes for exercise and physical activity phenotypes extracted from the literature (Loos et al., 2005; Lorentzon et al., 2001; Salmen et al., 2003; Simonen et al., 2003a; Stefan et al., 2002; Winnicki et al., 2004), we replicated association with the CYP19A1 gene in the Dutch sample and with the LEPR gene (rs12405556) in the American sample. The CYP19A1 gene is involved in the biosynthesis of estrogens from androgens and is suggested to play a role in bone mineral density and body fat regulation (Remes et al., 2003; Salmen et al., 2003; Tworoger et al., 2004). The CYP19A1 gene has previously been related to physical activity in study of early postmenopausal women (Salmen et al., 2003). Leptin is a hormone that has an important role in the regulation of energy balance. The LEPR gene, like the SGIP1 gene, is expressed in the hypothalamus (Park et al., 2006). The LEPR gene has frequently been linked to obesity and type II diabetes mellitus and is thought to influence food intake and energy expen-

diture (Liu et al., 2004; Park et al., 2006; Rankinen et al., 2006b). As for the SGIP1 gene the effects of LEPR were independent of BMI. This corroborates a study in a sample of 268 non-diabetic Pima Indians, in which the LEPR gene was found to be associated with 24 hour energy expenditure and physical activity, independent of adiposity (Stefan et al., 2002).

Associations of SNPs within several candidate linkage regions (Cai et al., 2006; De Moor et al., 2007a; Simonen et al., 2003b) were not replicated after multiple testing within each region had been taken into account. Still, one nominally significant SNP (rs10440457, $p=7.9 \cdot 10^{-3}$) in the 4q31.21 region is located only 359 kb away from the UCP1 gene. Uncoupling proteins are involved in metabolic energy expenditure (Oh et al., 2007; Ricquier, 2005). The 4q31.21 region has also been implicated for athlete status (De Moor et al., 2007b).

A number of recent genome-wide association studies show that the effect sizes of most loci that are identified for complex human diseases so far are small and very large samples including ten thousands of individuals instead of thousands of individuals are needed to detect these effects (Craddock et al., 2008; Loos et al., 2008; Weedon et al., 2008). Although we combined full GWA datasets from two relatively large samples, larger samples are likely to be needed to more robustly detect the small polygenic effects that account for the heritability of exercise behavior. A second limitation of our study is that parts of the relevant genomic variation may not have been captured. Variation in the final SNP set used may not have tagged all of the common variation in the genome. Moreover, our analyses were restricted to the autosomal genome, which excludes genes on the sex chromosomes and the mitochondrial DNA. The mitochondrial DNA is especially of interest because of its well-known involvement in energy metabolism and ATP generation (Saxena et al., 2006).

To conclude, this study found an association between genetic variants in the SGIP1, CYP19A1 and LEPR genes and voluntary exercise behavior. These effects on the drive to exercise were independent of BMI. We previously hypothesized that genetic effects on exercise ability, acute psychological effects of exercise and personality traits would account for the heritability of adult exercise behavior (De Geus & De Moor, 2008; De Moor et al., 2007a). The results of the current study suggest an important role for a different pathway, the hypothalamic regulation of the energy balance. Larger GWA studies are needed to identify the full palette of genetic variants influencing voluntary exercise behavior.





8

Exercise participation in adolescents and their parents: Evidence for genetic and generation specific environmental effects

This chapter is submitted as:

De Moor, M.H.M., Beem, A.L., Rebollo-Mesa, I., Stubbe, J.H., Willemsen, G., De Geus, E.J.C. & Boomsma, D.I. Exercise participation in adolescents and their parents: Evidence for genetic and generation specific environmental effects, *Behavior Genetics*, under review.

Abstract

Twin studies have shown that individual differences in adolescent exercise behavior are explained by additive genetic, shared environmental and unique environmental factors. The aim of this study was to test whether the shared environmental factors are best explained by the effects of cultural transmission of parents on their offspring, by generation specific environmental effects or by assortative mating. We used survey data on leisure-time exercise behavior collected between 1991-1995 in 3,525 adolescent twins and siblings (13 to 18 years) and 3,138 parents from 1,736 families registered at the Netherlands Twin Registry. Exercise participation (No/Yes, using a cut-off criterion of 4 metabolic equivalents and 60 minutes weekly) was based on questions on type, frequency and duration of exercise. We added exercise data from 2,962 adult twins, siblings and spouses collected in 2002-2004 who were age-matched with the parents (35-55 years). These data were used to investigate the causes of the spouse correlation ($r=0.41$, due to phenotypic assortment) and to estimate the heritability of exercise in the parental generation ($h^2=36\%$). A parent-offspring model including differences in variance decomposition across sex and generation was fitted to the dichotomous data using a threshold model. Results showed that 39.8% of the variance of exercise participation in adolescent boys was explained by additive genetic factors, 40.8% by generation specific shared environmental factors, 3.0% by vertical cultural transmission and 15.4% by unique environmental factors. In girls, 38.8% of the variance was explained by additive genetic factors, 49.0% by generation specific shared environmental factors and 12.3% by unique environmental factors. In conclusion, parental exercise behavior is of minor importance for adolescent exercise behavior.

Twin studies have repeatedly shown that individual differences in adult exercise behavior can be explained by a combination of genetic and unique environmental factors (Beunen & Thomis, 1999; Eriksson et al., 2006; Kujala et al., 2002; Lauderdale et al., 1997; Stubbe et al., 2006a) with heritability estimates ranging between 35% and 83%. In adolescents, shared environmental factors are of major importance. In a study of twins aged 13 to 20 years (Stubbe et al., 2005), exercise behavior in young adolescents (up to 16 years) was largely determined by shared environmental factors. The influence of these factors rapidly wanes when adolescents become young adults and genetic factors start to become of importance. Other twin studies (Carlsson et al., 2006; Maia et al., 2002) also show that family resemblance in adolescent leisure time exercise behavior is explained by a combination of additive genetic and shared environmental factors but that shared environmental factors cease to exert their influence in young adulthood.

The shared environment in adolescence can represent several different types of environmental factors that are shared between the twins. For example, it could consist of the influence of parents on their children's behavior (through their own physical activity levels or social support), environmental factors that are specific to the adolescent generation, such as the influence of siblings, friends, peers or the school, or a combination of these factors. All these factors have appeared in the literature as correlates of adolescent exercise behavior (Gustafson & Rhodes, 2006; Sallis et al., 2000). An alternative explanation for the shared environmental influences is non-random mating of the parents with respect to exercise behavior.

In the classical twin design, it is assumed that the correlation between additive genetic factors in monozygotic (MZ) twin pairs equals unity and in first degree relatives, such as dizygotic (DZ) twin pairs and non-twin sibling pairs, this correlation is 0.5 (Neale & Cardon, 1992). This value of 0.5 comes from the expected additive genetic correlation in first degree relatives under random mating (Falconer & Mackay, 1996). Non-random mating refers to the phenomenon that the phenotypes of spouses are correlated at the moment they meet. A number of studies have reported a significant spouse correlation for exercise behavior, ranging from 0.16 to 0.60 (Aarnio et al., 1997; Boomsma et al., 1989; Perusse et al., 1988; Perusse et al., 1989; Seabra et al., 2008).

There are several possible explanations for the spouse correlation in exercise behavior, including social interaction, social homogamy and phenotypic assortment (Heath & Eaves, 1985; Reynolds et al., 2006; van Leeuwen et al., 2008). Social interaction refers to the phenomenon that spouses mutually influence each other because they spend time together. This means that the phenotypes of the spouses are not necessarily correlated at the time they first meet, but they become more similar during the course of their relationship. Therefore, if social interaction explains the spouse correlation, it is expected that the spouse correlation increases when partners are together for a longer time. Social homogamy is the tendency that individuals coming from similar social backgrounds are more likely to meet and marry each other. This process would also induce a positive spouse correlation. A third type of explanation for a positive spouse correlation is phenotypic assortment, which occurs when individuals select each other based on the phenotype under study (or a correlated phenotype). If phenotypic assortment is present, this increases the additive genetic variance in the offspring generation and the covariance between additive genetic factors in first degree relatives. If this increased genetic resemblance is not accounted for in the twin model, and an additive genetic correlation of 0.5 is assumed, this leads to the overestimation of shared environmental effects in an ACE model and to overestimation of additive genetic effects in an AE model.

Extending the classical twin design with additional family members such as non-twin siblings and parents (Boomsma & Molenaar, 1987; Eaves et al., 1978; Fulker, 1989; Heath et al., 1985) makes it possible to decompose the shared environmental factors found in adolescent exercise behavior into the influences of the parental phenotype on offspring behavior (vertical cultural transmission), environmental effects that are shared among offspring but non-shared with the parents (horizontal cultural

transmission) and the effects of non-random mating. These effects can be separated by modeling the correlations between parents, between parents and their offspring and between twins and siblings. A significant spouse correlation suggests that some of the shared environmental variance may be explained by assortative mating. If the parent-offspring correlations are larger than what would be expected under genetic transmission alone and comparable to the DZ twin and sibling correlations, this suggests that the shared environmental factors found in adolescents are the result of vertical cultural transmission. A lower parent-offspring correlation compared with DZ twin and sibling correlations suggests that at least part of the environmental factors is shared between twins and siblings only.

There are a few studies that have examined the familial resemblance for exercise behavior in twins and their parents (Aarnio et al., 1997; Boomsma et al., 1989; Perusse et al., 1989) but all of these studies made the (a priori reasonable) assumption that the variance decomposition in parents and offspring is the same. However, as outlined above, we now know that there is a large shift from shared environmental to genetic factors in the transition from adolescence to adulthood. In this study, we readdressed parent-offspring resemblance, taking into account possible differences in variance decomposition between the generations. We used structural equation modeling techniques developed for extended kinships (Boomsma & Molenaar, 1987; Eaves et al., 1978; Fulker, 1989; Neale et al., 1994b; Phillips & Fulker, 1989) to test whether the shared environmental factors in adolescents are best explained by the effects of vertical cultural transmission, horizontal cultural transmission, assortative mating or a combination of these mechanisms. To this end, we made use of data on exercise participation collected from families registered at the Netherlands Twin Registry at different time points. Between 1991-1995 parent-offspring exercise data were collected from 3,525 adolescent twins and siblings and 3,138 parents. Between 2002-2004 exercise data were collected in 1,843 adult twins and siblings and 1,119 spouses that were in the same age range as the parents in the parent-offspring sample.

The age-matched adult twin sample allowed us to estimate the heritability of exercise in the parents in our parent-offspring sample. Twin-spouse pairs were used to test for different explanations of the spouse correlation. Data on duration of relationship in twin-spouse pairs were used to test whether the spouse correlation could be explained by social interaction. To test whether social homogamy or phenotypic assortment explained the spouse correlation, the twin-cotwin's spouse correlations were computed as a function of zygosity. This information was used to extend the parent-

offspring model to include differences in variance decomposition across generations. The parent-offspring model was further extended to account for sex differences in the variance decomposition of exercise. Exercise participation was a dichotomous phenotype and therefore we used a threshold model. In adolescents, the relatively small opposite-sex (DOS) twin correlation compared with the dizygotic (DZ) same-sex twin correlations suggests that the shared environmental factors influencing exercise in boys and girls are partly different (Stubbe et al., 2005).

We extended the parent-offspring model to examine which part of the variance of exercise participation in adolescent boys and girls is explained by additive genetic factors, by vertical cultural transmission effects of parents on offspring, by horizontal transmission through environmental factors shared among offspring, and by unique environmental factors.

Methods

Participants

This study was part of an on-going study on health, lifestyle and personality in twins and their family members (siblings, parents and spouses of twins) who are voluntarily registered with the Netherlands Twin Register (NTR) (Boomsma et al., 2002; Boomsma et al., 2006b). Adolescent and young adult twins were recruited through city councils during 1990 and 1992. Since 1991, every two to three years the participants receive a mailed questionnaire, including questions about exercise participation. The exercise data from the first three surveys (1991, 1993 and 1995) were used to obtain a dataset in adolescent twins. In 1991 and 1993, surveys were sent out to twins and their parents; in 1995 non-twin siblings also completed the questionnaire. We created a cross-sectional dataset with adolescent (13 to 18 years old) twins and siblings and their parents by selecting from each twin family the data of the most recent survey. If the most recent survey of a twin family contained missing data on exercise for one of the twins (incomplete twin pair) and an earlier survey contained complete information of a twin pair, data from this earlier survey were selected instead. This ensured a maximum of complete twin pairs in the dataset. We excluded half-siblings, non-biological siblings and parents, twins with missing zygosity and all subjects with missing data on exercise, sex or age (less than 2%). This resulted in a dataset with 3,360 twins, 165 siblings and 3,138 parents from 1,736 families. Of the 3,360 twins, there were 1,667 twin pairs of which both twins had valid exercise data, 292 were monozygotic

male (MZM), 239 dizygotic male (DZM), 393 monozygotic female (MZF), 266 dizygotic female (DZF) and 477 opposite-sex (DOS) pairs. Mean age of the twins was 16.4 (SD=1.1), mean age of the siblings was 16.0 (SD=1.5) and of the parents 45.5 (SD=4.6). Of all family members, 46.9% were men and 53.1% were women. We additionally used data collected in 2002 and 2004 from 1,242 twins and 601 siblings between 35 and 55 years (mean age 43.6, SD=6.2). Of the 1,242 twins, there were 465 twin pairs of which both twins had exercise data, 63 were MZM, 19 DZM, 234 MZF, 96 DZF and 53 DOS pairs. These data were used to compute the heritability of exercise in the age range of the parents of the adolescent offspring. Data from 1,115 twin-spouse pairs were used to test for different explanations of the spouse correlation for exercise participation.

Zygoty determination

Zygoty was determined by DNA typing for 29.1% of the adolescent same-sex twin pairs and for 36.6% of the adult same-sex pairs. For the other same-sex twin pairs, zygoty was based on eight items on physical similarity and the frequency of confusion of the twins by parents, other family members and strangers. Agreement between zygoty based on these items and zygoty based on DNA was 97% (Willemsen et al., 2005).

Measurements

Leisure-time exercise participation was measured with a number of questions. The first question "Do you participate in exercise regularly?" could be answered with 'Yes' or 'No'. If the participants responded affirmative, further information on type, frequency and duration of exercise was gathered. Reported non-leisure time activities, such as walking or biking to work, were not counted as exercise. All remaining exercise activities were assigned a metabolic equivalent (MET) value, using Ainsworth's Compendium of physical activity (Ainsworth et al., 2000). A MET score of 1 corresponds to the rate of energy expenditure when at rest (1 kcal/kg/h). Subjects were classified as regular exercisers if they participated in exercise with at least 4 MET for 60 minutes weekly for at least 10 months during the past year. Subjects were classified as non-exercisers otherwise. This dichotomous variable was used in the analyses.

Statistical analyses

We used structural equation modeling in Mx (Neale et al., 2006) for all analyses. Threshold models were fitted to the raw ordinal data using maximum likelihood. In threshold models it is assumed that the ordinal vari-

able has an underlying liability with a continuous and standard normal distribution. For dichotomous traits, there is one threshold that divides the liability distribution into two discrete categories (i.e., 'regular exerciser' or 'non-exerciser'). This threshold is based on the prevalence of the different categories in the population. We allowed for sex and generation differences in the threshold and tested whether the prevalence for exercise participation differs in fathers, mothers, sons and daughters. Within each sex by generation group, we modeled age as a covariate on the threshold to account for any remaining variability in prevalence of exercise as a function of age. The variance of the liability distribution of exercise participation is fixed at 1 on all types of relatives.

In a first set of analyses, we estimated the spouse correlations for exercise participation. In the adolescent offspring and their parents, we estimated the spouse correlation in the parents. In the sample of adult twins and siblings, we estimated the spouse correlations in the spouse-twin pairs. We tested whether these spouse correlations were equal. Next, we tested three different explanations for the spouse correlation: social interaction, social homogamy and phenotypic assortment (Heath & Eaves, 1985; Reynolds et al., 2006). If social interaction explains the spouse correlation, it is expected that the spouse correlation increases as a function of the duration of the relationship. Data on duration of the relationship were available in surveys 2002 and 2004 for twins and their spouses. If social homogamy processes drive the spouse correlation, it is expected that the correlation between a twin with the cotwin's spouse is the same for MZ and DZ twin pairs, since twins within a pair come from the same social background. It is further expected that these twin-cotwin's spouse correlations equal the twin-spouse correlations. If assortment is mainly phenotypic, the twin-cotwin's spouse correlations are expected to be the product of the twin correlation and the twin-spouse correlation (Heath & Eaves, 1985). If the phenotype is heritable, the MZ twin correlation is smaller than one but larger than the DZ twin correlation. Therefore, it is expected that the twin-spouse correlation is larger than the MZ twin-cotwin's spouse correlation, which in turn is larger than the DZ twin-cotwin's spouse correlation.

Secondly, we computed the heritability of exercise participation in adult twins and siblings of comparable age with the parents of adolescent twins, assuming that there were no birth cohort differences in heritability. We used standard twin models (ACE or ADE) with quantitative and qualitative sex differences in parameters as described elsewhere (Neale & Cardon, 1992). Thresholds were also allowed to be different in males and

females. An age effect was modeled on the thresholds. The results of this model were used in the analyses of the parent-offspring data of adolescent twins and siblings and their parents.

Thirdly, in a saturated model for the data from adolescent twins, siblings and their parents, four thresholds were estimated (for the fathers, mothers, sons and daughters), an age effect on the thresholds, and the tetrachoric correlations for exercise participation between the different types of relatives. A total of 10 correlations was estimated: 1 correlation between the parents, 4 parent-offspring correlations (father-son, father-daughter, mother-son and mother-daughter) and 5 twin and sibling correlations (MZM, DZM/male sibling, MZF, DZF/female sibling and DOS/opposite sex sibling). Correlations between DZ twins, between twins and their non-twin siblings and between non-twin siblings were constrained to be equal. We tested for sex differences in the parent-offspring and twin/sibling correlations.

Next, in a series of models, the liability variance of exercise participation was decomposed into genetic and environmental factors, while modeling the effects of assortative mating. We used the factor model described by Neale and colleagues (Neale et al., 1994b), which builds on the work from Fulker, Heath, Eaves and others (Eaves et al., 1978; Fulker, 1989; Heath et al., 1985; Phillips & Fulker, 1989). We extended this model to account for sex and generation differences in the variance decomposition of exercise. The path model is shown in Figure 8.1 for an opposite-sex twin pair. The Mx script, which is based on the script provided by Neale et al. (1994b), is given in Appendix B. The expectations for the covariances among parents and offspring are given in Appendix C. Additional siblings are omitted from the model for clarity of presentation. The variances of the phenotypes in the offspring were decomposed into additive genetic factors (A), shared environmental factors (C) and non-shared environmental factors (E). Different path loadings were allowed in boys (a_{SO} , c_{SO} , e_{SO}) and girls (a_{DA} , c_{DA} and e_{DA}) to model quantitative sex differences. Qualitative sex limitation was modeled in C by fixing the correlation between residual variances of C in same-sex pairs to 1, but estimating this correlation in opposite-sex pairs (bound between 0 and 1). The variances of the parental phenotypes were decomposed into A and E. The estimates of the path loadings in the fathers (a_{FA}^* , e_{FA}^*) and mothers (a_{MO}^* , e_{MO}^*) were fixed at the values obtained from the analysis in the age-matched adult twin sample. The phenotypic variances of the phenotype in fathers, mothers, sons and daughters were constrained at 1, which provides a scale for the liability distribution in the threshold model. Resemblance between parents and off-

If there is no assortative mating, 50% of the additive genetic variance in the offspring generation comes from both parents and the remaining 50% is due to recombination, as represented by the 0.5 residual variance of A in the offspring. Vertical cultural transmission was modeled as the effect of the parental phenotype on the environment that is shared among offspring. The effect of vertical cultural transmission was allowed to be different depending on the sex of both the parent and the child. The total variance of C was computed as the variance induced by the vertical cultural transmission effects plus the residual variance. The residual variance represents horizontal cultural transmission and was fixed at 1. A consequence of vertical cultural transmission is that A and C become correlated in the offspring. The spouse correlation was modeled as phenotypic assortment based on the first set of analyses in the adult twin-spouse pairs and is represented as a co-path in the model (Cloninger, 1980). This co-path implicitly represents the many correlations induced by assortment between genetic and environmental deviations of spouses (Eaves et al., 2005). This induces an additive genetic correlation between first degree relatives that is larger than 0.5.

The parent-offspring model was fitted to the data of adolescent twins and siblings and their parents. Different constraints were imposed to test for the significance of assortative mating and vertical cultural transmission. The fit of these models was evaluated by means of the log-likelihood ratio test (LRT). The difference in minus two times the log-likelihood (-2LL) between two nested models has a χ^2 distribution and the degrees of freedom (df) equals the difference in df between the two models. If the χ^2 -test yielded a p-value larger than 0.05 the fit of the constrained model was not significantly worse than the fit of the more complex model and the constrained model was kept as the most parsimonious and best fitting model.

Results

Prevalence of exercise participation in adolescents and adults

The prevalence of exercise participation in different age groups for men and women is given in Table 8.1. The percentage of regular exercisers in adolescents is larger than in adults ($\chi^2=13.16$, $\Delta df=2$, $p<0.01$). Further, the percentage of regular exercisers in adolescent boys is larger than in adolescent girls ($\chi^2=9.45$, $\Delta df=1$, $p<0.01$). The percentage of regular exercisers in adult men is not different from the percentage of exercisers in adult women ($\chi^2=0.89$, $\Delta df=1$, $p=0.35$). Within each group of sons, daughters, fathers and mothers, there is a significant decrease in prevalence of exercise participation with age ($\chi^2=47.73$, $\Delta df=1$, $p<0.001$).

Table 8.1. Prevalence of exercise participation (%) in adolescent twins and siblings and their parents (data from 1991, 1993 and 1995) and in adult twins, siblings and spouses (data from 2002, 2004)

Exercise participation	Adolescent twins and siblings Mean age 16.4 (SD=1.2)		Parents of adolescent twins Mean age 45.5 (SD=4.6)		Adult twins, siblings and spouses Mean age 43.4 (SD=6.1)	
	No	Yes	No	Yes	No	Yes
Men	31.7	68.3	60.8	39.2	50.7	49.3
Women	38.4	61.6	59.2	40.8	49.6	50.4
Total	35.3	64.7	59.9	40.1	50.1	49.9

Spouse correlations for exercise participation in adults

The spouse correlation in parents of adolescent twins is estimated at 0.41 (95% confidence interval (CI) 0.34 ; 0.48). The spouse correlation in the adult twin cohort is 0.44 (95% CI 0.37 ; 0.53). Obviously, these correlation were not different ($\chi^2=0.69$, $\Delta df=1$, $p=0.41$).

To investigate whether the spouse correlations are the result of social interaction, the correlations were computed as a function of duration of the relationship. The correlations are given in Table 8.2. There is no clear increase in spouse correlations as a function of length of the relationship. Constraining the twin-spouse correlations to be equal across different groups of duration of relationship was permitted, as indicated by the non-significant deterioration in fit ($\chi^2=6.59$, $\Delta df=3$, $p=0.09$). Thus, social interaction does not provide a good explanation for the spouse correlations.

To further investigate the causes of assortment, the correlations of twins with the cotwin's spouses as a function of zygosity were estimated

(see Table 8.2). The estimates of the MZ twin-cotwin's spouse correlations are larger than of the DZ twin-cotwin's spouse correlations, although these differences were not statistically significant in male twin pairs ($\chi^2=2.54$, $\Delta df=1$, $p=0.11$), female twins pairs ($\chi^2=2.07$, $\Delta df=1$, $p=0.15$) or all twin pairs ($\chi^2=3.80$, $\Delta df=1$, $p=0.05$). Further, the twin-cotwin's spouse correlations in MZ pairs were significantly lower than the twin-spouse correlation ($\chi^2=18.01$, $\Delta df=2$, $p<0.001$). Given this pattern of correlations, the correlation in spouses for exercise participation seems best explained by phenotypic assortment.

Table 8.2. Tetrachoric twin-spouse and twin-cotwin's spouse correlations for exercise participation as a function of duration of relationship and zygosity

Twin-spouse correlations as a function of duration of relationship					
	<5 years	5-10 years	10-15 years	>=15 years	
Number of pairs	245	360	198	255	
Tetrachoric correlation	0.61	0.34	0.42	0.47	
95% Confidence Interval	0.45 ; 0.74	0.18 ; 0.48	0.22 ; 0.59	0.30 ; 0.62	
Twin-cotwin's spouse correlations as a function of zygosity of the twin pair					
	MZM	DZM	MZF	DZF	DOS
Number of pairs	144	63	352	155	174
Tetrachoric correlation	0.20	-0.21	0.16	-0.04	0.04
95% Confidence Interval	-0.03 ; 0.42	-0.57 ; 0.23	0.02 ; 0.31	-0.29 ; 0.20	-0.18 ; 0.26

Twin-cotwin's spouse correlations in MZM=Monozygotic male twin pairs, DZM=Dizygotic male twin pairs, MZF=Monozygotic female twin pairs, DZF=Dizygotic female twin pairs and DOS=Dizygotic opposite-sex twin pairs

Heritability of exercise participation in adults

In adults, the MZM correlation was 0.38 (95% Confidence Interval (CI): -0.00; 0.68), the DZM/male sibling correlation was 0.03 (95% CI: -0.30; 0.41), the MZF correlation was 0.41 (95% CI: 0.22; 0.57), the DZF/female sibling correlation was 0.09 (95% CI: -0.09; 0.27) and the DOS/opposite sex sibling correlation was 0.20 (95% CI: -0.01; 0.38). There were no significant sex differences in the twin/sib correlations ($\chi^2=0.93$, $\Delta df=3$, $p=0.82$). The MZ correlation was 0.40 (95% CI: 0.23; 0.55) and the DZ correlation 0.12 (95% CI: -0.01 ; 0.26). An ADE model without sex differences indicated that 9.4% (95% CI: 0.00%; 48.0%) of the liability variance of exercise participation is explained by additive genetic factors, 30.8% by non-additive

genetic factors (95% CI: 0.00%; 55.0%) and 59.8% (95% CI: 45.0%; 76.6%) by non-shared environmental factors. Non-additive genetic factors were not significant ($\chi^2=0.95$, $\Delta df=1$, $p=0.33$). An AE model in which 36.0% (95% CI: 21.5%; 49.8%) of the variance in exercise was explained by additive genetic factors and 64.0% (95% CI: 50.2%; 78.5%) by non-shared environmental factors fitted the data well ($\chi^2=1.88$, $\Delta df=4$, $p=0.76$). The path loading of A on the phenotype was 0.60 and of E on the phenotype was 0.80. These values were used as fixed values for the decomposition of the phenotypic variance in the parents in the parent-offspring model.

Parent-offspring correlations for exercise participation

The parent-offspring, twin and sibling correlations are given in Table 8.3. The resemblance between parents and their offspring is larger when parents and offspring are of the same sex. Correlations between fathers and sons and mothers and daughters and also between fathers and daughters and mothers and sons could be constrained to be equal ($\chi^2=2.40$, $\Delta df=2$, $p=0.30$), but the father-son and mother-daughter correlations are significantly larger than the father-daughter and mother-son correlations ($\chi^2=10.58$, $\Delta df=1$, $p<0.01$). The father-son and mother-daughter correlation is 0.32 (95% CI 0.27 ; 0.38) and the father-daughter/mother-son correlation is 0.20 (95% CI 0.13 ; 0.26).

Table 8.3. Tetrachoric correlations for exercise participation between adolescent twins, siblings and parents

	Fa-Mo	Fa-So	Fa-Da	Mo-So	Mo-Da	MZM	DZM/ sib	MZF	DZF/ sib	DOS/ sib
Number of complete pairs	1,444	1,438	1,599	1,566	1,792	292	305	393	338	580
Tetrachoric correlation	0.41	0.36	0.21	0.18	0.29	0.85	0.66	0.88	0.69	0.45
95% CI, lower bound	0.34	0.27	0.12	0.09	0.20	0.76	0.52	0.81	0.56	0.34
95% CI, upper bound	0.48	0.44	0.30	0.27	0.36	0.91	0.78	0.92	0.79	0.56

Fa=Father, Mo=Mother, So=Son, Da=Daughter, MZM=Monozygotic male twin pairs, DZM/sib=Dizygotic male twin and non-twin sibling pairs, MZF=Monozygotic female twin pairs, DZF/sib=Dizygotic female twin and non-twin sibling pairs, DOS/sib=Dizygotic opposite-sex twin and non-twin sibling pairs, 95% CI=95% Confidence Interval

There were no sex differences in the MZ correlations ($\chi^2=0.77$, $\Delta df=1$, $p=0.38$) and DZ/sibling correlations ($\chi^2=1.47$, $\Delta df=1$, $p=0.23$), but the DOS/sibling correlation was significantly smaller than the same-sex DZ/sibling correlations ($\chi^2=14.69$, $\Delta df=2$, $p<0.001$), indicating that there are no quantitative sex differences, but there are qualitative sex differences for exercise participation in adolescents. This seems consistent with the sex differences found in the parent-offspring correlations. The MZ and DZ/sib correlations in boys and girls are high, confirming that shared environmental factors play an important role in exercise participation in adolescents. However, the MZ correlations are significantly larger than the DZ/sib correlations in boys ($\chi^2=6.37$, $\Delta df=1$, $p<0.05$) and in girls ($\chi^2=5.29$, $\Delta df=1$, $p<0.05$), indicating that genetic factors are also of importance.

Parent-offspring modeling of exercise participation

Model fitting results for the parent-offspring models are given in Tables 8.4 and 8.5. The vertical cultural transmission path from father to son was positive and significant ($\chi^2=5.62$, $\Delta df=1$, $p<0.05$). The path from mother to son was also significant but negative ($\chi^2=4.18$, $\Delta df=1$, $p<0.05$). The vertical cultural transmission paths from father and mother to daughter were not significant ($\chi^2=1.10$, $\Delta df=1$, $p=0.29$ and $\chi^2=0.72$, $\Delta df=1$, $p=0.40$, respectively). The correlation between residual variances of the shared environmental latent factor in opposite-sex pairs was estimated at 0.56 and significantly lower than 1 ($\chi^2=5.80$, $\Delta df=1$, $p<0.05$), suggesting that besides sex specific effects of the parental phenotype on boys and girls, there are also qualitative differences between boys and girls in the generation specific shared environmental factors. The additive genetic effects in boys and girls could not be omitted from the model ($\chi^2=17.36$, $\Delta df=2$, $p<0.001$). There were no significant quantitative sex differences in the path loadings a, c and e between boys and girls ($\chi^2=1.51$, $\Delta df=2$, $p=0.68$). The best fitting model showed that 39.8% of the variance of exercise participation in adolescent boys was explained by additive genetic factors, 40.8% by generation specific shared environmental factors, 3.0% by vertical cultural transmission and 15.4% by unique environmental factors. In girls, 38.8% of the variance was explained by additive genetic factors, 49.0% by generation specific shared environmental factors and 12.3% by unique environmental factors.

Table 8.4. Model fitting results for exercise participation in adolescent twins and siblings and their parents (1991, 1993, 1995)

Model	v.	-2LL	df	#par	χ^2	Δ df	p-value
1. Full genetic model	–	8036.35	6648	15	–	–	–
2. Drop t_{FA-SO}	1	8041.97	6649	14	5.62	1	<0.05
3. Drop t_{MO-SO}	1	8040.53	6649	14	4.18	1	<0.05
4. Drop t_{FA-DA}	1	8037.45	6649	14	1.10	1	0.29
5. Drop t_{MO-DA}	1	8037.07	6649	14	0.72	1	0.40
6. Drop t_{FA-SO} , t_{MO-SO} , t_{FA-DA} , t_{MO-DA}	1	8048.52	6652	11	12.17	4	<0.05
7. Drop t_{FA-SO} , t_{MO-SO} , t_{FA-DA} , t_{MO-DA} , c_{SO} , c_{DA}	1	8096.08	6655	8	59.73	7	<0.001
8. Drop $r_{C,OS}$	1	8042.15	6649	14	5.80	1	<0.05
9. Drop a_{SO} , a_{DA}	1	8053.71	6650	13	17.36	2	<0.001
10. $a_{SO}=a_{DA}$, $c_{SO}=c_{DA}$, $e_{SO}=e_{DA}$	1	8037.86	6650	13	1.51	2	0.68

v.=versus (indicates comparison model), -2LL=-2 Log-Likelihood, df=degrees of freedom, #par=number of free parameters, χ^2 =chi square, Δ df=difference in degrees of freedom between fitted and comparison model, D=Assortative mating path, a=additive genetic path, c=shared environmental path, d=non-additive genetic path, e=non-shared environmental path, t=cultural transmission path, FA=Father, MO=Mother, SO=Son, DA=Daughter

Table 8.5. Parameter estimates from parent-offspring models for exercise participation in adolescent twins and siblings and their parents (1991, 1993, 1995)

Model	D	t _{FA,SO}	t _{MO,SO}	t _{FA,DA}	t _{MO,DA}	var(A)	var(C _{SO})	cov(A _{SO} , C _{SO})	var(C _{DA})
1. Full genetic model	0.41	0.25	-0.22	-0.10	0.07	1.07	1.07	0.01	1.01
2. Drop t _{FA,SO}	0.41	-	-0.23	-0.11	0.08	1.07	1.05	-0.10	1.01
3. Drop t _{MO,SO}	0.40	0.24	-	-0.10	0.09	1.07	1.06	0.10	1.01
4. Drop t _{FA,DA}	0.41	0.26	-0.23	-	0.06	1.07	1.07	0.01	1.00
5. Drop t _{MO,DA}	0.41	0.26	-0.23	-0.08	-	1.07	1.07	0.01	1.01
6. Drop t _{FA,SO} , t _{MO,SO} , t _{FA,DA} , t _{MO,DA}	0.41	-	-	-	-	1.07	1	0	1
7. Drop t _{FA,SO} , t _{MO,SO} , t _{FA,DA} , t _{MO,DA} , c _{SO} , c _{DA}	0.39	-	-	-	-	1.07	1	0	1
8. r _{c,OS} =1	0.41	0.27	-0.58	-0.10	0.08	1.07	1.29	-0.13	1.01
9. Drop a _{SO} , a _{DA}	0.41	0.43	0.05	0.14	0.28	1.07	1.21	0.20	1.13
10. a _{SO} =a _{DA} , c _{SO} =c _{DA} , e _{SO} =e _{DA}	0.41	0.18	-0.21	-0.09	0.11	1.07	1.05	-0.01	1.01

Model	cov(A _{DA} , C _{DA})	a _{SO}	c _{SO}	e _{SO}	a _{DA}	c _{DA}	e _{DA}	r _{c,OS}
1. Full genetic model	-0.01	0.61	0.64	0.39	0.62	0.69	0.35	0.56
2. Drop t _{FA,SO}	-0.01	0.76	0.56	0.36	0.62	0.69	0.35	0.59
3. Drop t _{MO,SO}	-0.00	0.40	0.74	0.43	0.62	0.69	0.35	0.56
4. Drop t _{FA,DA}	0.02	0.61	0.64	0.39	0.56	0.72	0.36	0.54
5. Drop t _{MO,DA}	-0.04	0.61	0.64	0.39	0.66	0.67	0.34	0.56
6. Drop t _{FA,SO} , t _{MO,SO} , t _{FA,DA} , t _{MO,DA}	0	0.63	0.66	0.38	0.61	0.70	0.35	0.52
7. Drop t _{FA,SO} , t _{MO,SO} , t _{FA,DA} , t _{MO,DA} , c _{SO} , c _{DA}	0	0.90	-	0.38	0.91	-	0.39	-
8. r _{c,OS} =1	-0.01	0.84	0.38	0.37	0.61	0.69	0.35	1
9. Drop a _{SO} , a _{DA}	0.08	-	0.80	0.48	-	0.85	0.44	0.55
10. a _{SO} =a _{DA} , c _{SO} =c _{DA} , e _{SO} =e _{DA}	0.01	0.62	0.67	0.36	0.62	0.67	0.36	0.56

D=Assortative mating path, a=additive genetic path, c=shared environmental path, e=non-shared environmental path, t=cultural transmission path, FA=Father, MO=Mother, SO=Son, DA=Daughter, var(A)=variance of the additive genetic factor var(C)=variance of the shared environmental factor, cov(A,C)=gene-environment covariance

Discussion

In this study, we extended existing parent-twin offspring models to allow for differences in genetic and environmental effects across sex and generation. The model also allowed for cultural transmission to differ as a function of the sex of the parent and the offspring. We fitted this extended model to data on exercise behavior from adolescent MZ and DZ twins and their siblings and parents in order to determine to what extent the shared environmental factors typically found for exercise participation in adolescent twins are composed of the effects of the parental phenotype on the offspring's exercise behavior (vertical cultural transmission), generation-specific environmental factors (horizontal cultural transmission) and the effects of assortative mating.

First, data from adult twins, siblings, and their spouses (age matched to the parents in the parent-offspring sample) were used to investigate whether the spouse correlation for exercise participation was best explained by social interaction, social homogamy or phenotypic assortment and to determine the heritability of exercise participation in the parental generation. The spouse correlation for exercise participation was significant and relatively high (0.41). This correlation was not the result of social interaction between spouses, as the correlation did not increase as a function of duration of relationship. The correlations between adult twins and the cotwin's spouse were a function of zygosity. These results indicate that phenotypic assortment is the best explanation for the observed spouse correlations. This means that individuals who are both exercisers are more attracted to each other because they share similar interests (for example, participating in sports or being active). Our finding of a positive spouse correlation is consistent with previous studies reporting significant spouse correlations for exercise behavior ranging from 0.16 to 0.60 (Aarnio et al., 1997; Boomsma et al., 1989; Perusse et al., 1988; Perusse et al., 1989; Seabra et al., 2008). We are the first to demonstrate that this spouse correlation most likely arises from phenotypic assortment. In keeping with previous studies (De Geus et al., 2003; Stubbe et al., 2006a), the heritability in adults between 35 and 55 years old was estimated at 36%, without any evidence for sex differences in heritability, with the remaining variance due to unique environmental factors.

Next, a parent-offspring model was fitted to data from adolescent twins, their siblings and parents, in which the effects of phenotypic assortment and cultural transmission were included. The heritability of exercise participation in the parents was set at 36%. For adolescent boys, 39.8% of

the variance of exercise participation was explained by additive genetic factors, 40.8% by generation specific shared environmental factors, 3.0% by vertical cultural transmission and 15.4% by unique environmental factors. Only the exercise behavior of the father had a positive influence on the exercise behavior of the son. In girls, 38.8% of the phenotypic variance was explained by additive genetic factors, 49.0% by generation specific shared environmental factors and 12.3% by unique environmental factors. There was evidence that the environmental factors shared within the adolescent generation are partly different in boys and girls. Thus, the effects of parental exercise behavior on the exercise behavior of their children were only significant in boys, although the impact is minor compared with the influence of generation specific environmental and genetic factors.

The findings from the parent-offspring models are largely consistent with two previous studies in adolescent twins and their parents (Aarnio et al., 1997; Boomsma et al., 1989) with regard to the low parent-offspring correlation and moderate to large MZ and DZ twin correlations. The study from Aarnio et al. (1997), conducted in 3,254 twins of 16 years old and their parents, reported low parent-offspring correlations (0.05-0.10), high monozygotic (MZ) correlations (0.64-0.72) and moderate DZ correlations (0.22-0.45) for leisure time physical activity. These correlations also suggest that both genetic and shared environmental factors influence exercise behavior and that the influence of the parental phenotype on adolescent exercise behavior is minor. In a study of 90 twin pairs aged 14-20 years and their parents (Boomsma et al., 1989), it was shown that sports participation is influenced by additive genetic factors and non-shared environmental factors in boys, and additionally by shared environmental factors in girls. This study did not find evidence for vertical cultural transmission. In contrast, in a sample of 893 biological and adopted children (including twins) with a mean age of 15 years ($SD=3.3$) and their parents (Perusse et al., 1989) it was found that exercise participation is influenced by a combination of vertical and horizontal cultural transmission. Vertical transmission effects explained 12% of the variance in exercise participation. It was not tested whether these effects were sex specific. Level of habitual physical activity, however, was not influenced by vertical cultural transmission. Both measures were based on a three-day activity diary, which reflects daily total physical activity and may be different from our measure of past year's leisure time exercise behavior.

The low impact of vertical cultural transmission in boys and the absence in girls does not necessarily mean that parents cannot have any influence on their children's exercise behavior. Cultural transmission was

modeled as phenotypic: the phenotype of the parents needs to have a direct influence on the environment of the children. If parental influences that are unrelated to their own exercise behavior also have an impact on children's exercise behavior, these parental influences will be part of the residual shared environmental variance. Not much is known yet on the relationship between parental exercise behavior and other parental influences such as attitudes toward children's exercise behavior, support, and actual facilitation of exercise behavior (pay for equipment, drive to the playing field etc), although a review suggests that parental attitudes to exercise behavior are more strongly correlated to children's exercise behavior than the exercise behavior of the parents themselves (Gustafson & Rhodes, 2006). Data on parental exercise behavior, attitudes and social support in twin families are needed to test the hypothesis that parental influences unrelated to the parents own exercise behavior are of importance for exercise behavior in their children, while taking into account the effects of assortative mating and genetic transmission.

A potential limitation of the present study is that we obtained heritability estimates and spouse correlations for exercise participation in the parental generation from twins who had a comparable age range as the parents, but who came from an younger birth cohort (about 10 years difference). This could inadvertently have introduced a birth cohort related difference in heritability. To examine this possibility we compared heritability estimates in the age-matched adult twins to published values in adult twins that came from the same birth cohort as the parents of the adolescent twins (De Geus et al., 2003). Although this was a smaller sample (213 adult twin pairs), the heritability of exercise was estimated at 41% (De Geus et al., 2003), a value highly comparable to that found in the larger sample of adult twins in the current study. Thus, at least for heritability estimates, birth cohort effects were small at best.

A second limitation of this study is that in the parent-offspring model we allowed for different effects of additive genetic, shared environmental and unique environmental factors across generations (quantitative generation differences) but we assumed that the genetic factors affecting exercise in adolescence and adulthood were the same (no qualitative generation differences). A model with both vertical cultural transmission and qualitative differences across generations was not identified, since the information to estimate both parameters comes from the same parent-offspring correlation. The negative cultural transmission path from mother to son may however be an indication that qualitative generation differences in genetic effects on exercise behavior may be present. Future studies using

longitudinal data on exercise participation in adolescents followed up into young and middle adulthood are needed to test the hypothesis that different genetic factors influence exercise participation in the different stages of life. With a new wave of data being currently collected at the Netherlands Twin Register, it will become possible to estimate the longitudinal genetic correlation of exercise participation in individuals growing up from adolescents into adults who are in their thirties.

In conclusion, our study shows that individual differences in adolescent exercise participation are accounted for by a combination of additive genetic factors, shared environmental factors and non-shared environmental factors. The influence of parental exercise behavior on their children's exercise behavior was only significant for the influence of fathers on sons, explaining a small portion of the variance. Therefore, we can conclude that the influence of parental exercise behavior seems to be of minor importance for adolescent exercise behavior, but parental influences through social support mechanisms cannot be ruled out. Future research needs to include measures of parental attitudes and social support towards children's exercise behavior to resolve this. It also needs to focus on the generation specific environmental factors on adolescent exercise behavior (including peer behavior), as these, together with additive genetic factors, appear to be the largest contributors to adolescent exercise behavior.



9

Summary and general discussion



This thesis represents a genetic perspective on the association between exercise behavior and mental health. Longitudinal data (1991-2004) on regular exercise, symptoms of anxiety and depression, neuroticism, extraversion, sensation seeking, social problems and self-rated health were collected in twins and their family members registered at the Netherlands Twin Registry. The first part of this thesis described the population association between exercise behavior and symptoms of anxiety and depression and tested whether this association is causal or derives from a set of common genetic factors. The second part of this thesis aimed to further characterize the genetic basis of exercise behavior by applying genome-wide linkage and association techniques to exercise data from adults and by using advanced structural equation modeling of exercise data in parents and their adolescent offspring. This chapter provides a summary and discussion of the results, followed by some directions for future research.

Exercise behavior and symptoms of anxiety and depression

In chapters 2 to 4 it was reported that regular exercise in adults is associated with reduced symptoms of anxiety and depression, less neuroticism, more extraversion, more sensation seeking behaviors, fewer social problems and a better perceived health. These associations are small in size but hold for men and women of various ages (18 to 50 years). The results cor-

roborate previous findings on the relationship between exercise behavior and these variables (Allison et al., 2005; Arai & Hisamichi, 1998; Camacho et al., 1991; Farmer et al., 1988; Franken et al., 1994; Jack & Ronan, 1998; Potgieter & Bisschoff, 1990; Potgieter & Venter, 1995; Simonen et al., 2004; Weyerer, 1992).

To thoroughly test for causality in the association, I analyzed the cross-sectional and longitudinal association of exercise with symptoms of anxiety and depression for time intervals up to 11 years. A bivariate genetic model was used to test for causality in both the cross-sectional and longitudinal association. I argued that, if the association is fully explained by a causal effect of exercise on symptoms, *all* factors affecting exercise behavior should, indirectly and through the causal chain, also affect symptoms of anxiety and depression. In other words, under the causal hypothesis it is expected to find significant genetic and environmental correlations that together explain the association in a bivariate genetic model (since the causal chain is not explicitly modeled). Two additional methods were used to test for causality: the intra-pair differences method in MZ twins and longitudinal modeling of changes in exercise behavior and anxious and depressive symptoms in an individual over the years. All four methods did not point towards a causal effect of exercise on reduced symptoms of anxiety and depression. Rather, the association between exercise and these symptoms was best explained by a set of common genetic factors with opposite effects on exercise behavior and symptoms of anxiety, depression and neuroticism. Common genetic factors also explained the positive association between exercise behaviour and self-rated health (chapter 4), the latter reflecting different aspects of both physical and mental health (Eriksson et al., 2001; Idler & Benyamini, 1997).

Comparison with experimental studies on exercise, anxiety and depression

At first, these results seem at odds with findings from experimental studies demonstrating beneficial effects of exercise training on symptoms of anxiety and depression in clinical populations (Brosse et al., 2002; Dunn et al., 2005; Teychenne et al., 2008). There are several explanations for this discrepancy. First, population-based studies mainly include healthy subjects, whereas most experimental studies are conducted in subjects who suffer from an anxiety or depressive disorder. Clinical patients may respond differently to exercise and may have more room to improve than healthy subjects. Thus, there may be a floor effect of exercise on symptoms in the subjects who are included in population-based observational studies.

However, the effect of exercise on symptoms of anxiety and depression did not depend on the baseline levels of symptoms in our analyses, indicating that even in the subjects who initially score high on symptoms, an increase in exercise behavior does not reduce their symptoms.

A more relevant difference may be the setting of the exercise activities. In most experimental studies, exercise is monitored and part of a therapeutic program, whereas in this thesis the voluntary exercise behavior in people's every day life was studied. It may be that the anti-depressant effects only occur if exercise is prescribed and monitored within a therapeutic setting. In such a setting, the effects of exercising *per se* are confounded with positive feedback from supervisor, social interaction with other participants, and the often strong expectation of therapeutic efficacy of exercise (Blumenthal et al., 2007). The latter expectation is often (accidentally) amplified by the study recruitment procedure. It is reasonable to expect that a volunteer for "a study of exercise therapy for depression", if he or she is assigned to the exercise group, might tend to exhibit a favorable attitude toward exercise. Conversely, if a volunteer who signs up for "a study of exercise therapy for depression" is assigned to a non-exercise group (e.g., antidepressant medication, behavioral therapy, or a waiting list), it is reasonable to expect a certain degree of discontent and lack of belief in a good outcome.

A final difference between experimental and observational studies is that in experimental studies there may be a substantial selection bias, which is largely overcome in population-based observational studies. Experimental studies in clinical samples may only attract subjects who sought help for their problems and are willing to improve. In addition, exercise training studies may only attract subjects who are willing to engage in an exercise training program. These subjects may be the ones who have the experience that they are good at exercise, who like to exercise, or who have the strongest beliefs in the positive effects of exercise (Babyak et al., 2000). There may also be selective attrition; only those subjects who experience positive psychological effects complete the study.

Thus, the results from experimental studies cannot readily be generalized to the population at large. The results in this thesis show that the, often implicit, assumption in population-based studies and intervention programs that regular participation in exercise has beneficial psychological effects in *all* individuals in the population may not be valid. The results also imply that genetic factors need to be taken into account if we want to increase our understanding on the relationship between exercise and mental health. A major next step is to identify the genetic factors overlapping

between exercise behaviour, anxiety, and depression. Obvious areas to look for these pleiotropic genes are the neurobiological pathways in the brain that are likely to be involved in mood regulation, such as the norepinephrine, dopaminergic and serotonergic systems (Dunlop & Nemeroff, 2007; Hill et al., 1999; Lesch et al., 2003; Tremblay et al., 2002). Each of these pathways has been shown to be activated during exercise (Chaouloff, 1997; Dishman et al., 2006; Dishman, 1997).

One way to directly test for overlapping genes would be to test whether known genes for anxiety and depression are also associated with exercise behavior. This is however not an easy task. Although there have been numerous attempts to find genes for anxiety and depression, or related personality traits like neuroticism (Clement et al., 2002; Fullerton, 2006; Lopez-Leon et al., 2008; Middeldorp et al., 2008; Shifman et al., 2008; Sullivan et al., 2008), replication of association of genetic polymorphisms with these traits has been proven a difficult task. For example, a meta-analysis of gene-finding studies for major depressive disorder concludes that among the many susceptibility genes that have been studied, only six have been sufficiently replicated (the APOE, DRD4, GNB3, MTHFR, SLC6A3 and SLC6A4 genes) (Lopez-Leon et al., 2008). None of these genes showed a strong association to exercise behavior in the Dutch or American samples. Another strategy is to test for overlapping genes by first identifying the genes for exercise behavior, of which currently relatively little is known (Dishman et al., 2006; Rankinen et al., 2006a). Once identified, one could test in a next step whether these genes are also associated with mental health. In this thesis, I took this second approach by attempting to identify which genetic variants are associated with exercise behavior.

Genetics of exercise behavior

In chapters 5 to 7, several genomic regions and genetic variants were identified that are related to exercise behavior in adults. A genome-wide linkage scan showed that chromosomal region 19p13.3 was suggestively linked to exercise participation in Dutch adults ($LOD=2.18$). This region does not coincide with regions or genes found in previous smaller scaled linkage and association studies on exercise behavior and physical activity (Cai et al., 2006; Loos et al., 2005; Lorentzon et al., 2001; Salmen et al., 2003; Simonen et al., 2003a; Simonen et al., 2003b; Stefan et al., 2002; Winnicki et al., 2004), but it harbors a number of potentially interesting genes (the mus-

cle integrin-binding protein gene (MIBP), the thyroid receptor-interacting protein 10 gene (TRIP-10), the myosin IE gene (MYO1F), the endothelial differentiation G-protein coupled receptors 5 and 6 genes (EDG5 and EDG6), the thromboxane A2 receptor gene (TBXA2R) and the calponin-1 gene) that may be related to muscle performance and muscle blood flow.

A genome-wide linkage scan in British adult women showed that the chromosomal regions 3q22-q24 and 4q31-34 were linked to the maximum level of sports participation achieved. Suggestive linkages were found on chromosomes 3q22-q24 (at the sodium/hydrogen exchanger 9 (SLC9A9) gene) and 4q31-34 (near the fatty-acid binding protein 2 (FABP2) gene and the uncoupling protein 1 (UCP1) gene). The second region has been related to physical activity in a previous linkage study (Simonen et al., 2003b), an observation that is consistent with the hypothesis that some of the genetic influences on exercise behavior are mediated by exercise ability.

A genome-wide association analysis in Dutch and American adults revealed several novel SNPs in the SH3-domain GRB2-like (endophilin) interacting protein 1 (SGIP1) gene, the deoxyribonuclease II beta (DNASE2B) gene, the protease serine 16 (PRSS16) gene, the excision repair cross-complementing rodent repair deficiency, complementation group 2 (ERCC2) gene and the 3'-phosphoadenosine 5'-phosphosulfate synthase 2 (PAPSS2) gene that are associated with exercise participation. We replicated the association with two candidate genes for exercise behavior (Salmen et al., 2003; Stefan et al., 2002): the aromatase (CYP19A1) gene in the Dutch sample and the leptin receptor (LEPR) gene in the American sample. Associations of SNPs within several candidate linkage regions (Cai et al., 2006; De Moor et al., 2007a; Simonen et al., 2003b) were not replicated. Two of the genes found (SGIP1 and LEPR) are expressed in the hypothalamus and involved in the regulation of energy homeostasis (Park et al., 2006; Trevaskis et al., 2005). Their effects were independent of body mass index (BMI), suggesting a direct role of this pathway in the drive to exercise.

Our genome-wide association study, performed in two independent samples, constitutes the largest attempt in this field so far. However, overseeing the first wave of genome-wide association studies on complex traits and diseases, it is now clear that sample sizes must be increased much more (Craddock et al., 2008). For example, a recent genome-wide association analysis for adult height in 30,147 individuals identified 20 SNPs explaining only 3% of the variation in height (Weedon et al., 2008). This study illustrates that very large collaborative samples are needed to detect the small polygenic effects for complex traits such as height, and with-

out doubt also exercise behavior. Besides increased power to detect small effects, large collaborative samples have also shown to be useful to obtain more insight into the pathways through which genetic polymorphisms exert their influence on complex phenotypes (Craddock et al., 2008; McCarthy et al., 2008). An example of this is a recent genome-wide association study for type II diabetes combining data from multiple samples (Zeggini et al., 2007), in which association of the FTO gene was found in all samples except one in which the diabetes patients were selected based on low BMI. This showed that the effect of the FTO gene on type II diabetes was primarily mediated by adiposity (Frayling et al., 2007).

In addition to increases in sample size, an increase in the number and type of the genetic variants screened may also be required to fully characterize the heritability of exercise behavior. Many of the current genome-wide association studies are characterized by analyzing a set of common SNPs on the autosomal genome and then focusing on the most significant hits. Chapter 7 was no exception. There is good reason to suspect that additional analysis of the sex chromosomes and the mitochondrial DNA may yield a more complete picture. The genome-wide linkage scan for exercise participation reported in chapter 5 suggests that the genes affecting exercise in men and women are partly different. Analysis of sex chromosomes may shed more light on this sex heterogeneity in the genetic effects on exercise. Also, mitochondrial variants might well impact on exercise behavior. The mitochondrial DNA plays an important role in energy metabolism and variation at different sites of the mitochondria has already been related to several human diseases linked to exercise, such as type II diabetes (Lowell & Shulman, 2005; Saxena et al., 2006).

Other types of genomic variation such as insertion deletion polymorphisms, copy number variations (CNVs), methylation of the DNA and rare mutations should also be considered (Bodmer & Bonilla, 2008; McCarthy et al., 2008). For example, the 287-bp AluI repeat insertion-deletion polymorphism of the angiotensin converting enzyme (ACE) gene has been related to endurance performance, muscle efficiency and physical activity (Williams et al., 2000; Winnicki et al., 2004; Woods et al., 2000). The variation of SNPs in a genome-wide association study may not adequately capture the variation of other types of polymorphisms (Burgner et al., 2003) and therefore some important signals may be missed. Developments in technologies and methodologies to genotype and analyze these different types of genomic variation at a genome-wide basis will make it possible to extend the genetic analyses of exercise behavior and related traits.

Finally, new methodologies are needed that combine information

about genetic variation and known biological pathways and environmental factors relevant to exercise phenotypes. Methods have already been developed that incorporate information from previous linkage scans or candidate gene studies to prioritize specific regions or genes (Curtis et al., 2007; Li et al., 2008; Roeder et al., 2006). More recently a different approach was suggested to prioritize specific regions or genes by combining the p-values of individual SNPs for clusters of genes that have the same biological pathways (Wang et al., 2007). This method reduces the total number of tests and shifts the traditional focus on the most significant genes, to a focus on the most significant networks of genes that are involved in the same biological pathway or function. Of course, these methods all rely heavily on what is already known about the biology of the phenotype, but they may nevertheless provide a useful approach to genome-wide association analysis because it is more hypothesis-driven.

Hypotheses about genes for exercise

Pending the results from larger scaled GWA studies, with a more complete set of genetic markers, we can already speculate on the biological pathways that may prove important for exercise behavior. The first is homeostatic control over energy expenditure. Genes that affect the hypothalamic regulation of energy homeostasis have mainly been related to obesity and related metabolic diseases, such as type II diabetes mellitus (Liu et al., 2004; Park et al., 2006). The data reported in this thesis suggests that the regulation of energy homeostasis may also be related to the drive to spend energy and be physically active independently of BMI. This finding is consistent with the findings from two previous candidate gene association studies in which it was found that the LEPR and MC4R genes were related to physical activity independent of body composition phenotypes such as BMI (Loos et al., 2005; Stefan et al., 2002).

A second biological pathway through which genes may affect exercise behavior is through exercise ability. It is well-known that physical performance is genetically determined (Bouchard & Malina, 1998). Numerous twin studies show that different aspects of cardiorespiratory fitness and skeletal muscular strength and performance are heritable (Bouchard et al., 1998; Bouchard et al., 1999; De Mars et al., 2007; Perusse et al., 2001; Thomis et al., 1998), and some progress has been made in identifying the genetic variants that contribute to physical performance phenotypes (Rankinen et al., 2006a). These genetic variants help explain why one person is good at exercise and the other is not. The experience of being good at exercise, which probably starts early in life when children compare their

own ability and performance with that from friends and peers, may have a rewarding effect and this may stimulate an individual to continue to engage in exercise activities later in life. The experience of not being good at exercise and performing worse than the peer group may have punishing effects and may on the long term prevent an individual from regular participation in exercise.

A third biological pathway through which genes may affect exercise behavior is through the acute effects of exercise. Individuals who experience more rewarding than aversive effects after an acute bout of exercise are more likely to repeat their behavior and become regular lifetime exercisers. These rewarding effects may be physiological and psychological in nature (e.g., feeling 'energetic', less stressed, experiencing feelings of mastery and competence). In contrast, individuals for whom the rewarding effects do not outweigh the aversive effects will be less inclined to continue to exercise and are at higher risk to become and stay sedentary. A study in which persistent exercisers, recent adopters, fitness program dropouts and persistent sedentary individuals were extensively interviewed, suggested that regular exercisers differed from sedentary individuals mainly in that they enjoyed the exercise itself and felt that something was missing in their life when they did not regularly exercise (Gauvin, 1990). Similar to the genetic influences on changes in physical fitness as a result of exercise training (Bouchard et al., 1999; Perusse et al., 2001), it is hypothesized that the acute psychological effects of exercise training are also partly under genetic control.

A fourth pathway that could mediate the relationship between genes and exercise behavior is personality. It has been shown in chapter 2 that regular exercisers differ from non-exercisers on extraversion and sensation seeking. I have not yet tested whether the association with these traits reflects causal effects of personality on exercise behavior, but this is a distinct possibility. Individuals who are inclined to seek experience, thrills and adventure, or to enrich their social environment, may achieve this by participation in (competitive) sports. Sensation seeking behavior and extraversion are influenced by genetic factors (Johnson & Krueger, 2004; Stoel et al., 2006; Viken et al., 1994). Since it is well-known that the dopaminergic system in the brain is involved in experiencing pleasure and reward (Dunlop & Nemeroff, 2007) dopaminergic genes have been implied as an important source of this heritability. The finding that the observed decline of physical activity with age can be explained by depleted dopamine release and loss of dopamine receptors in the brain in human and non-human species (Ingram, 2000) is consistent with the hypothesis that

dopamine is involved in long-term exercise participation. Moreover, the dopamine 2 receptor gene is one of the few genes that has been implicated in physical activity in humans (Simonen et al., 2003a). Thus, one could speculate that genetic variants that are involved in the extraversion and sensation seeking are also involved in the drive to exercise.

Shared environment in adolescent exercise behavior

In contrast to the contribution of genetic factors to exercise behavior in adults (Beunen & Thomis, 1999; Eriksson et al., 2006; Kujala et al., 2002; Lauderdale et al., 1997; Stubbe et al., 2006a), exercise behavior in adolescents is largely influenced by shared environmental factors (Carlsson et al., 2006; Maia et al., 2002; Stubbe et al., 2005). In chapter 8, I used a parent-offspring design, including parents and siblings of adolescent twins, in order to determine whether the shared environment in adolescents is best explained by the influence of parental exercise behavior on their offspring's exercise behavior, by environmental factors specific to the adolescent generation, or by the effects of assortative mating. Since there are substantial differences in variance decomposition of exercise in both generations and also across sex, I extended the parent-offspring model to account for sex and generation differences in variance decomposition of exercise. Furthermore, data from adult twins and spouses were used to test for different causes of the spouse correlation (Heath & Eaves, 1985). It was found that the spouse correlation did not increase as a function of duration of the relationship. The spouse correlation was most likely because of phenotypic assortment.

The shared environment in adolescents mainly consists of generation-specific influences. There was little evidence for the influence of parental exercise behavior on their children's exercise behavior, except for the influence of fathers' behavior on sons' sports participation. The model tested the importance of parental exercise behavior on offspring exercise behavior. Parental influences that are unrelated to a parent's own exercise behavior, for example through social support mechanisms, cannot be ruled out. Future research needs to include measures of parental attitudes and social support towards children's exercise behavior to resolve whether there is cultural transmission of for example attitudes from parents to offspring. It also needs to focus on the generation specific environmental factors on adolescent exercise behavior (including peer behavior), as these, together

with additive genetic factors, appear to be the largest contributors to adolescent exercise behavior. It would further be of interest to examine the association between exercise behavior and well-being in adolescents and to test for causality. Since other factors contribute to regular exercise participation in adolescence than in adulthood, the mechanisms explaining the association between exercise and mental health may also be different.

Gene by exercise interaction

This thesis provides evidence for common genetic factors influencing exercise behavior and symptoms of anxiety and depression. The genetic influences on anxiety and depression may however also be *moderated* by exercise. Put differently, the effects of exercise on reduced symptoms of anxiety and depression may depend on an individual's genotype. This would mean that the causal effect of exercise on improved mood only holds for subgroups of the population. The possibility of gene by exercise interaction is consistent with the absence of a causal effect in the population at large and the presence of a causal effect in smaller subgroups of the population. It could also help explain the individual differences in the rewarding versus aversive psychological effects in response to exercise. As far as I know, there are no studies that investigated the possibility of gene by exercise interaction on reduced symptoms of anxiety and depression.

To test whether the genetic factors influencing mental health problems such as anxiety and depression can also be moderated by exercise, a research design in which twins participate in an exercise training program is needed. This design has already been successfully applied to the physiological effects of exercise training (Bouchard et al., 1999; Bouchard et al., 2000; Perusse et al., 2001), but these studies did not report psychological effects, such as changes in symptoms of anxiety and depression. A randomized controlled trial including extensive measurements at baseline and follow-up will provide more insight into the potential role of gene-exercise interaction in the association between exercise behavior and anxiety and depression. By including a wide range of measurements such as (changes in) cardiorespiratory fitness, muscle strength, anxiety, depression and attitudes towards exercise, it becomes possible to test the hypothesized mediating and moderating mechanisms between genetic variation at one hand and exercise and mental health at the other hand. Structural equation modeling techniques in twin pairs can for example be used to test whether the heritability of symptoms of anxiety and depression is a function of exercise

training (Purcell, 2002). Genotyping of twins will make it possible to study the interaction effect of measured genetic polymorphisms with exercise training on anxious and depressive symptoms (Fulker et al., 1999; van der Sluis et al., 2008a). Recent power studies show that a few hundred twin pairs would be needed to reliably detect such gene by exercise interaction effects (van der Sluis et al., 2008a; van der Sluis et al., 2008b).

Future perspective

The findings in this thesis have implications for population-based exercise intervention and prevention programs. Currently, findings on population associations between exercise and mental health are directly translated to these programs, assuming that the population association reflects a causal effect of exercise on well-being and assuming that exercise has the same beneficial effect on *all* participants. The results from this thesis do not support these two assumptions and urge for a change in perspective from less 'population-based' to more 'personalized' exercise intervention strategies. This requires an increased understanding of the pathways from genes to exercise behavior and of the differences in genetic sensitivity to the mental health benefits of exercise.



Nederlandse samenvatting
(Dutch summary)



Waarom doen sommige personen wel aan sport en anderen niet? In hoeverre speelt genetische aanleg voor sportgedrag een rol? Hoe hangt sportgedrag samen met geestelijke gezondheidsklachten, zoals angst en depressie? Zijn sporters minder angstig en depressief dan niet-sporters en hoe kunnen we dit verklaren? Deze vragen vormen de basis van dit proefschrift, getiteld "Sportgedrag en geestelijke gezondheid: een genetisch perspectief". In het eerste deel van het proefschrift wordt het verband beschreven tussen sportgedrag en verschillende aspecten van geestelijke gezondheid, waaronder angst en depressie, in de Nederlandse populatie. Daarnaast is de vraag onderzocht hoe het verband tussen sportgedrag en geestelijke gezondheid verklaard kan worden. Naast de gangbare causale hypothese dat sporten een betere geestelijke gezondheid veroorzaakt, is ook een alternatieve hypothese getoetst. Deze alternatieve hypothese stelt dat het verband tussen sportgedrag en geestelijke gezondheid verklaard wordt door een derde, onderliggende, factor. Dit kan een genetische factor zijn die zowel sportgedrag en geestelijke gezondheid beïnvloedt. Het tweede deel van dit proefschrift heeft tot doel de genetische basis van sportgedrag verder te karakteriseren. Allereerst is er systematisch gekeken of er locaties aan te wijzen zijn op het menselijk genoom waar de genen liggen die sportgedrag en sportniveau beïnvloeden. Vervolgens is getoetst of er specifieke genetische varianten zijn die een invloed hebben op sportgedrag. Tot slot is onderzocht of ouders naast hun genen ook hun sportcultuur doorgeven aan hun adolescente kinderen.

In dit proefschrift is gebruik gemaakt van vragenlijstgegevens die tussen 1991 en 2004 zijn verzameld bij het Nederlands Tweelingen Register

(NTR). Om de 2 à 3 jaar ontvangen adolescente en volwassen tweelingen en hun familieleden per post een vragenlijst over hun leefstijl, geestelijke en lichamelijke gezondheid en persoonlijkheid. In totaal hebben ruim 20.000 personen ooit meegedaan aan dit onderzoek. Ongeveer de helft van deze personen is tweeling. De andere helft is familie van een tweeling (broers en zussen, ouders, partners en kinderen). In dit proefschrift zijn de gegevens geanalyseerd van deze personen over sportgedrag, depressieve symptomen, angstklachten, neuroticisme, extraversie, spanningsbehoefte, sociale problemen en zelfbeoordeelde gezondheid. De gegevens over sportniveau uit hoofdstuk 6 komen van Britse tweelingen die ingeschreven staan bij het TwinsUK Adult Twin Registry. Voor de studie in hoofdstuk 7 is naast gegevens van personen uit het NTR ook gebruik gemaakt van gegevens van een Amerikaanse steekproef. Dit hoofdstuk geeft een overzicht van de belangrijkste resultaten uit dit proefschrift.

Sportgedrag en symptomen van angst en depressie

In de hoofdstukken 2 tot en met 4 is vastgesteld dat volwassenen die regelmatig sporten gemiddeld minder last hebben van depressieve en angstige klachten, minder neurotisch zijn en minder sociale problemen hebben. Ook zijn sporters gemiddeld meer extravert, hebben zij meer behoefte aan spanning en zij beoordelen hun gezondheid vaker als goed. Verder zijn personen die regelmatig aan sport doen een paar jaar later nog steeds minder angstig en depressief. Dat geldt voor zowel kortere periodes van 2 en 4 jaar, als voor langere periodes van 7, 9 of zelfs 11 jaar. Het verband tussen sport en angstige en depressieve symptomen is, zowel cross-sectioneel als longitudinaal, echter niet sterk met effectgroottes die variëren tussen de -0.14 en -0.29 standaarddeviaties en correlaties die lopen van -0.06 tot -0.14. Dat wil zeggen dat er ook personen zijn die niet aan sport doen maar helemaal geen last hebben van angsten, depressies of gerelateerde problemen. Andersom komt ook voor; personen die veel aan sport zijn kunnen desondanks angstig of depressief zijn. Het verband tussen sportgedrag en geestelijke gezondheid geldt voor mannen en vrouwen van alle leeftijden (18 tot en met 50 jaar). Deze bevindingen komen overeen met eerdere studies naar het verband tussen sportgedrag en deze variabelen.

Vervolgens is met behulp van 3 verschillende methodes getoetst of het verband tussen sportgedrag en symptomen van angst en depressie verklaard kan worden door een causaal effect van sporten op angst en

depressie of door een set genetische factoren die zowel een invloed heeft op inactief gedrag als op het risico op angst en depressie. Met behulp van tweelingdata kan vastgesteld worden of een eigenschap erfelijk is. Ook kan onderzocht worden of het verband tussen twee erfelijke eigenschappen verklaard kan worden door dezelfde erfelijke factoren. Uit de eerste methode (een bivariaat genetisch model) blijkt dat de erfelijke factoren die een rol spelen bij angstige en depressieve klachten voor een deel overlappen met de erfelijke factoren voor sportgedrag. Deze bevinding verklaart de correlatie die wordt waargenomen op populatieniveau tussen sport en geestelijke gezondheid. Als het hele verband tussen sport en geestelijke gezondheid verklaard wordt door een set van onderliggende genetische factoren volgt daaruit een intrigerende voorspelling voor het verband tussen sport en geestelijke gezondheid bij eenige tweelingen. Als de tweeling die meer aan sport doet minder angstig of depressief is dan de tweelingbroer of -zus die niet aan sport doet, kan dit niet verklaard worden door genetische verschillen tussen deze personen en heeft sport mogelijk een causaal effect op minder angst en depressie. Als echter de persoon die meer aan sport doet niet minder angstig of depressief is dan zijn of haar niet-sportende tweelinghelft dan duidt dit erop dat genetische factoren verantwoordelijk zijn voor het verband tussen sportgedrag en geestelijke gezondheid. Immers, op het moment dat voor genetische factoren gecorrigeerd wordt, verdwijnt het verband. Uit deze tweede methode blijkt dat het verband tussen sportgedrag en geestelijke gezondheid verklaard wordt door gedeelde genetische factoren. Uit de derde gebruikte methode (een longitudinale analyse) blijkt dat wanneer personen met de jaren meer zijn gaan sporten dit niet leidt tot minder depressieve klachten. Omgekeerd blijkt ook niet het geval: personen die gestopt zijn met sporten worden niet depressiever. Sportgedrag is dus niet per sé de oorzaak voor minder angsten en depressies, maar erfelijke factoren beïnvloeden zowel iemands behoefte om regelmatig aan sport te doen als iemands geestelijke gezondheid. Sportgedrag kan gezien worden als een van de uitingen van een goede mentale gezondheid. Als personen stoppen met sporten verandert dit hun genetische aanleg voor het ontwikkelen van angsten en depressies niet en zullen zij door het stoppen met sporten ook niet onmiddellijk angstig of depressief worden.

Deze bevindingen staan in sterk contrast met de resultaten uit eerdere experimentele onderzoeken, waaruit blijkt dat een sportprogramma angstige en depressieve klachten kan verminderen. Een verklaring voor deze verschillen is dat wij sportgedrag in de vrije tijd hebben bestudeerd, dat wordt geïnitieerd door de persoon zelf. De eerdere experimentele onder-

zoeken analyseerden de effecten van begeleid sporten, vaak als onderdeel van een therapie. Een andere verklaring is dat sporten wel een causaal effect op geestelijke gezondheid kan hebben, maar dat dit effect maar bij een deel van de mensen optreedt, waardoor het niet zichtbaar wordt als je een steekproef uit de totale bevolking neemt. Dat het maar bij een deel van de mensen optreedt kan worden verklaard door verschillen in de genetische aanleg die de ene persoon wel maar de andere niet gevoelig maken voor de psychologische effecten van sporten. Deze genetische aanleg zou mogelijk kunnen overlappen met de genetische aanleg voor sportgedrag.

Genetische aanleg voor sportgedrag

In de hoofdstukken 5 tot en met 8 is de genetische basis van sportgedrag verder gekarakteriseerd. In hoofdstuk 5 is vastgesteld dat er op chromosoom 19 een specifieke regio is (p13.3) waar genen liggen die mogelijk sportgedrag beïnvloeden via hun effecten op spierkracht en de bloedstroming naar de spieren. In hoofdstuk 6 zijn 2 regio's op het menselijk genoom ontdekt (op de chromosomen 3 en 4) die gekoppeld zijn aan het maximale sportniveau dat personen ooit bereikt hebben. De regio op chromosoom 4 is in een eerdere studie ook al gekoppeld aan lichamelijke activiteit, wat consistent is met de hypothese dat de invloed van genen op sportgedrag voor een deel de genetische aanleg weergeeft om goed te zijn in sport.

In hoofdstuk 7 zijn een aantal verdere genetische varianten geïdentificeerd die mogelijk samenhangen met sportgedrag. Deze varianten liggen niet binnen de chromosomale regio's die al in de hoofdstukken 5 en 6 waren geïdentificeerd, maar verschaffen nieuwe inzichten in hoe genetische varianten een invloed kunnen hebben op sportgedrag. Twee genen (SGIP1 en LEPR) komen tot expressie in de hypothalamus van de hersenen en spelen een rol bij de regulatie van de energiebalans in het lichaam (de balans tussen energie-inname en -verbranding). De varianten in deze genen beïnvloeden het sportgedrag onafhankelijk van lichaamsgewicht, wat duidt op een directe rol van de hypothalamus in de behoefte om lichamelijk actief te zijn.

Uit hoofdstuk 8 blijkt dat de gelijkens tussen ouders en hun kinderen (tussen 13 en 18 jaar) bijna volledig verklaard kan worden doordat ouders hun genen doorgeven aan hun kinderen. De directe invloed van het sportgedrag van de ouders op dat van hun kinderen lijkt minimaal; alleen voor zoons lijkt er een effect van het sportgedrag van de vader te zijn. Naast genetische factoren zijn omgevingsinvloeden die specifiek zijn voor

adolescenten bepalend voor het sportgedrag van adolescenten. Dit kunnen invloeden zijn van bijvoorbeeld vrienden, leeftijdsgenoten en de school.

Toekomstig onderzoek

De resultaten van dit proefschrift kunnen belangrijke implicaties hebben voor bevolkingsbrede sportinterventie- en preventieprogramma's. Op dit moment worden onderzoeksresultaten over het verband tussen sport en geestelijke gezondheid vaak regelrecht vertaald naar interventie- en preventieprogramma's. Hierbij wordt aangenomen dat sportgedrag de oorzaak is voor een verbeterde stemming en dat dit (in dezelfde mate) geldt voor alle deelnemers. Meer inzicht in individuele verschillen in psychologische effecten van sportgedrag en de rol van genetische factoren hierin leidt hopelijk in de toekomst tot sport interventies die meer toegesneden zijn op de specifieke behoeftes en vermogens van het individu.



References



- Aarnio, M., Winter, T., Kujala, U. M., & Kaprio, J. (1997). Familial aggregation of leisure-time physical activity - A three generation study. *International Journal of Sports Medicine*, 18, 549-556.
- Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30, 97-101.
- Abecasis, G. R., Cherny, S. S., Cookson, W. O. C., & Cardon, L. R. (2001). GRR: graphical representation of relationship errors. *Bioinformatics*, 17, 742-743.
- Abecasis, G. R. & Wigginton, J. E. (2005). Handling marker-marker linkage disequilibrium: Pedigree analysis with clustered markers. *American Journal of Human Genetics*, 77, 754-767.
- Achenbach, T. M. (1990). *The Young Adult Self Report*. Burlington, VT: University of Vermont, Dept of Psychiatry.
- Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J. et al. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise*, 32, S498-S516.
- Albright, A., Franz, M., Hornsby, G., Kriska, A., Marrero, D., Ullrich, I. et al. (2000). Exercise and type 2 diabetes. *Medicine and Science in Sports and Exercise*, 32, 1345-1360.
- Allison, K. R., Adlaf, E. M., Irving, H. M., Hatch, J. L., Smith, T. F., Dwyer, J. J. M. et al. (2005). Relationship of vigorous physical activity to psychologic distress among adolescents. *Journal of Adolescent Health*, 37, 164-166.
- Altshuler, D., Brooks, L. D., Chakravarti, A., Collins, F. S., Daly, M. J., & Donnelly, P. (2005). A haplotype map of the human genome. *Nature*, 437, 1299-1320.
- Arai, Y. & Hisamichi, S. (1998). Self-reported exercise frequency and personality: A population-based study in Japan. *Perceptual and Motor Skills*, 87, 1371-1375.
- Babyak, M., Blumenthal, J. A., Herman, S., Khatri, P., Doraiswamy, M., Moore, K. et al. (2000). Exercise treatment for major depression: Maintenance of therapeutic benefit at 10 months. *Psychosomatic Medicine*, 62, 633-638.
- Barbour, K. A. & Blumenthal, J. A. (2002). Exercise training and depression in older adults. *Neurobiology of aging*, 26, S118-S123.
- Barrett, J. C. & Cardon, L. R. (2006). Evaluating coverage of genome-wide association studies. *Nature Genetics*, 38, 659-662.
- Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21, 263-265.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for measuring depression. *Archives of General Psychiatry*, 4, 561-571.
- Berlin, J. A. & Colditz, G. A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *American Journal of Epidemiology*, 132, 612-628.

References

- Beunen, G. & Thomis, M. (1999). Genetic determinants of sports participation and daily physical activity. *International Journal of Obesity*, 23, S55-S63.
- Birney, E., Andrews, D., Caccamo, M., Chen, Y., Clarke, L., Coates, G. et al. (2006). Ensembl 2006. *Nucleic Acids Research*, 34, D556-D561.
- Blumenthal, J. A., Babyak, M. A., Doraiswamy, P. M., Watkins, L., Hoffman, B. M., Barbour, K. A. et al. (2007). Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosomatic Medicine*, 69, 587-596.
- Bodmer, W. & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics*, 40, 695-701.
- Boomsma, D. I., Beem, A. L., van den, B. M., Dolan, C. V., Koopmans, J. R., Vink, J. M. et al. (2000). Netherlands twin family study of anxious depression (NETSAD). *Twin Research*, 3, 323-334.
- Boomsma, D. I., Cacioppo, J. T., Slagboom, P. E., & Posthuma, D. (2006a). Genetic linkage and association analysis for loneliness in Dutch twin and sibling pairs points to a region on chromosome 12q23-24. *Behavior Genetics*, 36, 137-146.
- Boomsma, D. I., De Geus, E. J. C., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J. et al. (2006b). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics*, 9, 849-857.
- Boomsma, D. I. & Molenaar, P. C. M. (1987). Constrained maximum-likelihood analysis of familial resemblance of twins and their parents. *Acta Geneticae Medicae et Gemellologiae*, 36, 29-39.
- Boomsma, D. I., Vandenberg, M. B. M., Orlebeke, J. F., & Molenaar, P. C. M. (1989). Resemblances of parents and twins in sports participation and heart-rate. *Behavior Genetics*, 19, 123-141.
- Boomsma, D. I., Vink, J. M., van Beijsterveldt, T. C., De Geus, E. J., Beem, A. L., Mulder, E. J. et al. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research*, 5, 401-406.
- Boomsma, D. I., Willemsen, G., Sullivan, P. F., Heutink, P., Meijer, P., Sondervan, D. et al. (2008). Genome-wide association of major depression: Description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *European Journal of Human Genetics*, 16, 335-342.
- Bouchard, C., An, P., Rice, T., Skinner, J. S., Wilmore, J. H., Gagnon, J. et al. (1999). Familial aggregation of VO₂max response to exercise training: Results from the HERITAGE Family Study. *Journal of Applied Physiology*, 87, 1003-1008.
- Bouchard, C., Daw, E. W., Rice, T., Perusse, L., Gagnon, J., Province, M. A. et al. (1998). Familial resemblance for VO₂max in the sedentary state: The HERITAGE Family Study. *Medicine and Science in Sports and Exercise*, 30, 252-258.
- Bouchard, C. & Malina, R. M. (1998). Genetics of physiological fitness and motor performance. *Exercise and Sport Sciences Reviews*, 11, 306-339.

- Bouchard, C., Rankinen, T., Chagnon, Y. C., Rice, T., Perusse, L., Gagnon, J. et al. (2000). Genomic scan for maximal oxygen uptake and its response to training in the HERITAGE Family Study. *Journal of Applied Physiology*, 88, 551-559.
- Brosse, A. L., Sheets, E. S., Lett, H. S., & Blumenthal, J. A. (2002). Exercise and the treatment of clinical depression in adults - Recent findings and future directions. *Sports Medicine*, 32, 741-760.
- Brown, W. J., Ford, J. H., Burton, N. W., Marshall, A. L., & Dobson, A. J. (2005). Prospective study of physical activity and depressive symptoms in middle-aged women. *American Journal of Preventive Medicine*, 29, 265-272.
- Brutsaert, T. D. & Parra, E. J. (2006). What makes a champion? Explaining variation in human athletic performance. *Respiratory Physiology & Neurobiology*, 151, 109-123.
- Burgner, D., Rockett, K., Ackermann, H., Hull, J., Usen, S., Pinder, M. et al. (2003). Haplotypic relationship between SNP and microsatellite markers at the NOS2A locus in two populations. *Genes and Immunity*, 4, 506-514.
- Burstrom, B. & Fredlund, P. (2001). Self rated health: Is it as good a predictor of subsequent mortality among adults in lower as well as in higher social classes? *Journal of Epidemiology and Community Health*, 55, 836-840.
- Byrne, A. & Byrne, D. G. (1993). The effect of exercise on depression, anxiety and other mood states - A review. *Journal of Psychosomatic Research*, 37, 565-574.
- Cai, G. W., Cole, S. A., Butte, N., Bacino, C., Diego, V., Tan, K. et al. (2006). A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. *Obesity Research*, 14, 1596-1604.
- Camacho, T. C., Roberts, R. E., Lazarus, N. B., Kaplan, G. A., & Cohen, R. D. (1991). Physical activity and depression: Evidence from the Alameda County Study. *American Journal of Epidemiology*, 134, 220-231.
- Carlsson, S., Andersson, T., Lichtenstein, P., Michaelsson, K., & Ahlbom, A. (2006). Genetic effects on physical activity: Results from the Swedish twin registry. *Medicine and Science in Sports and Exercise*, 38, 1396-1401.
- Caspersen, C. J., Pereira, M. A., & Curran, K. M. (2000). Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Medicine and Science in Sports and Exercise*, 32, 1601-1609.
- Caspi, A. & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, 12, 49-66.
- Chaouloff, F. (1997). Effects of acute physical exercise on central serotonergic systems. *Medicine and Science in Sports and Exercise*, 29, 58-62.
- Clement, Y., Calatayud, F., & Belzung, C. (2002). Genetic basis of anxiety-like behaviour: A critical review. *Brain Research Bulletin*, 57, 57-71.
- Cloninger, C. R. (1980). Interpretation of intrinsic and extrinsic structural relations by path analysis: Theory and applications to assortative mating. *Genetical Research*, 36, 133-145.

- Cohen, J. (1969). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press.
- Cooper-Patrick, L., Ford, D. E., Mead, L. A., Chang, P. P., & Klag, M. J. (1997). Exercise and depression in midlife: A prospective study. *American Journal of Public Health, 87*, 670-673.
- Costa, P. T., Jr. & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38*, 668-678.
- Costa, P. T. & McCrae, R. R. (1992). *Professional manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor-Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- Craddock, N., O'Donovan, M. C., & Owen, M. J. (2008). Genome-wide association studies in psychiatry: Lessons from early studies of non-psychiatric and psychiatric phenotypes. *Molecular Psychiatry, 13*, 649-653.
- Craft, L. L. & Landers, D. M. (1998). The effect of exercise on clinical depression and depression resulting from mental illness: A meta-analysis. *Journal of Sport and Exercise Psychology, 20*, 339-357.
- Curtis, D., Vine, A. E., & Knight, J. (2007). A pragmatic suggestion for dealing with results for candidate genes obtained from genome wide association studies. *BMC Genetics, 8*.
- De Geus, E. J., van Doornen, L. J., & Orlebeke, J. F. (1993). Regular exercise and aerobic fitness in relation to psychological make-up and physiological stress reactivity. *Psychosomatic Medicine, 55*, 347-363.
- De Geus, E. J. C., Boomsma, D. I., & Snieder, H. (2003). Genetic correlation of exercise with heart rate and respiratory sinus arrhythmia. *Medicine and Science in Sports and Exercise, 35*, 1287-1295.
- De Geus, E. J. C. & De Moor, M. H. M. (2008). A genetic perspective on the association between exercise and mental health. *Mental Health and Physical Activity*, in press.
- De Mars, G., Thomis, M. A. I., Windelinckx, A., Van Leemputte, M., Maes, H. H., Blimkie, C. J. et al. (2007). Covariance of isometric and dynamic arm contractions: Multivariate genetic analysis. *Twin Research and Human Genetics, 10*, 180-190.
- De Moor, M. H. M., Beem, A. L., Stubbe, J. H., Boomsma, D. I., & De Geus, E. J. C. (2006). Regular exercise, anxiety, depression and personality: A population-based study. *Preventive Medicine, 42*, 273-279.
- De Moor, M. H. M., Boomsma, D. I., Stubbe, J. H., Willemsen, G., & De Geus, E. J. C. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of General Psychiatry, 65*, 897-905.

- De Moor, M. H. M., Posthuma, D., Hottenga, J. J., Willemsen, A. H. M., Boomsma, D. I., & De Geus, E. J. C. (2007a). Genome-wide linkage scan for exercise participation in Dutch sibling pairs. *European Journal of Human Genetics*, 15, 1252-1259.
- De Moor, M. H. M., Spector, T. D., Cherkas, L. F., Falchi, M., Hottenga, J. J., Boomsma, D. I. et al. (2007b). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics*, 10, 812-820.
- De Silva, M. G., Elliott, K., Dahl, H. H., Fitzpatrick, E., Wilcox, S., Delatycki, M. et al. (2003). Disruption of a novel member of a sodium/hydrogen exchanger family and DOCK3 is associated with an attention deficit hyperactivity disorder-like phenotype. *Journal of Medical Genetics*, 40, 733-740.
- Deng, H. W., Deng, H. Y., Liu, Y. J., Liu, Y. Z., Xu, F. H., Shen, H. et al. (2002). A genomewide linkage scan for quantitative-trait loci for obesity phenotypes. *American Journal of Human Genetics*, 70, 1138-1151.
- Dishman, R. K., Berthoud, H. R., Booth, F. W., Cotman, C. W., Edgerton, R., Fleshner, M. R. et al. (2006). Neurobiology of exercise. *Obesity*, 14, 345-356.
- Dishman, R. K. (1997). Brain monoamines, exercise, and behavioral stress: Animal models. *Medicine and Science in Sports and Exercise*, 29, 63-74.
- Distel, M. A., Ligthart, L., Willemsen, G., Nyholt, D. R., Trull, T. J., & Boomsma, D. I. (2007). Personality, health and lifestyle in a questionnaire family study: A comparison between highly cooperative and less cooperative families. *Twin Research and Human Genetics*, 10, 348-353.
- Duffy, D. L. (2006). An integrated genetic map for linkage analysis. *Behavior Genetics*, 36, 4-6.
- Duffy, D. L. & Martin, N. G. (1994). Inferring the direction of causation in cross-sectional twin data: Theoretical and empirical considerations. *Genetic Epidemiology*, 11, 483-502.
- Dunlop, B. W. & Nemeroff, C. B. (2007). The role of dopamine in the pathophysiology of depression. *Archives of General Psychiatry*, 64, 327-337.
- Dunn, A. L., Trivedi, M. H., Kampert, J. B., Clark, C. G., & Chambliss, H. O. (2005). Exercise treatment for depression - Efficacy and dose response. *American Journal of Preventive Medicine*, 28, 1-8.
- Eaves, L. J., Last, K. A., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human behavior. *Heredity*, 41, 249-320.
- Eaves, L. J., Silberg, J. L., & Maes, H. H. (2005). Revisiting the children of twins: Can they be used to resolve the environmental effects of dyadic parental treatment on child behavior? *Twin Research and Human Genetics*, 8, 283-290.
- Eriksson, I., Unden, A. L., & Elofsson, S. (2001). Self-rated health. Comparisons between three different measures. Results from a population study. *International Journal of Epidemiology*, 30, 326-333.

- Eriksson, M., Rasmussen, F., & Tynelius, P. (2006). Genetic factors in physical activity and the equal environment assumption - The Swedish young male twins study. *Behavior Genetics*, 36, 238-247.
- Eysenck, H. J. & Eysenck, S. B. G. (1964). *Eysenck Personality Inventory*. San Diego, CA: Educational and Industrial Testing Service.
- Falconer, D. S. & Mackay, T. F. C. (1996). *Introduction to quantitative genetics*. (4 ed.) Essex: Pearson Education Limited.
- Farmer, M. E., Locke, B. Z., Moscicki, E. K., Dannenberg, A. L., Larson, D. B., & Radloff, L. S. (1988). Physical activity and depressive symptoms: The NHANES I epidemiologic follow-up study. *American Journal of Epidemiology*, 128, 1340-1351.
- Feij, J. A., Dekker, P. H., Koopmans, J. R., & Boomsma, D. I. (1997). Nieuwe normen en stabiliteitsgegevens voor de Spanningsbehoefte lijst (SBL). *Nederlands Tijdschrift voor de Psychologie*, 52, 131-134.
- Feij, J. A. & Van Zuilen, R. W. (1984). *Handleiding bij de Spanningsbehoefte lijst (SBL)*. Lisse: Swets & Zeitlinger.
- Ferreira, M. A. R. (2004). Linkage analysis: Principles and methods for the analysis of human quantitative traits. *Twin Research*, 7, 513-530.
- Folkens, C. H. & Sime, W. E. (1981). Physical fitness training and mental health. *American Psychologist*, 36, 373-389.
- Franken, R. E., Hill, R., & Kierstead, J. (1994). Sport interest as predicted by the personality measures of competitiveness, mastery, instrumentality, expressivity, and sensation seeking. *Personality and Individual Differences*, 17, 467-476.
- Franz, S. I. & Hamilton, G. V. (1905). The effects of exercise upon the retardation in conditions of depression. *American Journal of Insanity*, 62, 239-256.
- Frayling, T. M., Timpson, N. J., Weedon, M. N., Zeggini, E., Freathy, R. M., Lindgren, C. M. et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, 316, 889-894.
- Fulker, D. W. (1989). Extensions of the classical twin method. In B.S.Weir, E. J. Eisen, Goodman.M.M., & G. Namkoong (Eds.), *Proceedings of the second international conference on quantitative genetics* (pp. 395-406). Sunderland, MA: Sinauer Associates, Inc.
- Fulker, D. W., Cherny, S. S., Sham, P. C., & Hewitt, J. K. (1999). Combined linkage and association sib-pair analysis for quantitative traits. *American Journal of Human Genetics*, 64, 259-267.
- Fullerton, J. (2006). New approaches to the genetic analysis of neuroticism and anxiety. *Behavior Genetics*, 36, 147-161.

- Gauvin, L. (1990). An experiential perspective on the motivational features of exercise and life-style. *Canadian Journal of Sport Sciences-Revue Canadienne des Sciences du Sport*, 15, 51-58.
- Gauvin, L. & Spence, J. C. (1996). Physical activity and psychological well-being: Knowledge base, current issues, and caveats. *Nutrition Review*, 54, S53-S65.
- Ge, D. L., Zhang, K. L., Need, A. C., Martin, O., Fellay, J., Urban, T. J. et al. (2008). WGAViewer: Software for genomic annotation of whole genome association studies. *Genome Research*, 18, 640-643.
- Goldfarb, A. H. & Jamurtas, A. Z. (1997). Beta-endorphin response to exercise - An update. *Sports Medicine*, 24, 8-16.
- Greist, J. H., Klein, M. H., Eischens, R. R., Faris, J., Gurman, A. S., & Morgan, W. P. (1979). Running As Treatment for Depression. *Comprehensive Psychiatry*, 20, 41-54.
- Gustafson, S. L. & Rhodes, R. E. (2006). Parental correlates of physical activity in children and early adolescents. *Sports Medicine*, 36, 79-97.
- Haase, A., Steptoe, A., Sallis, J. F., & Wardle, J. (2004). Leisure-time physical activity in university students from 23 countries: associations with health beliefs, risk awareness, and national economic development. *Preventive Medicine*, 39, 182-190.
- Haseman, J. K. & Elston, R. C. (1972). Investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, 2, 3-19.
- Heath, A. C. & Eaves, L. J. (1985). Resolving the effects of phenotype and social background on mate selection. *Behavior Genetics*, 15, 15-30.
- Heath, A. C., Kendler, K. S., Eaves, L. J., & Markell, D. (1985). The resolution of cultural and biological inheritance - Informativeness of different relationships. *Behavior Genetics*, 15, 439-465.
- Heath, A. C., Kessler, R. C., Neale, M. C., Hewitt, J. K., Eaves, L. J., & Kendler, K. S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics*, 23, 29-50.
- Hettema, J. M., Neale, M. C., & Kendler, K. S. (2001). A review and meta-analysis of the genetic epidemiology of anxiety disorders. *American Journal of Psychiatry*, 158, 1568-1578.
- Hill, S. Y., Zezza, N., Wipprecht, G., Locke, J., & Neiswanger, K. (1999). Personality traits and dopamine receptors (D2 and D4): Linkage studies in families of alcoholics. *American Journal of Medical Genetics*, 88, 634-641.
- Hirschhorn, J. N. & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6, 95-108.
- Idler, E. L. & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, 38, 21-37.

- Ikeda, T., Mabuchi, A., Fukuda, A., Hiraoka, H., Kawakami, A., Yamamoto, S. et al. (2001). Identification of sequence polymorphisms in two sulfation-related genes, PAPSS2 and SLC26A2, and an association analysis with knee osteoarthritis. *Journal of Human Genetics*, 46, 538-543.
- Ingram, D. K. (2000). Age-related decline in physical activity: Generalization to non-humans. *Medicine and Science in Sports and Exercise*, 32, 1623-1629.
- Jack, S. J. & Ronan, K. R. (1998). Sensation seeking among high- and low-risk sports participants. *Personality and Individual Differences*, 25, 1063-1083.
- Johnson, W. & Krueger, R. F. (2004). Genetic and environmental structure of adjectives describing the domains of the Big Five Model of personality: A nationwide US twin study. *Journal of Research in Personality*, 38, 448-472.
- Kaplan, G. A., Strawbridge, W. J., Cohen, R. D., & Hungerford, L. R. (1996). Natural history of leisure-time physical activity and its correlates: Associations with mortality from all causes and cardiovascular disease over 28 years. *American Journal of Epidemiology*, 144, 793-797.
- King, A. C., Blair, S. N., Bild, D. E., Dishman, R. K., Dubbert, P. M., Marcus, B. H. et al. (1992). Determinants of physical activity and interventions in adults. *Medicine and Science in Sports and Exercise*, 24, S221-S236.
- King, A. C., Taylor, C. B., Haskell, W. L., & Debusk, R. F. (1989). Influence of regular aerobic exercise on psychological health - A randomized, controlled trial of healthy middle-aged adults. *Health Psychology*, 8, 305-324.
- Kong, X., Murphy, K., Raj, T., He, C., White, P. S., & Matise, T. C. (2004). A combined linkage-physical map of the human genome. *American Journal of Human Genetics*, 75, 1143-1148.
- Koopmans, J. R. (1997). The relation between sensation seeking, alcohol use and smoking: A multivariate genetic analysis. In *The genetics of health-related behaviors: A study of adolescent twins and their parents* (pp. 71-87). Amsterdam: Vrije Universiteit.
- Kritz-Silverstein, D., Barrett-Connor, E., & Corbeau, C. (2001). Cross-sectional and prospective study of exercise and depressed mood in the elderly - The Rancho Bernardo Study. *American Journal of Epidemiology*, 153, 596-603.
- Kujala, U. M., Kaprio, J., & Koskenvuo, M. (2002). Modifiable risk factors as predictors of all-cause mortality: The roles of genetics and childhood environment. *American Journal of Epidemiology*, 156, 985-993.
- Lander, E. & Kruglyak, L. (1995). Genetic dissection of complex traits - Guidelines for interpreting and reporting linkage results. *Nature Genetics*, 11, 241-247.
- Lauderdale, D. S., Fabsitz, R., Meyer, J. M., Sholinsky, P., Ramakrishnan, V., & Goldberg, J. (1997). Familial determinants of moderate and intense physical activity: A twin study. *Medicine and Science in Sports and Exercise*, 29, 1062-1068.

- Lawlor, D. A. & Hopker, S. W. (2001). The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *British Medical Journal*, 322, 763-767.
- Leinonen, R., Kaprio, J., Jylha, M., Tolvanen, A., Koskenvuo, M., Heikkinen, E. et al. (2005). Genetic influences underlying self-rated health in older female twins. *Journal of the American Geriatrics Society*, 53, 1002-1007.
- Lesch, K. P., Zeng, Y., Reif, A., & Gutknecht, L. (2003). Anxiety-related traits in mice with modified genes of the serotonergic pathway. *European Journal of Pharmacology*, 480, 185-204.
- Li, C., Li, M. Y., Lange, E. M., & Watanabe, R. M. (2008). Prioritized subset analysis: Improving power in genome-wide association studies. *Human Heredity*, 65, 129-141.
- Liu, Y. J., Rocha-Sanchez, S. M. S., Liu, P. Y., Long, J. R., Lu, Y., Elze, L. et al. (2004). Tests of linkage and/or association of the LEPR gene polymorphisms with obesity phenotypes in Caucasian nuclear families. *Physiological Genomics*, 17, 101-106.
- Long, B. C. & Vanstavel, R. (1995). Effects of exercise training on anxiety - A meta-analysis. *Journal of Applied Sport Psychology*, 7, 167-189.
- Loos, R. J. F., Lindgren, C. M., Li, S. X., Wheeler, E., Zhao, J. H., Prokopenko, I. et al. (2008). Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature Genetics*, 40, 768-775.
- Loos, R. J. F., Rankinen, T., Tremblay, A., Perusse, L., Chagnon, Y., & Bouchard, C. (2005). Melanocortin-4 receptor gene and physical activity in the Quebec Family Study. *International Journal of Obesity*, 29, 420-428.
- Lopez-Leon, S., Janssens, A. C. J. W., Ladd, A. M. G. Z., Del-Favero, J., Claes, S. J., Oostra, B. A. et al. (2008). Meta-analyses of genetic studies on major depressive disorder. *Molecular Psychiatry*, 13, 772-785.
- Lorentzon, M., Lorentzon, R., Lerner, U. H., & Nordstrom, P. (2001). Calcium sensing receptor gene polymorphism, circulating calcium concentrations and bone mineral density in healthy adolescent girls. *European Journal of Endocrinology*, 144, 257-261.
- Lowell, B. B. & Shulman, G. I. (2005). Mitochondrial dysfunction and type 2 diabetes. *Science*, 307, 384-387.
- MacArthur, D. & North, K. (2005). Genes and human elite athletic performance. *Human Genetics*, 116, 331-339.
- MacArthur, D. G. & North, K. N. (2007). ACTN3: A genetic influence on muscle function and athletic performance. *Exercise and Sport Sciences Reviews*, 35, 30-34.
- MacArthur, D. G., Seto, J. T., Raftery, J. M., Quinlan, K. G., Huttley, G. A., Hook, J. W. et al. (2007). Loss of ACTN3 gene function alters mouse muscle metabolism and shows evidence of positive selection in humans. *Nature Genetics*, Published online.

- Maia, J. A. R., Thomis, M., & Beunen, G. (2002). Genetic factors in physical activity levels - A twin study. *American Journal of Preventive Medicine*, 23, 87-91.
- Malina, R. M. (1996). Tracking of physical activity and physical fitness across the lifespan. *Research Quarterly for Exercise and Sport*, 67, S48-S57.
- Manolio, T. A., Rodriguez, L. L., Brooks, L., Abecasis, G., Ballinger, D., Daly, M. et al. (2007). New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nature Genetics*, 39, 1045-1051.
- Marchini, J., Howie, B., Myers, S., McVean, G., & Donnelly, P. (2007). A new multi-point method for genome-wide association studies by imputation of genotypes. *Nature Genetics*, 39, 906-913.
- Martinez-Gonzalez, M. A., Martinez, J. A., Hu, F. B., Gibney, M. J., & Kearney, J. (1999). Physical inactivity, sedentary lifestyle and obesity in the European Union. *International Journal of Obesity*, 23, 1192-1201.
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A. et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, 9, 356-369.
- Medland, S. E. (2004). Alternate parameterization for scalar and non-scalar sex-limitation models in Mx. *Twin Research*, 7, 299-305.
- Medland, S. E. (2005). Parameterization of sex-limited autosomal linkage analysis for Mx. *Twin Research and Human Genetics*, 8, 569-573.
- Meulenbelt, I., Droog, S., Trommelen, G. J. M., Boomsma, D. I., & Slagboom, P. E. (1995). High-yield noninvasive human genomic DNA isolation method for genetic-studies in geographically dispersed families and populations. *American Journal of Human Genetics*, 57, 1252-1254.
- Middeldorp, C. M., Birley, A. J., Cath, D. C., Gillespie, N. A., Willemsen, G., Statham, D. J. et al. (2005). Familial clustering of major depression and anxiety disorders in Australian and Dutch twins and siblings. *Twin Research and Human Genetics*, 8, 609-615.
- Middeldorp, C. M., Cath, D. C., van den Berg, M., Beem, A. L., Van Dyck, R., & Boomsma, D. I. (2006). The association of personality with anxious and depressive psychopathology. In T. Canli (Ed.), *The biological basis of personality and individual differences* (pp. 251-272). New York: Guilford Press.
- Middeldorp, C. M., Hottenga, J. J., Slagboom, P. E., Sullivan, P. F., De Geus, E. J. C., Posthuma, D. et al. (2008). Linkage on chromosome 14 in a genome-wide linkage study of a broad anxiety phenotype. *Molecular Psychiatry*, 13, 84-89.
- Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, 16, 1215.
- Mobily, K. E., Rubenstein, L. M., Lemke, J. H., Ohara, M. W., & Wallace, R. B. (1996). Walking and depression in a cohort of older adults: The Iowa 65+ Rural Health Study. *Journal of Aging and Physical Activity*, 4, 119-135.

- Moore, G. E., Shuldiner, A. R., Zmuda, J. M., Ferrell, R. E., Mccole, S. D., & Hagberg, J. M. (2001). Obesity gene variant and elite endurance performance. *Metabolism-Clinical and Experimental*, 50, 1391-1392.
- Moran, C. N., Vassilopoulos, C., Tsiokanos, A., Jamurtas, A. Z., Bailey, M. E. S., Montgomery, H. E. et al. (2006). The associations of ACE polymorphisms with physical, physiological and skill parameters in adolescents. *European Journal of Human Genetics*, 14, 332-339.
- Morgan, W. P. & Horstman, D. H. (1976). Anxiety reduction following acute physical activity. *Medicine and Science in Sports and Exercise*, 8, 62.
- Muthen, B. O. & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology 1995*, Vol 25, 25, 267-316.
- Muthén, L. K. & Muthén, B. O. (2007). *Mplus User's Guide*. (4 ed.) Los Angeles, CA: Muthén & Muthén.
- Neale, B., Ferreira, M. A. R., Medland, S. E., & Posthuma, D. (2007). *Statistical genetics: Gene mapping through linkage and association* (1 ed.). London: Taylor and Francis.
- Neale, M. C. (2000). QTL mapping with sib-pairs: The flexibility of Mx. In T.D.Spector, H. Snieder, & A. J. MacGregor (Eds.), *Advances in twin and sib-pair analysis* (pp. 219-243). London: Greenwich Medical Media Ltd.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical Modeling*. (6 ed.) Richmond, VA 23298: VCU Box 900126.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical Modeling*. (7 ed.) Richmond, Virginia: Department of Psychiatry, Virginia Commonwealth University.
- Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Dordrecht: Kluwer Academic Publishers.
- Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994a). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics*, 24, 239-258.
- Neale, M. C., Walters, E. E., Eaves, L. J., Maes, H. H., & Kendler, K. S. (1994b). Multivariate genetic analysis of twin-family data on fears - Mx models. *Behavior Genetics*, 24, 119-139.
- Norman, A., Bellocco, R., Vaida, F., & Wolk, A. (2002). Total physical activity in relation to age, body mass, health and other factors in a cohort of Swedish men. *International Journal of Obesity*, 26, 670-675.
- Norris, S., Carrol, D., & Cochrane, R. (1992). The effects of physical activity and exercise training on psychological stress and well-being in an adolescent population. *Journal of Psychosomatic Research*, 36, 55-65.
- North, T. C., McCullagh, P., & Tran, Z. V. (1990). Effect of exercise on depression. *Exercise and Sport Sciences Reviews*, 18, 379-415.

- Oh, K. S., Kim, E. Y., Yoon, M., & Lee, C. M. (2007). Swim training improves leptin receptor deficiency-induced obesity and lipid disorder by activating uncoupling proteins. *Experimental and Molecular Medicine*, 39, 385-394.
- Paparini, A., Ripani, M., Giordano, G. D., Santoni, D., Pigozzi, F., & Romano-Spica, V. (2007). ACTN3 genotyping by real-time PCR in the Italian population and athletes. *Medicine and Science in Sports and Exercise*, 39, 810-815.
- Park, K. S., Shin, H. D., Park, B. L., Cheong, H. S., Cho, Y. M., Lee, H. K. et al. (2006). Polymorphisms in the leptin receptor (LEPR) - putative association with obesity and T2DM. *Journal of Human Genetics*, 51, 85-91.
- Peltonen, L., Perola, M., Boomsma, D., Christensen, K., Evans, A., Harris, J. et al. (2003). GENOMEUTWIN - an integrated project of eight large twin cohorts for genetic studies on complex traits. *American Journal of Human Genetics*, 73, 397.
- Perusse, L., Gagnon, J., Province, M. A., Rao, D. C., Wilmore, J. H., Leon, A. S. et al. (2001). Familial aggregation of submaximal aerobic performance in the HERITAGE Family study. *Medicine and Science in Sports and Exercise*, 33, 597-604.
- Perusse, L., Leblanc, C., & Bouchard, C. (1988). Familial resemblance in lifestyle components - Results from the Canada Fitness Survey. *Canadian Journal of Public Health-Revue Canadienne de Sante Publique*, 79, 201-205.
- Perusse, L., Rankinen, T., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2003). The human gene map for performance and health-related fitness phenotypes: The 2002 update. *Medicine and Science in Sports and Exercise*, 35, 1248-1264.
- Perusse, L., Tremblay, A., Leblanc, C., & Bouchard, C. (1989). Genetic and environmental influences on level of habitual physical activity and exercise participation. *American Journal of Epidemiology*, 129, 1012-1022.
- Phillips, K. & Fulker, D. W. (1989). Quantitative genetic analysis of longitudinal trends in adoption designs with application to IQ in the Colorado Adoption Project. *Behavior Genetics*, 19, 621-658.
- Posthuma, D., Beem, A. L., De Geus, E. J. C., van Baal, G. C. M., von Hjelmberg, J. B., Lachine, I. et al. (2003). Theory and practice in quantitative genetics. *Twin Research*, 6, 361-376.
- Posthuma, D. & Boomsma, D. I. (2000). A note on the statistical power in extended twin designs. *Behavior Genetics*, 30, 147-158.
- Potgieter, J. R. & Bisschoff, F. (1990). Sensation seeking among medium- and low-risk sports participants. *Perceptual and Motor Skills*, 71, 1203-1206.
- Potgieter, J. R. & Venter, R. E. (1995). Relationship between adherence to exercise and scores on extraversion and neuroticism. *Perceptual and Motor Skills*, 81, 520-522.
- Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, 5, 554-571.

- Rankinen, T., Bray, M. S., Hagberg, J. M., Perusse, L., Roth, S. M., Wolfarth, B. et al. (2006a). The human gene map for performance and health-related fitness phenotypes: The 2005 update. *Medicine and Science in Sports and Exercise*, 38, 1863-1888.
- Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2001). The human gene map for performance and health-related fitness phenotypes. *Medicine and Science in Sports and Exercise*, 33, 855-867.
- Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2002). The human gene map for performance and health-related fitness phenotypes: the 2001 update. *Medicine and Science in Sports and Exercise*, 34, 1219-1233.
- Rankinen, T., Perusse, L., Rauramaa, R., Rivera, M. A., Wolfarth, B., & Bouchard, C. (2004). The human gene map for performance and health-related fitness phenotypes: The 2003 update. *Medicine and Science in Sports and Exercise*, 36, 1451-1469.
- Rankinen, T., Zuberi, A., Chagnon, Y. C., Weisnagel, S. J., Argyropoulos, G., Walts, B. et al. (2006b). The human obesity gene map: The 2005 update. *Obesity*, 14, 529-644.
- Rebollo, I., De Moor, M. H. M., Dolan, C. V., & Boomsma, D. I. (2006). Phenotypic factor analysis of family data: Correction of the bias due to dependency. *Twin Research and Human Genetics*, 9, 367-376.
- Remes, T., Vaisanen, S. B., Mahonen, A., Huuskonen, J., Kroger, H., Jurvelin, J. S. et al. (2003). Aerobic exercise and bone mineral density in middle-aged Finnish men: a controlled randomized trial with reference to androgen receptor, aromatase, and estrogen receptor alpha gene polymorphisms. *Bone*, 32, 412-420.
- Rettew, D. C., Vink, J. M., Willemsen, G., Doyle, A., Hudziak, J. J., & Boomsma, D. I. (2006). The genetic architecture of neuroticism in 3301 Dutch adolescent twins as a function of age and sex: A study from the Dutch Twin Register. *Twin Research and Human Genetics*, 9, 24-29.
- Reynolds, C. A., Barlow, T., & Pedersen, N. L. (2006). Alcohol, tobacco and caffeine use: Spouse similarity processes. *Behavior Genetics*, 36, 201-215.
- Rhodes, R. E. & Smith, N. E. I. (2006). Personality correlates of physical activity: a review and meta-analysis. *British Journal of Sports Medicine*, 40, 958-965.
- Ricquier, D. (2005). Respiration uncoupling and metabolism in the control of energy expenditure. *Proceedings of the Nutrition Society*, 64, 47-52.
- Roeder, K., Bacanu, S. A., Wasserman, L., & Devlin, B. (2006). Using linkage genome scans to improve power of association in genome scans. *American Journal of Human Genetics*, 78, 243-252.
- Rutten, A., Abel, T., Kannas, L., von Lengerke, T., Luschen, G., Diaz, J. A. R. et al. (2001). Self reported physical activity, public health, and perceived environment: results from a comparative European study. *Journal of Epidemiology and Community Health*, 55, 139-146.
- S-Plus (2001). *S-Plus 6 for Windows User's Guide*. Seattle, WA: Insightful Corporation.

References

- Sallis, J. F., Haskell, W. L., Fortmann, S. P., Vranizan, K. M., Taylor, C. B., & Solomon, D. S. (1986). Predictors of adoption and maintenance of physical activity in a community sample. *Preventive Medicine*, 15, 331-341.
- Sallis, J. F., Prochaska, J. J., & Taylor, W. C. (2000). A review of correlates of physical activity of children and adolescents. *Medicine and Science in Sports and Exercise*, 32, 963-975.
- Salmen, T., Heikkinen, A. M., Mahonen, A., Kroger, H., Komulainen, M., Pallonen, H. et al. (2003). Relation of aromatase gene polymorphism and hormone replacement therapy to serum estradiol levels, bone mineral density, and fracture risk in early postmenopausal women. *Annals of Medicine*, 35, 282-288.
- Salmon, P. (2001). Effects of physical exercise on anxiety, depression, and sensitivity to stress: A unifying theory. *Clinical Psychology Review*, 21, 33-61.
- Saxena, R., de Bakker, P. I. W., Singer, K., Mootha, V., Burt, N., Hirschhorn, J. N. et al. (2006). Comprehensive association testing of common mitochondrial DNA variation in metabolic disease. *American Journal of Human Genetics*, 79, 54-61.
- Scully, D., Kremer, J., Meade, M. M., Graham, R., & Dudgeon, K. (1998). Physical exercise and psychological well being: A critical review. *British Journal of Sports Medicine*, 32, 111-120.
- Seabra, A. F., Mendonca, D. M., Goring, H. H. H., Thomis, M. A., & Maia, J. A. (2008). Genetic and environmental factors in familial clustering in physical activity. *European Journal of Epidemiology*, 23, 205-211.
- Sham, P. (1998). *Statistics in human genetics*. London: Arnold applications of statistics.
- Shifman, S., Bhomra, A., Smiley, S., Wray, N. R., James, M. R., Martin, N. G. et al. (2008). A whole genome association study of neuroticism using DNA pooling. *Molecular Psychiatry*, 13, 302-312.
- Simonen, R., Levalahti, E., Kaprio, J., Videman, T., & Battie, M. C. (2004). Multivariate genetic analysis of lifetime exercise and environmental factors. *Medicine and Science in Sports and Exercise*, 36, 1559-1566.
- Simonen, R. L., Perusse, L., Rankinen, T., Rice, T., Rao, D. C., & Bouchard, C. (2002). Familial aggregation of physical activity levels in the Quebec family study. *Medicine and Science in Sports and Exercise*, 34, 1137-1142.
- Simonen, R. L., Rankinen, T., Perusse, L., Leon, A. S., Skinner, J. S., Wilmore, J. H. et al. (2003a). A dopamine D2 receptor gene polymorphism and physical activity in two family studies. *Physiology & Behavior*, 78, 751-757.
- Simonen, R. L., Rankinen, T., Perusse, L., Rice, T., Rao, D. C., Chagnon, Y. et al. (2003b). Genome-wide linkage scan for physical activity levels in the Quebec family study. *Medicine and Science in Sports and Exercise*, 35, 1355-1359.
- Spector, T. D., Cicuttini, F., Baker, J., Loughlin, J., & Hart, D. (1996). Genetic influences on osteoarthritis in women: A twin study. *British Medical Journal*, 312, 940-943.

- Spector, T. D. & Williams, F. M. K. (2006). The UK Adult Twin Registry (TwinsUK). *Twin Research and Human Genetics*, 9, 899-906.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *STAI Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Stefan, N., Vozarova, B., Del Parigi, A., Ossowski, V., Thompson, D. B., Hanson, R. L. et al. (2002). The Gln223Arg polymorphism of the leptin receptor in Pima Indians: influence on energy expenditure, physical activity and lipid metabolism. *International Journal of Obesity*, 26, 1629-1632.
- Stephens, T. (1988). Physical activity and mental health in the United States and Canada - Evidence from 4 population surveys. *Preventive Medicine*, 17, 35-47.
- Step toe, A. (2006). Depression and physical activity. In A. Steptoe (Ed.), *Depression and physical illness* (pp. 348-368). Cambridge University Press.
- Step toe, A. & Butler, N. (1996). Sports participation and emotional wellbeing in adolescents. *Lancet*, 347, 1789-1792.
- Step toe, A., Edwards, S., Moses, J., & Mathews, A. (1989). The effects of exercise training on mood and perceived coping ability in anxious adults from the general population. *Journal of Psychosomatic Research*, 33, 537-547.
- Step toe, A., Wardle, J., Fuller, R., Holte, A., Justo, J., Sanderman, R. et al. (1997). Leisure-time physical exercise: prevalence, attitudinal correlates, and behavioral correlates among young Europeans from 21 countries. *Preventive Medicine*, 26, 845-854.
- Stoel, R. D., De Geus, E. J. C., & Boomsma, D. I. (2006). Genetic analysis of sensation seeking with an extended twin design. *Behavior Genetics*, 36, 229-237.
- Strawbridge, W. J., Deleger, S., Roberts, R. E., & Kaplan, G. A. (2002). Physical activity reduces the risk of subsequent depression for older adults. *American Journal of Epidemiology*, 156, 328-334.
- Stubbe, J. H., Boomsma, D. I., & De Geus, E. J. C. (2005). Sports participation during adolescence: A shift from environmental to genetic factors. *Medicine and Science in Sports and Exercise*, 37, 563-570.
- Stubbe, J. H., Boomsma, D. I., Vink, J. M., Cornes, B. K., Martin, N. G., Skytthe, A. et al. (2006a). Genetic influences on exercise participation: A comparative study in adult twin samples from seven countries. *PLoS ONE*, 1, e22. doi:10.1371/journal.pone.0000022-<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0000022>.
- Stubbe, J. H. & De Geus, E. J. (2006b). Genetics of exercise behavior. In *Handbook of behavior genetics*, in press.
- Stubbe, J. H., De Moor, M. H. M., Boomsma, D., & De Geus, E. J. (2007). The association between exercise participation and well-being: A co-twin study. *Preventive Medicine*, 44, 148-152.

References

- Sullivan, P. F., De Geus, E. J. C., Willemsen, G., James, M. R., Smit, J. H., Zandbelt, T. et al. (2008). Genomewide association for major depressive disorder: a possible role for the presynaptic protein piccolo, *Molecular Psychiatry*, in press.
- Sullivan, P. F., Neale, M. C., & Kendler, K. S. (2000). Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry*, 157, 1552-1562.
- Svedberg, P., Gatz, M., Lichtenstein, P., Sandin, S., & Pedersen, N. L. (2005). Self-rated health in a longitudinal perspective: A 9-year follow-up twin study. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 60, S331-S340.
- Svedberg, P., Lichtenstein, P., & Pedersen, N. L. (2001). Age and sex differences in genetic and environmental factors for self-rated health: A twin study. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, 56, S171-S178.
- Teychenne, M., Ball, K., & Salmon, J. (2008). Physical activity and likelihood of depression in adults: A review. *Preventive Medicine*, 46, 397-411.
- Thomis, M. A. I., Beunen, G. P., Maes, H. H., Blimkie, C. J., Van Leemputte, M., Claessens, A. L. et al. (1998). Strength training: importance of genetic factors. *Medicine and Science in Sports and Exercise*, 30, 724-731.
- Todd, J. A. & Robinson, R. J. (2003). Osteoporosis and exercise. *Postgraduate Medical Journal*, 79, 320-323.
- Tremblay, L. K., Naranjo, C. A., Cardenas, L., Herrmann, N., & Busto, U. E. (2002). Probing brain reward system function in major depressive disorder - Altered response to dextroamphetamine. *Archives of General Psychiatry*, 59, 409-416.
- Trevaskis, J., Walder, K., Foletta, V., Kerr-Bayles, L., McMillan, J., Cooper, A. et al. (2005). Src homology 3-domain growth factor receptor-bound 2-like (endophilin) interacting protein 1, a novel neuronal protein that regulates energy balance. *Endocrinology*, 146, 3757-3764.
- Twisk, J. W. R., Kemper, H. C. G., & van Mechelen, W. (2000). Tracking of activity and fitness and the relationship with cardiovascular disease risk factors. *Medicine and Science in Sports and Exercise*, 32, 1455-1461.
- Twoogor, S. S., Chubak, J., Aiello, E. J., Yasui, Y., Ulrich, C. M., Farin, F. M. et al. (2004). The effect of CYP19 and COMT polymorphisms on exercise-induced fat loss in postmenopausal women. *Obesity Research*, 12, 972-981.
- Ul Haque, M. F., King, L. M., Krakow, D., Cantor, R. M., Rusiniak, M. E., Swank, R. T. et al. (1998). Mutations in orthologous genes in human spondyloepimetaphyseal dysplasia and the brachymorphic mouse. *Nature Genetics*, 20, 157-162.
- Van der Gils, V. W. V., Hildebrandt, V. H., & Otten, F. (2002). Sporten, Wandelen en Fietsen in Nederland. In W.T.M.Ooijendijk, V. H. Hildebrandt, & M. Stiggelbout (Eds.), *Trendrapport Bewegen en Gezondheid 2001/2002* (pp. 39-62). Hoofddorp: TNO Arbeid.
- Van der Meulen, R. (2003). Beoefening. In K.Breedveld (Ed.), *Rapportage Sport 2003* (pp. 71-97). Den Haag: Sociaal Cultureel Planbureau.

- Van der Sluis, S., Dolan, C. V., Neale, M. C., & Posthuma, D. (2008a). A general test for gene-environment interaction in sib pair-based association analysis of quantitative traits. *Behavior Genetics*, 38, 372-389.
- Van der Sluis, S., Dolan, C. V., Neale, M. C., & Posthuma, D. (2008b). Power calculations using exact data simulation: A useful tool for genetic study designs. *Behavior Genetics*, 38, 202-211.
- Van Gool, C. H., Kempen, G. I. J. M., Penninx, B. W. J. H., Deeg, D. J. H., Beekman, A. T. F., & Van Eijk, J. T. M. (2003). Relationship between changes in depressive symptoms and unhealthy lifestyles in late middle aged and older persons: results from the Longitudinal Aging Study Amsterdam. *Age and Ageing*, 32, 81-87.
- Van Leeuwen, M., Van den Berg, S. M., & Boomsma, D. I. (2008). A twin-family study of general IQ. *Learning and Individual Differences*, 18, 76-88.
- Van Mechelen, W., Twisk, J. W. R., Post, G. B., Snel, J., & Kemper, H. C. G. (2000). Physical activity of young people: the Amsterdam Longitudinal Growth and Health Study. *Medicine and Science in Sports and Exercise*, 32, 1610-1616.
- Viken, R. J., Rose, R. J., Kaprio, J., & Koskenvuo, M. (1994). A developmental genetic analysis of adult personality: Extraversion and neuroticism from 18 to 59 years of age. *Journal of Personality and Social Psychology*, 66, 722-730.
- Vink, J. M., Willemsen, G., Engels, R. C. M. E., & Boomsma, D. I. (2003). Smoking status of parents, siblings and friends: Predictors of regular smoking? Findings from a longitudinal twin-family study. *Twin Research*, 6, 209-217.
- Vink, J. M., Willemsen, G., Stubbe, J. H., Middeldorp, C. M., Ligthart, R. S. L., Baas, K. D. et al. (2004). Estimating non-response bias in family studies: Application to mental health and lifestyle. *European Journal of Epidemiology*, 19, 623-630.
- Wang, K., Li, M. Y., & Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. *American Journal of Human Genetics*, 81, 1278-1283.
- Weedon, M. N., Lango, H., Lindgren, C. M., Wallace, C., Evans, D. M., Mangino, M. et al. (2008). Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, 40, 575-583.
- Weyerer, S. (1992). Physical inactivity and depression in the community. Evidence from the Upper Bavarian Field Study. *International Journal of Sports Medicine*, 13, 492-496.
- Wilde, G. J. S. (1970). *Neurotische Labiliteit Gemeten Volgens de Vragenlijstmethode (The Questionnaire Method as a Means of Measuring Neurotic Instability)*. Amsterdam: Van Rossen.
- Willemsen, G., Posthuma, D., & Boomsma, D. I. (2005). Environmental factors determine where the Dutch live: Results from the Netherlands Twin Register. *Twin Research and Human Genetics*, 8, 312-317.

References

- Williams, A. G., Dhamrait, S. S., Wootton, P. T. E., Day, S. H., Hawe, E., Payne, J. R. et al. (2004). Bradykinin receptor gene variant and human physical performance. *Journal of Applied Physiology*, 96, 938-942.
- Williams, A. G., Rayson, M. P., Jubb, M., World, M., Woods, D. R., Hayward, M. et al. (2000). Physiology - The ACE gene and muscle performance. *Nature*, 403, 614.
- Wilmore, J. H. & Costill, D. L. (2004). *Physiology of sport and exercise*. Champaign (IL): Human Kinetics.
- Wilson, S. G., Reed, P. W., Bansal, A., Chiano, M., Lindersson, M., Langdown, M. et al. (2003). Comparison of genome screens for two independent cohorts provides replication of suggestive linkage of bone mineral density to 3p21 and 1p36. *American Journal of Human Genetics*, 72, 144-155.
- Winnicki, M., Accurso, V., Hoffmann, M., Pawlowski, R., Dorigatti, F., Santonastaso, M. et al. (2004). Physical activity and angiotensin-converting enzyme gene polymorphism in mild hypertensives. *American Journal of Medical Genetics*, 125A, 38-44.
- Wise, L. A., Adams-Campbell, L. L., Palmer, J. R., & Rosenberg, L. (2006). Leisure time physical activity in relation to depressive symptoms in the Black Women's Health Study. *Annals of Behavioral Medicine*, 32, 68-76.
- Wolfarth, B., Bray, M. S., Hagberg, J. M., Perusse, L., Rauramaa, R., Rivera, M. A. et al. (2005). The human gene map for performance and health-related fitness phenotypes: The 2004 update. *Medicine and Science in Sports and Exercise*, 37, 881-903.
- Wolfarth, B., Rivera, M. A., Oppert, J. M., Boulay, M. R., Dionne, F. T., Chagnon, M. et al. (2000). A polymorphism in the alpha(2a)-adrenoceptor gene and endurance athlete status. *Medicine and Science in Sports and Exercise*, 32, 1709-1712.
- Woods, D. R., Humphries, S. E., & Montgomery, H. E. (2000). The ACE I/D polymorphism and human physical performance. *Trends in Endocrinology and Metabolism*, 11, 416-420.
- Yang, N., MacArthur, D. G., Gulbin, J. P., Hahn, A. G., Beggs, A. H., Eastal, S. et al. (2003). ACTN3 genotype is associated with human elite athletic performance. *American Journal of Human Genetics*, 73, 627-631.
- Zeggini, E., Weedon, M. N., Lindgren, C. M., Frayling, T. M., Elliott, K. S., Lango, H. et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316, 1336-1341.
- Zubenko, G. S., Maher, B., Hughes, H. B., Zubenko, W. N., Stiffler, J. S., Kaplan, B. B. et al. (2003). Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 123B, 1-18.
- Zuckerman, M. (1971). Dimensions of Sensation Seeking. *Journal of Consulting and Clinical Psychology*, 36, 45.





Appendix A

Overview of all exercise-related questions that have been included in the surveys from 1991 to 2004



Survey 1991

Probeer de volgende vragen zo nauwkeurig mogelijk te beantwoorden voor de afgelopen 3 maanden. Als je nu een bijzondere periode doormaakt waardoor je leven heel anders is dan normaal, beantwoord de vragen dan voor de 3 maanden die vlak voor deze periode liggen.

1. Doe je aan sport? (Bedoeld wordt bijvoorbeeld zwemmen, voetballen, maar ook dansen, yoga, ballet, conditietraining)
 - nee > ga door met vraag 10
 - ja
2. Ben je lid van een sportvereniging?
 - nee > ga door met vraag 9!
 - ja

Bij de volgende vragen gaat het om het soort sport en niet om de naam van de sportclub. Als je lid bent van meerdere clubs vermeld dan de twee belangrijkste sporten.

	Eerst sport	Tweede sport
3. Welke sport(en) beoefen je in clubverband?	_____	_____
4. Hoe lang beoefen je deze sport al?	___ jaar	___ jaar
5. Hoe vaak per week train je?	___ keer per week	___ keer per week
6. Hoe lang train je per keer?	___ minuten	___ minuten
7. Hoe vaak speel je een wedstrijd?	___ keer per week	___ keer per week
8. Hoe lang duurt een wedstrijd?	___ minuten	___ minuten

9. Welke sport(en) beoefen je niet in clubverband? (b.v. fietsen, wandelen)

	Hoe vaak en hoe lang?
a) _____	_____
b) _____	_____

10. Heb je andere regelmatig terugkerende activiteiten die lichamelijke beweging vereisen? (b.v. vrijwilligerswerk, lid van een vereniging)
 - nee
 - ja

11. Zo ja, wat voor soort activiteiten, en hoe vaak?

12. Vind je jezelf in vergelijking met anderen
 - minder actief
 - ongeveer even actief
 - meer actief

Appendix A

13. Als je cijfers mag geven van 1 tot 10 die aangeven hoe actief je mensen vindt, wat voor cijfer geef je dan aan: (hoe hoger het cijfer hoe actiever) [voor tweelingen]

- a) jezelf _____
- b) je tweelingbroer/zus _____
- c) je vader _____
- d) je moeder _____

14. Als u cijfers mag geven van 1 tot 10 die aangeven hoe actief u de volgende mensen vindt, wat voor cijfer geeft U dan aan: (hoe hoger het cijfer hoe actiever) [voor ouders]

- a) uzelf _____
- b) vader van de tweeling _____
- c) oudste van de tweeling _____
- d) jongste van de tweeling _____

De volgende vragen gaan over speciale talenten die je mogelijk hebt. De eerste antwoord mogelijkheid beschrijft een uitzonderlijk talent. De derde mogelijkheid beschrijft een gemiddelde, niet goed en niet slecht. Er zijn maar weinig mensen met uitzonderlijke talenten. De meeste mensen zullen de derde of vierde mogelijkheid omcirkelen. Maar de mensen die wel talentvol zijn, kunnen dat hier aangeven. Kies steeds maar één mogelijkheid.

15. Sport

- 1) Je bent atletisch gevormd en zeer vaardig in één of meer sporten. Geef aan:

- 2) Je bent beter dan gemiddeld in één of meer sporten.
- 3) Je speelt voor je plezier één of meer sporten zonder uitzonderlijke prestaties.
- 4) Je speelt geen sport en hebt er geen talent voor.

Appendix A

6. Hoe is je lichamelijke uithoudingsvermogen?

- 1) slecht
- 2) matig
- 3) redelijk
- 4) goed
- 5) uitstekend

7. Hoeveel van je vrienden/vriendinnen doen regelmatig aan sport? [voor tweelingen]

- 1) niemand
- 2) een paar
- 3) ongeveer de helft
- 4) de meesten
- 5) allemaal

Appendix A

6. Hoeveel van je vrienden/vriendinnen doen regelmatig aan sport?

[voor tweelingen en broers/zussen]

- 1) niemand
- 2) een paar
- 3) ongeveer de helft
- 4) de meesten
- 5) allemaal

6. Hoe is u lichamelijke uithoudingsvermogen? [voor ouders]

- 1) slecht
- 2) matig
- 3) redelijk
- 4) goed
- 5) uitstekend

Survey 1997

1. Doet u regelmatig aan sport?

- 1) nee
- 2) ja, namelijk 1) _____ (soort sport)
- 2) _____ (soort sport)
- 3) _____ (soort sport)

2. Als u terugdenkt aan de afgelopen 6 maanden, hoe vaak bent u in uw vrije tijd tenminste 20 minuten per keer lichamelijk actief geweest? (bv fietsen, wandelen, zwemmen, joggen, dansen, tuinieren).

- 1) nooit
- 2) minder dan één keer per maand
- 3) ongeveer één keer per maand
- 4) ongeveer 2-3 keer per maand
- 5) 1-2 keer per week
- 6) 3 keer per week of vaker

3. Bent u tenminste één keer per week in uw vrije tijd zo lichamelijk actief dat u ervan gaat zweten?

- 1) nee
- 2) ja, namelijk 1) één keer per week
- 2) twee keer per week
- 3) drie keer per week
- 4) vier keer per week

Survey 2002

Sport

1. Doet u regelmatig aan sport?

- 1) nee > door naar effecten van sporten (vraag 3)
 2) ja

2. Geef hieronder aan welke sport(en) u beoefent. Geef aan hoeveel jaren, hoeveel maanden per jaar, hoe vaak per maand en hoelang u gemiddeld per week deze sport(en) beoefent. Tel de tijd van de trainingen en wedstrijden bij elkaar op.

naam van de sport	aantal jaren	aantal maanden per jaar	aantal keren per maand	gemiddelde tijd per week
1. _____	___ jaren	___ maanden	___ keer	___ minuten
2. _____	___ jaren	___ maanden	___ keer	___ minuten
3. _____	___ jaren	___ maanden	___ keer	___ minuten
4. _____	___ jaren	___ maanden	___ keer	___ minuten
5. _____	___ jaren	___ maanden	___ keer	___ minuten

Effecten van sporten

3. De onderstaande stellingen gaan over mogelijke effecten van regelmatig sporten.

Wilt u aangeven of u het eens of oneens bent met deze stellingen?

	absoluut mee oneens	mee oneens	mee eens	absoluut mee eens
a) Door te sporten ziet je lichaam er beter uit.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
b) Je voelt je beter als je regelmatig sport.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
c) Sporten geeft je meer energie.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
d) Sporten geeft je een gevoel dat je iets bereikt.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
e) Sporten houdt de geest actief.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
f) Sporten is goed voor je hart.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
g) Sporten is goed voor je gemoedstoestand.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
h) Mensen sporten om gezond te blijven.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
i) Door te sporten voel je je minder gestresst en gespannen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4
j) Door te sporten kom je in contact met andere mensen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4

Lichamelijke activiteit

4. Neem een normale week in de afgelopen maand in uw gedachten. Wilt u aangeven hoeveel dagen per week u de onderstaande activiteiten verrichtte en hoelang u daar dan gemiddeld op zo'n dag mee bezig was? Indien u een bepaalde activiteit niet heeft gedaan kunt u deze activiteit overslaan.

- | | aantal dagen per week | gemiddelde tijd per dag |
|---|-----------------------|-------------------------|
| I. Fietsen | | |
| a) doordeeweeks
(maandag tot en met vrijdag) | ___ dagen | ___ minuten |
| b) in het weekend | ___ dagen | ___ minuten |
| II. Wandelen | | |
| a) doordeeweeks
(maandag tot en met vrijdag) | ___ dagen | ___ minuten |
| b) in het weekend | ___ dagen | ___ minuten |
| III. Lichamelijke activiteit in het huishouden | | |
| a) licht en matig inspannend huishoudelijk
werk (staand werk zoals koken, afwassen,
strijken, kind in bad doen en/of lopend
werk zoals stofzuigen) | ___ dagen | ___ minuten |
| b) zwaar inspannend huishoudelijk werk
(vloer schrobben en met zware
boodschappen lopen) | ___ dagen | ___ minuten |
| IV. Lichamelijke activiteit op werk en/of school | | |
| a) matig inspannend werk (staand werk met
af en toe lopen, zoals baliewerk,
kapper en schilder) | ___ dagen | ___ minuten |
| b) zwaar inspannend werk (werk waarbij
regelmatig zware dingen worden
opgetild, zoals verhuizer en stukadoor) | ___ dagen | ___ minuten |
| V. Overige lichamelijke activiteiten | | |
| a) tuinieren | ___ dagen | ___ minuten |
| b) klussen/doe-het-zelven | ___ dagen | ___ minuten |
| c) dansen | ___ dagen | ___ minuten |
| 5. Bent u tenminste één keer per week in uw vrije tijd zo lichamenlijk actief dat u ervan gaat zweten? | | |
| <input type="checkbox"/> 1) nee | | |
| <input type="checkbox"/> 2) ja, namelijk | | |
| <input type="checkbox"/> 1) één keer per week | | |
| <input type="checkbox"/> 2) twee keer per week | | |
| <input type="checkbox"/> 3) drie keer per week | | |
| <input type="checkbox"/> 4) meer dan drie keer per week | | |
| 6. Als u terugdenkt aan de afgelopen 6 maanden, hoe vaak bent u in uw vrije tijd tenminste 20 minuten per keer lichamenlijk actief geweest? (bv fietsen, wandelen, zwemmen, joggen, dansen, tuinieren). | | |
| <input type="checkbox"/> 1) nooit | | |
| <input type="checkbox"/> 2) minder dan één keer per maand | | |
| <input type="checkbox"/> 3) ongeveer één keer per maand | | |
| <input type="checkbox"/> 4) ongeveer 2-3 keer per maand | | |
| <input type="checkbox"/> 5) 1-2 keer per week | | |
| <input type="checkbox"/> 6) 3 keer per week of vaker | | |

Belemmeringen voor lichamelijke activiteit

7. Hoe vaak wordt u door het volgende gehinderd om lichamenlijk actief te worden of te gaan sporten?

	nooit	zelden	af en toe	vaak	heel vaak
a) Ik ben onzeker over mijn uiterlijk als ik actief ben.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
b) Ik heb geen interesse in lichamelijke activiteit.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
c) Ik heb geen zelfdiscipline of wilskracht.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
d) Ik heb er geen tijd voor.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
e) Ik heb er de energie niet voor.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
f) Ik heb niemand om samen mee te sporten.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
g) Ik beleef geen plezier aan sport of lichamelijke activiteit.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
h) Ik wil niet falen, dus ik probeer het niet.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
i) Ik heb niet de vereiste sportbenodigdheden.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
j) Ik vind het weer vaak te slecht.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
k) Ik heb te weinig sportieve vaardigheden.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
l) Ik ben te moe om te sporten.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
m) Ik heb te weinig kennis over hoe ik moet sporten.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
n) Ik heb een slechte gezondheid.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
o) Ik ben bang voor blessures.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
p) Ik vind bewegen zwaar.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
q) Ik heb geen goed bereikbare sportfaciliteiten in de buurt.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
r) Ik ben te dik.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
s) Ik vind sporten saai.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
t) Ik heb werkverplichtingen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
u) Ik heb sociale verplichtingen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
v) Ik heb familieverplichtingen.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
w) Ik vind sporten te duur.	<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5

Survey 2004

1. Doet u regelmatig aan sport?

1) nee > door naar vraag 3

2) ja

2. Wilt u hieronder invullen welke sport(en) u beoefent? Geef per sport aan hoeveel jaren u deze al beoefent, hoeveel keer per week u de sport beoefent en hoelang u gemiddeld per keer deze sport beoefent.

naam van de sport	aantal jaren	aantal maanden per jaar	aantal keren per maand	gemiddelde tijd per week
1. _____	___ jaren	___ maanden	___ keer	___ minuten
2. _____	___ jaren	___ maanden	___ keer	___ minuten
3. _____	___ jaren	___ maanden	___ keer	___ minuten
4. _____	___ jaren	___ maanden	___ keer	___ minuten

3. Bent u tenminste één keer per week in uw vrije tijd zo lichamelijk actief dat u ervan gaat zweten?

1) nee

2) ja, namelijk

1) één keer per week

2) twee keer per week

3) drie keer per week

4) meer dan drie keer per week

4. Als u terugdenkt aan de afgelopen 6 maanden, hoe vaak bent u in uw vrije tijd tenminste 20 minuten per keer lichamelijk actief geweest? (bv fietsen, zwemmen, joggen, dansen, tuinieren).

1) nooit

2) minder dan één keer per maand

3) ongeveer één keer per maand

4) ongeveer 2-3 keer per maand

5) 1-2 keer per week

6) 3 keer per week of vaker





Appendix B

Mx script for parent-offspring model
with sex and generation differences
in variance decomposition



```

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
! SCRIPT NAME      :
! Gen_par_off_sp_adolescents_sexdiff_ACE_AE_assmat_Mxbib.mx
! GOAL: To model parent-offspring data with assortative mating and cultural
! transmission allowing for sex and generation differences in variance
! decomposition
! DATA            : binary
! INPUT            : raw data
! UNI/BI/MULTI     : uni
! DATA-GROUPS     : MZM DZM MZF DZF DOS, including parents and
! siblings
! MEANS MODEL      : different thresholds for fathers, mothers, sons and
! daughters, same age effect on each threshold
! VARIANCE COVARIANCE MODEL(S)    ## 1.ACE 2.AE 3.CE 4.E

! Downloading Mx software: http://www.vcu.edu/mx
! Mx script's library: http://www.psy.vu.nl/mxbib
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! Phenotypic Assortment Model based on factor model described by Neale
! BG (1994), with cultural transmission
! The model assumes a fixed AE model for the parental generation (adults)
! and an ACE model for the offspring generation (adolescents)
! Quantitative sex differences in both generations
! Qualitative sex differences in offspring modelled in C (same environment?)
! Quantitative differences between generations
! Qualitative differences between generations modelled in A (same genes?)

! MHM de Moor, Biological Psychology, VU Amsterdam
! (MHM.de.Moor@psy.vu.nl)
! Univariate, binary data
! Script for the analysis of two parents, twin offspring, plus a
! maximum of two additional male and two female siblings

! Note: genetic correlation across generations is not identified when
! cultural transmission is estimated

#define nvar 1           ! one phenotype per subject
#define nthres 1        ! number of thresholds
#define nind 8          ! number of individuals per family

```

!!!!!!

!!MODEL

!!!!!!

Group 1: Mother-Father covariance

Data Calculation NGroups=31

Matrices

D Full nvar nvar free ! assortative mating delta paths

P Symm nvar nvar fix ! within person covariance

! FATHER=MOTHER=SON=DAUGHTER (Rp), Note: can be modified if necessary

Compute P*D*P /

Start 1.0 P 1 1

start 0.4 D 1 1

Option No_output

End

Group 2: Calculate T Genotype-Phenotype covariance SON

Data Calculation

Matrices

C Low nvar nvar free ! common environment paths

G Symm nvar nvar free ! additive genetic covariance (Ra)

A Low nvar nvar free ! additive genetic paths

S Full nvar nvar free ! A-C covariance

Compute G*A' + S*C' /

Start 1 G 1 1

Start .62 A 1 1

Start .63 C 1 1

Option No_output

End

Group 3: Calculate T Genotype-Phenotype covariance DAUGHTER

Data Calculation

Matrices

Y Low nvar nvar free ! common environment paths

I Symm nvar nvar = G(2) ! additive genetic cov (Ra) SON=DAUGHTER

X Low nvar nvar free ! additive genetic paths

T Full nvar nvar free ! A-C covariance

Compute I*X' + T*Y' /

Start .62 X 1 1

Start .71 Y 1 1

```
Option No_output
End
```

Group 4: Father-Son Covariance

Data Calculation

Matrices

```
D Full nvar nvar = D(1)           ! assortative mating delta paths
P Symm nvar nvar = P(1)          ! within person covariance (Rp)
W Full nvar nvar = %E1           ! spouse covariance
U Low nvar nvar fix              ! additive genetic paths FATHER
V Low nvar nvar fix              ! additive genetic paths MOTHER
F Full nvar nvar free            ! father-son cultural transmission
M Full nvar nvar free            ! mother-son cultural transmission
Z Full nvar nvar fix             ! genetic correlation between generations
A Low nvar nvar = A(2)           ! additive genetic paths SON
C Low nvar nvar = C(2)           ! common environment paths SON
H Full 1 1                        ! scalar .5
I Iden nvar nvar                 ! identity matrix
Compute (P*F' + W*M')*C' + I*U*((H*Z)@A')+(P*D')*V*((H*Z)@A') /
!fix paths parents to values obtained from adult twin analysis
Start 0.6 U 1 1
Start 0.6 V 1 1
Start 0.2 F 1 1
Start -0.2 M 1 1
Start 1 Z 1 1                     !fix if cultural transm is estimated
Matrix H 0.5
Bound 0 1 Z 1 1
Option No_output
End
```

Group 5: Mother-Son covariance

Data Calculation

Matrices

```
D Full nvar nvar = D(1)           ! assortative mating delta paths
P Symm nvar nvar = P(1)          ! within person covariance (Rp)
W Full nvar nvar = %E1           ! spouse covariance
U Low nvar nvar = U(4)          ! additive genetic paths FATHER
V Low nvar nvar = V(4)          ! additive genetic paths MOTHER
F Full nvar nvar = F(4)          ! father-son cultural transmission
M Full nvar nvar = M(4)          ! mother-son cultural transmission
Z Full nvar nvar = Z(4)          ! genetic correlation between generations
```

```

A Low nvar nvar = A(2)      ! additive genetic paths SON
C Low nvar nvar = C(2)      ! common environment paths SON
H Full 1 1 = H(4)           ! scalar .5
I Iden nvar nvar           ! identity matrix
Compute (P*M' + W*F')*C' + I*V*((H*Z)@A')+(P*D')*U*((H*Z)@A') /
Option No_output
End

```

Group 6: Father-Daughter Covariance

Data Calculation

Matrices

```

D Full nvar nvar = D(1)      ! assortative mating delta paths
P Symm nvar nvar = P(1)      ! within person covariance (Rp)
W Full nvar nvar = %E1       ! spouse covariance
U Low nvar nvar = U(4)       ! additive genetic paths FATHER
V Low nvar nvar = V(4)       ! additive genetic paths MOTHER
E Full nvar nvar free        ! father-daughter cultural transmission
L Full nvar nvar free        ! mother-daughter cultural transmission
Z Full nvar nvar = Z(4)      ! genetic correlation between generations
X Low nvar nvar = X(3)       ! additive genetic paths DAUGHTER
Y Low nvar nvar = Y(3)       ! common environment paths DAUGHTER
H Full 1 1 = H(4)           ! scalar .5
I Iden nvar nvar           ! identity matrix
Compute (P*E' + W*L')*Y' + I*U*((H*Z)@X')+(P*D')*V*((H*Z)@X') /
Start -0.01 E 1 1 L 1 1
Option No_output
End

```

Group 7: Mother-Daughter Covariance

Data Calculation

Matrices

```

D Full nvar nvar = D(1)      ! assortative mating delta paths
P Symm nvar nvar = P(1)      ! within person covariance (Rp)
W Full nvar nvar = %E1       ! spouse covariance
U Low nvar nvar = U(4)       ! additive genetic paths FATHER
V Low nvar nvar = V(4)       ! additive genetic paths MOTHER
E Full nvar nvar = E(6)      ! father-daughter cultural transmission
L Full nvar nvar = L(6)      ! mother-daughter cultural transmission
Z Full nvar nvar = Z(4)      ! genetic correlation between generations
X Low nvar nvar = X(3)       ! additive genetic paths DAUGHTER

```

```

Y Low nvar nvar = Y(3)      ! common environment paths DAUGHTER
H Full 1 1 = H(4)          ! scalar .5
I Iden nvar nvar          ! identity matrix
Compute (P*L' + W'*E')*Y' + I*V'*((H*Z)@X')+(P*D')*U'*((H*Z)@X') /
Option No_output
End

```

Group 8: MZM Twin Covariance

Data Calculation

Matrices

```

R Symm nvar nvar free      ! common environment covariance SON
C Low nvar nvar = C(2)     ! common environment paths SON
G Symm nvar nvar = G(2)    ! additive genetic covariance SON
A Low nvar nvar = A(2)     ! additive genetic paths SON
S Full nvar nvar = S(2)    ! A-C covariance SON
N Low nvar nvar fix        ! non-additive paths SON
Compute A*G*A' + C*R*C' + A*S*C' + C*S*A' + N*N' /
Start 1.07 R 1 1
!Start 0.4 N 1 1
Option No_output
End

```

Group 9: DZM Twin Covariance

Data Calculation

Matrices

```

R Symm nvar nvar = R(8)    ! common environment covariance SON
D Full nvar nvar = D(1)    ! assortative mating delta paths
I Ident nvar nvar          ! matrix with 1s for variance of A in parents
Z Full nvar nvar = Z(4)    ! genetic correlation between generations
C Low nvar nvar = C(2)     ! common environment paths SON
H Full 1 1 = H(4)         ! scalar, .5
A Low nvar nvar = A(2)     ! additive genetic paths SON
S Full nvar nvar = S(2)    ! A-C covariance SON
U Low nvar nvar = U(4)     ! additive genetic paths FATHER
V Low nvar nvar = V(4)     ! additive genetic paths MOTHER
N Low nvar nvar = N(8)     ! non-additive paths SON
Compute (H*Z*Z)@A*(I+H@(U*(D'+D)*V'))*A' + C*R*C' + A*S*C' +
C*S*A' + H@H@N*N'/
Option No_output
End

```

Group 10: sibMM Covariance

Data Calculation

Matrices

```

R Symm nvar nvar = R(8)      ! common environment covariance SON
D Full nvar nvar = D(1)      ! assortative mating delta paths
I Ident nvar nvar            ! matrix with 1s for variance of A in parents
Z Full nvar nvar = Z(4)      ! genetic correlation between generations
C Low nvar nvar = C(2)       ! common environment paths SON
H Full 1 1 = H(4)           ! scalar, .5
A Low nvar nvar = A(2)       ! additive genetic paths SON
S Full nvar nvar = S(2)      ! A-C covariance SON
U Low nvar nvar = U(4)       ! additive genetic paths FATHER
V Low nvar nvar = V(4)       ! additive genetic paths MOTHER
N Low nvar nvar = N(8)       ! non-additive paths SON
Compute (H*Z*Z)@A*(I+H@(U*(D'+D)*V'))*A' + C*R*C' + A*S*C' +
C*S*A' + H@H@N*N' /
Option No_output
End

```

Group 11: MZF Twin Covariance

Data Calculation

Matrices

```

Q Symm nvar nvar free        ! common environment covariance DAUGHTER
Y Low nvar nvar = Y(3)       ! common environment paths DAUGHTER
I Symm nvar nvar = I(3)      ! additive genetic covariance DAUGHTER
X Low nvar nvar = X(3)       ! additive genetic paths DAUGHTER
T Full nvar nvar = T(3)      ! A-C covariance DAUGHTER
B Low nvar nvar fix          ! non-additive paths DAUGHTER
Compute X*I*X' + Y*Q*Y' + X*T*Y' + Y*T*X' + B*B' /
Start 1.03 Q 1 1
!Start 0.4 B 1 1
Option No_output
End

```

Group 12: DZF Twin Covariance

Data Calculation

Matrices

```

Q Symm nvar nvar = Q(11)     ! common environment covariance DAUGHTER
D Full nvar nvar = D(1)      ! assortative mating delta paths
Z Full nvar nvar = Z(4)      ! genetic correlation between generations

```

```

Y Low nvar nvar = Y(3)      ! common environment paths DAUGHTER
I Ident nvar nvar          ! matrix with 1s for var of A in parents
H Full 1 1 = H(4)         ! scalar, .5
X Low nvar nvar = X(3)    ! additive genetic paths DAUGHTER
T Full nvar nvar = T(3)   ! A-C covariance DAUGHTER
U Low nvar nvar = U(4)    ! additive genetic paths FATHER
V Low nvar nvar = V(4)    ! additive genetic paths MOTHER
B Low nvar nvar = B(11)   ! non-additive paths DAUGHTER
Compute (H*Z*Z)@X*(I+H@(V*(D'+D)*U'))*X' + Y*Q*Y' + X*T*Y' +
Y*T'*X' + H@H@B*B'/
Option No_output
End

```

Group 13: sibFF Covariance

Data Calculation

Matrices

```

Q Symm nvar nvar = Q(11)   ! common environment covariance DAUGHTER
D Full nvar nvar = D(1)    ! assortative mating delta paths
Z Full nvar nvar = Z(4)    ! genetic correlation between generations
Y Low nvar nvar = Y(3)    ! common environment paths DAUGHTER
I Ident nvar nvar          ! matrix with 1s for var of A in parents
H Full 1 1 = H(4)         ! scalar, .5
X Low nvar nvar = X(3)    ! additive genetic paths DAUGHTER
T Full nvar nvar = T(3)   ! A-C covariance DAUGHTER
U Low nvar nvar = U(4)    ! additive genetic paths FATHER
V Low nvar nvar = V(4)    ! additive genetic paths MOTHER
B Low nvar nvar = B(11)   ! non-additive paths DAUGHTER
Compute (H*Z*Z)@X*(I+H@(V*(D'+D)*U'))*X' + Y*Q*Y' + X*T*Y' +
Y*T'*X' + H@H@B*B'/
Option No_output
End

```

Group 14: DOS Twin Covariance

Data Calculation

Matrices

```

O Symm nvar nvar free      ! corr between res var C males + females
P Symm nvar nvar = P(1)   ! within person covariance (Rp)
D Full nvar nvar = D(1)   ! assortative mating delta paths
U Low nvar nvar = U(4)    ! additive genetic paths FATHER
V Low nvar nvar = V(4)    ! additive genetic paths MOTHER

```

```

I Ident nvar nvar          ! matrix with 1s for variance of A in parents
Z Full nvar nvar = Z(4)    ! genetic correlation between generations
H Full 1 1 = H(4)         ! scalar, .5
R Symm nvar nvar = R(8)   ! common environment covariance SON
C Low nvar nvar = C(2)    ! common environment paths SON
A Low nvar nvar = A(2)    ! additive genetic paths SON
N Low nvar nvar = N(8)    ! non-additive paths SON
F Full nvar nvar = F(4)   ! father-son cultural transmission
M Full nvar nvar = M(4)   ! mother-son cultural transmission
Q Symm nvar nvar = Q(11)  ! common environment covariance DAUGHTER
Y Low nvar nvar = Y(3)    ! common environment paths DAUGHTER
X Low nvar nvar = X(3)    ! additive genetic paths DAUGHTER
B Low nvar nvar = B(11)   ! non-additive paths DAUGHTER
E Full nvar nvar = E(6)   ! father-daughter cultural transmission
L Full nvar nvar = L(6)   ! mother-daughter cultural transmission
Compute (H*Z*Z)@A*(I+H@(U*(D'+D)*V'))*X' + C*(F*(E+D*P*L)+M*(L
+D*P*E)+(I)*O*(I))*Y' +A*((H*Z)@(U*E+U*D*P*L+V*L+V*D*P*E))*Y' +
C*((H*Z)@(F*U+F*P*D*V+M*V+M*P*D*U))*X' +H@H@N*B'/
Start .5 O 1 1
Bound -1 1 O 1 1
Option No_output
End

```

Group 15: sibOS Covariance

Data Calculation

Matrices

```

O Symm nvar nvar = O(14)  ! corr between res var C males + females
P Symm nvar nvar = P(1)   ! within person covariance (Rp)
D Full nvar nvar = D(1)   ! assortative mating delta paths
U Low nvar nvar = U(4)    ! additive genetic paths FATHER
V Low nvar nvar = V(4)    ! additive genetic paths MOTHER
I Ident nvar nvar        ! matrix with 1s for variance of A in parents
Z Full nvar nvar = Z(4)   ! genetic correlation between generations
H Full 1 1 = H(4)         ! scalar, .5
R Symm nvar nvar = R(8)   ! common environment covariance SON
C Low nvar nvar = C(2)    ! common environment paths SON
A Low nvar nvar = A(2)    ! additive genetic paths SON
N Low nvar nvar = N(8)    ! non-additive paths SON
F Full nvar nvar = F(4)   ! father-son cultural transmission
M Full nvar nvar = M(4)   ! mother-son cultural transmission

```

```

Q Symm nvar nvar = Q(11)      ! common environment covariance DAUGHTER
Y Low  nvar nvar = Y(3)       ! common environment paths DAUGHTER
X Low  nvar nvar = X(3)       ! additive genetic paths DAUGHTER
B Low  nvar nvar = B(11)      ! non-additive paths DAUGHTER
E Full nvar nvar = E(6)       ! father-daughter cultural transmission
L Full nvar nvar = L(6)       ! mother-daughter cultural transmission
Compute (H*Z*Z)@A*(I+H@(U*(D'+D)*V'))*X' + C*(F*(E+D*P*L)+M*(L
+D*P*E)+(I)*O*(I))*Y' +A*((H*Z)@(U*E+U*D*P*L+V*L+V*D*P*E))*Y' +
C*((H*Z)@(F*U+F*P*D*V+M*V+M*P*D*U))*X' +H@H@N*B/
Option No_output
End

```

```

!!!!!!!!!!!!

```

```

!!CONSTRAINTS

```

```

!!!!!!!!!!!!

```

Group 16: Genetic Constraint (Equation 3) SON=DAUGHTER

Data Constraint

Matrices

```

D Full nvar nvar = D(1)      ! assortative mating delta paths
G Symm nvar nvar = G(2)     ! additive genetic covariance SON
J Symm nvar nvar = I(3)     ! additive genetic covariance DAUGHTER
U Low  nvar nvar = U(4)     ! additive genetic paths FATHER
V Low  nvar nvar = V(4)     ! additive genetic paths MOTHER
Z Full nvar nvar = Z(4)     ! genetic correlation between generations
H Full 1 1 = H(4)           ! scalar, .5
I Iden nvar nvar            ! to form segr var, .5I + var(A) parents
Constraint G =H@((H*Z*Z)@I+(H*Z*Z)@I+(H*Z*Z)@(U*(D'+D)*V')+I) /
Option Rsidual
End

```

Group 17: A-C Constraint (Equation 2) SON

Data Constraint

Matrices

```

D Full nvar nvar = D(1)      ! assortative mating delta paths
F Full nvar nvar = F(4)     ! father-son cultural transmission
H Full 1 1 = H(4)           ! scalar, .5
M Full nvar nvar = M(4)     ! mother-son cultural transmission
P Symm nvar nvar = P(1)     ! within person covariance (Rp)
S Full nvar nvar = S(2)     ! A-C covariance SON

```

```

U Low nvar nvar = U(4)      ! additive genetic paths FATHER
V Low nvar nvar = V(4)      ! additive genetic paths MOTHER
Z Full nvar nvar = Z(4)      ! genetic correlation between generations
Constraint S=(H*Z)@U*(F'+D*P*M') + ((H*Z)@V*(M'+D*P*F')) /
Option Rsidual
End

```

Group 18: A-C Constraint (Equation 2) DAUGHTER

Data Constraint

Matrices

```

D Full nvar nvar = D(1)      ! assortative mating delta paths
E Full nvar nvar = E(6)      ! father-daughter cultural transmission
H Full 1 1 = H(4)           ! scalar, .5
L Full nvar nvar = L(6)      ! mother-daughter cultural transmission
P Symm nvar nvar = P(1)      ! within person covariance (Rp)
T Full nvar nvar = T(3)      ! A-C covariance DAUGHTER
U Low nvar nvar = U(4)      ! additive genetic paths FATHER
V Low nvar nvar = V(4)      ! additive genetic paths MOTHER
Z Full nvar nvar = Z(4)      ! genetic correlation between generations
Constraint T=(H*Z)@U*(E'+D*P*L') + ((H*Z)@V*(L'+D*P*E')) /
Option Rsidual
End

```

Group 19: Phenotypic Variance Constraint (Equation 5) SON

Constraint

Matrices

```

R Symm nvar nvar = R(8)      ! common environment covariance
C Low nvar nvar = C(2)       ! common environment paths
G Symm nvar nvar = G(2)      ! additive genetic covariance (Ra)
J Low nvar nvar free         ! specific environment paths
A Low nvar nvar = A(2)       ! additive genetic paths
P Symm nvar nvar = P(1)      ! within person covariance (Rp)
S Full nvar nvar = S(2)      ! A-C covariance
N Low nvar nvar = N(8)
Constraint P=(A*G*A' + C*R*C' + J*J' + A*S*C' + C*S*A' + N*N') /
Option Rsidual
Start .3 J 1 1
End

```

Group 20: Phenotypic Variance Constraint (Equation 5) DAUGHTER
Constraint

Matrices

Q Symm nvar nvar = Q(11) ! common environment covariance
 Y Low nvar nvar = Y(3) ! common environment paths
 I Symm nvar nvar = I(3) ! additive genetic covariance (Ra)
 K Low nvar nvar free ! specific environment paths
 X Low nvar nvar = X(3) ! additive genetic paths
 P Symm nvar nvar = P(1) ! within person covariance (Rp)
 T Full nvar nvar = T(3) ! A-C covariance
 B Low nvar nvar = B(11)

Constraint $P=(X*I*X' + Y*Q*Y' + K*K' + X*T*Y' + Y*T*X' + B*B') /$

Option Rsidual

Start .3 K 1 1

End

Group 21: Phenotypic Variance Constraint (Equation 5) FATHER
Constraint

Matrices

U Low nvar nvar = U(4) ! additive genetic paths FATHER
 E Low nvar nvar free ! specific environment paths FATHER
 I Ident nvar nvar ! matrix with 1s for var(A) in parents
 P Symm nvar nvar = P(1) ! within person covariance (Rp)

Constraint $P=(U*I*U' + E*E') /$

Option Rsidual

Start .7 E 1 1

End

Group 22: Phenotypic Variance Constraint (Equation 5) MOTHER
Constraint

Matrices

V Low nvar nvar = V(4) ! additive genetic paths MOTHER
 E Low nvar nvar free ! specific environment paths MOTHER
 I Ident nvar nvar ! matrix with 1s for var(A) in parents
 P Symm nvar nvar = P(1) ! within person covariance (Rp)

Constraint $P=(V*I*V' + E*E') /$

Option Rsidual

Start .7 E 1 1

End

Group 23: Common Environment Constraint (Equation 4) SON

Constraint

Matrices

B Iden nvar nvar ! common environment residual variance=1

R Symm nvar nvar = R(8) ! common environment covariance

F Full nvar nvar = F(4) ! father-son cultural transmission

M Full nvar nvar = M(4) ! mother-son cultural transmission

P Symm nvar nvar = P(1) ! within person covariance

W Full nvar nvar = %E1 ! spouse covariance

Constraint $R = (M*P*M' + F*P*F' + M*W*F' + F*W'*M' + B) /$

Option Rsidual

End

Group 24: Common Environment Constraint (Equation 4) DAUGHTER

Constraint

Matrices

B Iden nvar nvar ! common environment residual variance=1

Q Symm nvar nvar = Q(11) ! common environment covariance

E Full nvar nvar = E(6) ! father-daughter cultural transmission

L Full nvar nvar = L(6) ! mother-daughter cultural transmission

P Symm nvar nvar = P(1) ! within person covariance

W Full nvar nvar = %E1 ! spouse covariance

Constraint $Q = (L*P*L' + E*P*E' + L*W*E' + E*W'*L' + B) /$

Option Rsidual

End

!!!!!!

!!DATA

!!!!!!

G25 - MZM Twins and parents

Data NInput_vars= 26

Missing=-1

ORDinal_data file=sport_1995_1993_1991_mx.dat

LABELS trappreg twzyg

age1 sex1 sp1

age2 sex2 sp2

age31 sex31 sp31

age41 sex41 sp41

age11 sex11 sp11

```
age12 sex12 sp12
age16 sex16 sp16
age17 sex17 sp17
```

```
Select if twzyg = 1;           ! select MZM's
Select age1 sp1
age2 sp2
age31 sp31
age41 sp41
age11 sp11
age12 sp12
age16 sp16
age17 sp17 ;
```

```
Definition age1 age2 age31 age41 age11 age12 age16 age17;
```

```
Begin Matrices;
```

```
P Symm nvar nvar = P1           ! Within person covariances
A Full nvar nvar = %E1         ! Spouse covariances
B Full nvar nvar = %E4         ! Father-son covariances
C Full nvar nvar = %E5         ! Mother-son covariances
D Full nvar nvar = %E6         ! Father-daughter covariances
E Full nvar nvar = %E7         ! Mother-daughter covariances
F Full nvar nvar = %E8         ! MZM twin covariances
G Full nvar nvar = %E9         ! DZM twin covariances
H Full nvar nvar = %E10        ! sibMM covariances
I Full nvar nvar = %E11        ! MZF covariances
J Full nvar nvar = %E12        ! DZF covariances
K Full nvar nvar = %E13        ! sibFF covariances
L Full nvar nvar = %E14        ! DOS covariances
M Full nvar nvar = %E15        ! sibOS covariances
```

```
T Full nthres nind Free ;       !Thresholds
N Lower nthres nthres Fixed ;
O Full nthres nind Fixed;       !Beta first covariate
R Full nthres nind Fixed;       !Observed first covariate
End Matrices;
```

```
Value 1 N 1 1 - N nthres nthres
!Specify different thresholds for sons, daughters, fathers and mothers
```

Appendix B

Specify T 20 20 40 50 20 20 30 30
!Specify 1 Beta for sons, daughters, fathers and mothers
Specify O 80 80 80 80 80 80 80 80
Specify R age1 age2 age31 age41 age11 age12 age16 age17

Thresholds N*T+O.R;

Covariance (P | F | B | C | H | H | M | M _
F | P | B | C | H | H | M | M _
B | B | P | A | B | B | D | D _
C | C | A | P | C | C | E | E _
H | H | B | C | P | H | M | M _
H | H | B | C | H | P | M | M _
M | M | D | E | M | M | P | K _
M | M | D | E | M | M | K | P) ;

Start -0.65 T 1 1 - T 1 nind
Start 2.1 O 1 1 - O 1 nind

Bound -3 3 T 1 1 - T 1 nind
Option Rsiduals
End

Group 26 - DZM twins and parents
Data NInput_vars= 26
Missing=-1
ORDinal_data file=sport_1995_1993_1991_mx.dat
LABELS trappreg twzyg
age1 sex1 sp1
age2 sex2 sp2
age31 sex31 sp31
age41 sex41 sp41
age11 sex11 sp11
age12 sex12 sp12
age16 sex16 sp16
age17 sex17 sp17

Select if twzyg = 2; ! select DZM's
Select age1 sp1
age2 sp2

```

age31 sp31
age41 sp41
age11 sp11
age12 sp12
age16 sp16
age17 sp17 ;

```

```

Definition age1 age2 age31 age41 age11 age12 age16 age17;

```

```

Begin Matrices;

```

```

P Symm nvar nvar = P1           ! Within person covariances
A Full nvar nvar = %E1          ! Spouse covariances
B Full nvar nvar = %E4          ! Father-son covariances
C Full nvar nvar = %E5          ! Mother-son covariances
D Full nvar nvar = %E6          ! Father-daughter covariances
E Full nvar nvar = %E7          ! Mother-daughter covariances
F Full nvar nvar = %E8          ! MZM twin covariances
G Full nvar nvar = %E9          ! DZM twin covariances
H Full nvar nvar = %E10         ! sibMM covariances
I Full nvar nvar = %E11         ! MZF covariances
J Full nvar nvar = %E12         ! DZF covariances
K Full nvar nvar = %E13         ! sibFF covariances
L Full nvar nvar = %E14         ! DOS covariances
M Full nvar nvar = %E15         ! sibOS covariances

```

```

T Full nthres nind Free ;        !Thresholds
N Lower nthres nthres Fixed ;
O Full nthres nind Fixed;        !Beta first covariate
R Full nthres nind Fixed;        !Observed first covariate
End Matrices;

```

```

Value 1 N 1 1 - N nthres nthres
!Specify different thresholds for sons, daughters, fathers and mothers
Specify T 20 20 40 50 20 20 30 30
!Specify 1 Beta for sons, daughters, fathers and mothers
Specify O 80 80 80 80 80 80 80 80
Specify R age1 age2 age31 age41 age11 age12 age16 age17

```

```

Thresholds N*T+O.R;

```

Appendix B

```
Covariance ( P | G | B | C | H | H | M | M _  
G | P | B | C | H | H | M | M _  
B | B | P | A | B | B | D | D _  
C | C | A | P | C | C | E | E _  
H | H | B | C | P | H | M | M _  
H | H | B | C | H | P | M | M _  
M | M | D | E | M | M | P | K _  
M | M | D | E | M | M | K | P ) ;
```

Start -0.65 T 1 1 - T 1 nind

Start 2.1 O 1 1 - O 1 nind

Bound -3 3 T 1 1 - T 1 nind

Option Rsidual

End

G27 - MZF Twins and parents

Data NInput_vars= 26

Missing=-1

ORDinal_data file=sport_1995_1993_1991_mx.dat

LABELS trappreg twzyg

age1 sex1 sp1

age2 sex2 sp2

age31 sex31 sp31

age41 sex41 sp41

age11 sex11 sp11

age12 sex12 sp12

age16 sex16 sp16

age17 sex17 sp17

Select if twzyg = 3; ! select MZF's

Select age1 sp1

age2 sp2

age31 sp31

age41 sp41

age11 sp11

age12 sp12

age16 sp16

age17 sp17 ;

```

Definition age1 age2 age31 age41 age11 age12 age16 age17;
Begin Matrices;
P Symm nvar nvar = %P1           ! Within person covariances
A Full nvar nvar = %E1           ! Spouse covariances
B Full nvar nvar = %E4           ! Father-son covariances
C Full nvar nvar = %E5           ! Mother-son covariances
D Full nvar nvar = %E6           ! Father-daughter covariances
E Full nvar nvar = %E7           ! Mother-daughter covariances
F Full nvar nvar = %E8           ! MZM twin covariances
G Full nvar nvar = %E9           ! DZM twin covariances
H Full nvar nvar = %E10          ! sibMM covariances
I Full nvar nvar = %E11          ! MZF covariances
J Full nvar nvar = %E12          ! DZF covariances
K Full nvar nvar = %E13          ! sibFF covariances
L Full nvar nvar = %E14          ! DOS covariances
M Full nvar nvar = %E15          ! sibOS covariances

T Full nthres nind Free ;         !Thresholds
N Lower nthres nthres Fixed ;
O Full nthres nind Fixed;         !Beta first covariate
R Full nthres nind Fixed;         !Observed first covariate
End Matrices;

Value 1 N 1 1 - N nthres nthres
!Specify different thresholds for sons, daughters, fathers and mothers
Specify T 30 30 40 50 20 20 30 30
!Specify 1 Beta for sons, daughters, fathers and mothers
Specify O 80 80 80 80 80 80 80 80
Specify R age1 age2 age31 age41 age11 age12 age16 age17

Thresholds N*T+O.R;

Covariance ( P | I | D | E | M | M | K | K _
              I | P | D | E | M | M | K | K _
              D | D | P | A | B | B | D | D _
              E | E | A | P | C | C | E | E _
              M | M | B | C | P | H | M | M _
              M | M | B | C | H | P | M | M _
              K | K | D | E | M | M | P | K _
              K | K | D | E | M | M | K | P ) ;

```

Appendix B

Start -0.65 T 1 1 - T 1 nind
Start 2.1 O 1 1 - O 1 nind

Bound -3 3 T 1 1 - T 1 nind
Option Rsidual
End

Group 28 - DZF twins and parents
Data NInput_vars= 26
Missing=-1
ORDinal_data file=sport_1995_1993_1991_mx.dat
LABELS trappreg twzyg
age1 sex1 sp1
age2 sex2 sp2
age31 sex31 sp31
age41 sex41 sp41
age11 sex11 sp11
age12 sex12 sp12
age16 sex16 sp16
age17 sex17 sp17

Select if twzyg = 4; ! select DZF's

Select age1 sp1
age2 sp2
age31 sp31
age41 sp41
age11 sp11
age12 sp12
age16 sp16
age17 sp17 ;

Definition age1 age2 age31 age41 age11 age12 age16 age17;

Begin Matrices;
P Symm nvar nvar = P1 ! Within person covariances
A Full nvar nvar = %E1 ! Spouse covariances
B Full nvar nvar = %E4 ! Father-son covariances
C Full nvar nvar = %E5 ! Mother-son covariances
D Full nvar nvar = %E6 ! Father-daughter covariances

```

E Full nvar nvar = %E7      ! Mother-daughter covariances
F Full nvar nvar = %E8      ! MZM twin covariances
G Full nvar nvar = %E9      ! DZM twin covariances
H Full nvar nvar = %E10     ! sibMM covariances
I Full nvar nvar = %E11     ! MZF covariances
J Full nvar nvar = %E12     ! DZF covariances
K Full nvar nvar = %E13     ! sibFF covariances
L Full nvar nvar = %E14     ! DOS covariances
M Full nvar nvar = %E15     ! sibOS covariances

```

```

T Full nthres nind Free ;    !Thresholds
N Lower nthres nthres Fixed ;
O Full nthres nind Fixed;    !Beta first covariate
R Full nthres nind Fixed;    !Observed first covariate
End Matrices;

```

```

Value 1 N 1 1 - N nthres nthres
!Specify different thresholds for sons, daughters, fathers and mothers
Specify T 30 30 40 50 20 20 30 30
!Specify 1 Beta for sons, daughters, fathers and mothers
Specify O 80 80 80 80 80 80 80 80
Specify R age1 age2 age31 age41 age11 age12 age16 age17

```

```

Thresholds N*T+O.R;

```

```

Covariance ( P|J|D|E|M|M|K|K_
             J|P|D|E|M|M|K|K_
             D|D|P|A|B|B|D|D_
             E|E|A|P|C|C|E|E_
             M|M|B|C|P|H|M|M_
             M|M|B|C|H|P|M|M_
             K|K|D|E|M|M|P|K_
             K|K|D|E|M|M|K|P) ;

```

```

Start -0.65 T 1 1 - T 1 nind

```

```

Start 2.1 O 1 1 - O 1 nind

```

```

Bound -3 3 T 1 1 - T 1 nind

```

```

Option Rsidual

```

```

End

```

Group 29 - DOS-MF twins and parents

Data NInput_vars= 26

Missing=-1

ORDinal_data file=sport_1995_1993_1991_mx.dat

LABELS trappreg twzyg

age1 sex1 sp1

age2 sex2 sp2

age31 sex31 sp31

age41 sex41 sp41

age11 sex11 sp11

age12 sex12 sp12

age16 sex16 sp16

age17 sex17 sp17

Select if twzyg = 5;

! select DOS-MF's

Select age1 sp1

age2 sp2

age31 sp31

age41 sp41

age11 sp11

age12 sp12

age16 sp16

age17 sp17 ;

Definition age1 age2 age31 age41 age11 age12 age16 age17;

Begin Matrices;

P Symm nvar nvar = P1

! Within person covariances

A Full nvar nvar = %E1

! Spouse covariances

B Full nvar nvar = %E4

! Father-son covariances

C Full nvar nvar = %E5

! Mother-son covariances

D Full nvar nvar = %E6

! Father-daughter covariances

E Full nvar nvar = %E7

! Mother-daughter covariances

F Full nvar nvar = %E8

! MZM twin covariances

G Full nvar nvar = %E9

! DZM twin covariances

H Full nvar nvar = %E10

! sibMM covariances

I Full nvar nvar = %E11

! MZF covariances

J Full nvar nvar = %E12

! DZF covariances

K Full nvar nvar = %E13

! sibFF covariances

```

L Full nvar nvar = %E14      ! DOS covariances
M Full nvar nvar = %E15      ! sibOS covariances

T Full nthres nind Free ;      !Thresholds
N Lower nthres nthres Fixed ;
O Full nthres nind Fixed;      !Beta first covariate
R Full nthres nind Fixed;      !Observed first covariate
End Matrices;

Value 1 N 1 1 - N nthres nthres
!Specify different thresholds for sons, daughters, fathers and mothers
Specify T 20 30 40 50 20 20 30 30
!Specify 1 Beta for sons, daughters, fathers and mothers
Specify O 80 80 80 80 80 80 80 80
Specify R age1 age2 age31 age41 age11 age12 age16 age17

Thresholds N*T+O.R;

Covariance ( P | L | B | C | H | H | M | M _
              L | P | D | E | M | M | K | K _
              B | D | P | A | B | B | D | D _
              C | E | A | P | C | C | E | E _
              H | M | B | C | P | H | M | M _
              H | M | B | C | H | P | M | M _
              M | K | D | E | M | M | P | K _
              M | K | D | E | M | M | K | P ) ;

Start -0.65 T 1 1 - T 1 nind
Start 2.1 O 1 1 - O 1 nind

Bound -3 3 T 1 1 - T 1 nind
Option Rsidual
End

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!! SUMMARIZE RESULTS
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

Group 30: Print parameter estimates SON and FATHER
Data calc

```

Matrices

```

A Low nvar nvar = A(2)      ! additive genetic paths SON
O Full nthres nind = B(29)  ! Beta coeff age on thresholds DOS group
C Low nvar nvar = C(2)      ! common environment paths SON
D Full nvar nvar = D(1)     ! assort mating delta paths from group 1
E Low nvar nvar = E(21)     ! specific environment paths FATHER
F Full nvar nvar = F(4)     ! cultural transmission FATHER-SON
G Symm nvar nvar = G(2)    ! additive genetic covariance (Ra)
J Low nvar nvar = J(19)     ! specific environment paths SON
M Full nvar nvar = M(4)     ! cultural transmission MOTHER-SON
N Low nvar nvar = N(8)     ! non-additive genetic paths SON
P Symm nvar nvar = P(1)    ! within person covariance (Rp)
R Symm nvar nvar = R(8)    ! common environment var (Rc) SON
S Full nvar nvar = S(2)    ! A-C covariance SON
T Full nthres nind = T(29) ! threshold son, daughter, father, mother
O Full nvar nvar = O(14)   ! covariance between C SON and DAUGHTER
U Low nvar nvar = U(4)     ! additive genetic paths FATHER
Z Full nvar nvar = Z(4)    ! genetic correlation between generations

```

Compute

```

\vech(a)\vech(n)\vech(c)\vech(j)\m2v(d)\vech(g)\vech(r)\
\m2v(s)\m2v(m)\m2v(f) /
End

```

Group 31: Print parameter estimates DAUGHTERS and MOTHERS

Data calc

Matrices

```

A Low nvar nvar = X(3)      ! additive genetic paths DAUGHTER
C Low nvar nvar = Y(3)      ! common environment paths DAUGHTER
D Full nvar nvar = D(1)     ! assort mating delta paths from group 1
E Low nvar nvar = E(22)     ! specific environment paths MOTHER
F Full nvar nvar = E(6)     ! cultural transmission FATHER-DAUGHTER
G Symm nvar nvar = I(3)    ! additive genetic covariance (Ra)
J Low nvar nvar = K(20)     ! specific environment paths DAUGHTER
M Full nvar nvar = L(6)     ! cultural transmission MOTHER-DAUGHTER
N Low nvar nvar = B(11)    ! non-additive genetic paths DAUGHTER
P Symm nvar nvar = P(1)    ! within person covariance (Rp)
R Symm nvar nvar = Q(11)   ! common environment var (Rc) DAUGHTER
S Full nvar nvar = T(3)    ! A-C covariance DAUGHTER
O Full nvar nvar = O(14)   ! covariance between C SON and DAUGHTER
V Low nvar nvar = V(4)     ! additive genetic paths MOTHER

```

Z Full nvar nvar = Z(4) ! genetic correlation between generations

Compute

\vech(a)\vech(n)\vech(c)\vech(j)\m2v(d)\vech(g)\vech(r)_
 \m2v(s)\m2v(m)\m2v(f) /

Option Rsidual nd=3

!Option df=-4 !To be used when cultural transmission@0

Option func 1.e-10

Option TH=-5

Option multiple issat

End

Save gen_adol.mxs

Get gen_adol.mxs

! drop cultural transmission fa-so

Drop F 4 1 1

End

Get gen_adol.mxs

! drop cultural transmission fa-da

Drop E 6 1 1

End

Get gen_adol.mxs

! drop cultural transmission mo-so

Drop M 4 1 1

End

Get gen_adol.mxs

! drop cultural transmission mo-da

Drop L 6 1 1

End

Get gen_adol.mxs

! drop cultural transmission same-sex effects P-O

Drop F 4 1 1

Drop L 6 1 1

End

Appendix B

```
Get gen_adol.mxs
! drop cultural transmission opposite-sex effects P-O
Drop M 4 1 1
Drop E 6 1 1
End
```

```
Get gen_adol.mxs
! drop all cultural transmission
Drop F 4 1 1
Drop M 4 1 1
Drop E 6 1 1
Drop L 6 1 1
End
```

```
Get gen_adol.mxs
! drop all cultural transmission and C
Drop F 4 1 1
Drop M 4 1 1
Drop E 6 1 1
Drop L 6 1 1
Drop C 2 1 1
Drop Y 3 1 1
Drop O 14 1 1
End
```

```
Get gen_adol.mxs
! drop A in males and females
Drop A 2 1 1
Drop X 3 1 1
End
```

```
Get gen_adol.mxs
! drop assortment
Drop D 1 1 1
End
```

```
Get gen_adol.mxs
! no qual sex diff C
Drop @1 O 14 1 1
End
```

```
Get gen_adol.mxs  
! no sex differences  
EQ A 2 1 1 X 3 1 1  
EQ C 2 1 1 Y 3 1 1  
EQ J 19 1 1 K 20 1 1  
!EQ N 8 1 1 B 11 1 1  
EQ F 4 1 1 E 6 1 1  
EQ M 4 1 1 L 6 1 1  
Drop @1 O 14 1 1  
End
```



Appendix C

Expected covariances among
parents and offspring based
on parent-offspring model specified
in Appendix B



Note that it is assumed that the phenotypes are standardized (variances=1). G1 to G24 refer to the groups in the Mx script (see Appendix B) where the covariances are specified. The same notation for parameters is used as in Figure 8.1.

G1 – Father-Mother covariance

$$\text{cov}(P_{FA}, P_{MO}) = D$$

G2 – Phenotype-Genotype covariance Son

$$\text{cov}(P_{SO}, A_{SO}) = a_{SO} * \text{var}(A_{SO}) + c_{SO} * \text{cov}(A_{SO}, C_{SO})$$

G3 – Phenotype-Genotype covariance Daughter

$$\text{cov}(P_{DA}, A_{DA}) = a_{DA} * \text{var}(A_{DA}) + c_{DA} * \text{cov}(A_{DA}, C_{DA})$$

G4 – Father-Son covariance

$$\text{cov}(P_{FA}, P_{SO}) = c_{SO} * (t_{FA,SO} + D * t_{MO,SO}) + 0.5 * a_{FA} * a_{SO} + 0.5 * a_{MO} * a_{SO} * D$$

G5 – Mother-Son covariance

$$\text{cov}(P_{MO}, P_{SO}) = c_{SO} * (t_{MO,SO} + D * t_{FA,SO}) + 0.5 * a_{MO} * a_{SO} + 0.5 * a_{FA} * a_{SO} * D$$

G6 – Father-Daughter covariance

$$\text{cov}(P_{FA}, P_{DA}) = c_{DA} * (t_{FA,DA} + D * t_{MO,DA}) + 0.5 * a_{FA} * a_{DA} + 0.5 * a_{MO} * a_{DA} * D$$

G7 – Mother-Daughter covariance

$$\text{cov}(P_{MO}, P_{DA}) = c_{DA} * (t_{MO,DA} + D * t_{FA,DA}) + 0.5 * a_{MO} * a_{DA} + 0.5 * a_{FA} * a_{DA} * D$$

G8 – MZM Twin covariance

$$\text{cov}(MZM) = a_{SO}^2 * \text{var}(A_{SO}) + c_{SO}^2 * \text{var}(C_{SO}) + 2 * a_{SO} * c_{SO} * \text{cov}(A_{SO}, C_{SO})$$

G9 – DZM Twin covariance

$$\text{cov}(DZM) = 0.5 * a_{SO}^2 * (1 + a_{FA} * a_{MO} * D) + c_{SO}^2 * \text{var}(C_{SO}) + 2 * a_{SO} * c_{SO} * \text{cov}(A_{SO}, C_{SO})$$

G10 – Male sibling covariance

$$\text{cov}(P_{SO}, P_{SO}) = 0.5 * a_{SO}^2 * (1 + a_{FA} * a_{MO} * D) + c_{SO}^2 * \text{var}(C_{SO}) + 2 * a_{SO} * c_{SO} * \text{cov}(A_{SO}, C_{SO})$$

G11 – MZF Twin covariance

$$\text{cov}(MZF) = a_{DA}^2 * \text{var}(A_{DA}) + c_{DA}^2 * \text{var}(C_{DA}) + 2 * a_{DA} * c_{DA} * \text{cov}(A_{DA}, C_{DA})$$

G12 – DZF Twin covariance

$$\text{cov}(DZF) = 0.5 * a_{DA}^2 * (1 + a_{FA} * a_{MO} * D) + c_{DA}^2 * \text{var}(C_{DA}) + 2 * a_{DA} * c_{DA} * \text{cov}(A_{DA}, C_{DA})$$

G13 – Female sibling covariance

$$\text{cov}(P_{DA}, P_{DA}) = 0.5 * a_{DA}^2 * (1 + a_{FA} * a_{MO} * D) + c_{DA}^2 * \text{var}(C_{DA}) + 2 * a_{DA} * c_{DA} * \text{cov}(A_{DA}, C_{DA})$$

G14 – DOS Twin covariance

$$\begin{aligned} \text{cov}(DOS) = & 0.5 * a_{SO} * a_{DA} * (1 + a_{FA} * a_{MO} * D) + \\ & c_{SO} * c_{DA} * (t_{FA,SO} (t_{FA,DA} + D * t_{MO,DA}) + t_{MO,SO} (t_{MO,DA} + D * t_{FA,DA}) + r_{C,OS}) + \\ & 0.5 * a_{SO} * c_{DA} * (a_{FA} * t_{FA,DA} + a_{FA} * D * t_{MO,DA} + a_{MO} * t_{MO,DA} + a_{MO} * D * t_{MO,DA}) + \\ & 0.5 * c_{SO} * a_{DA} * (a_{FA} * t_{FA,SO} + a_{MO} * D * t_{FA,SO} + a_{MO} * t_{MO,SO} + a_{FA} * D * t_{MO,SO}) \end{aligned}$$

G15 – Opposite-sex sibling covariance

$$\begin{aligned} \text{cov}(P_{SO}, P_{DA}) = & 0.5 * a_{SO} * a_{DA} * (1 + a_{FA} * a_{MO} * D) + \\ & c_{SO} * c_{DA} * (t_{FA,SO} (t_{FA,DA} + D * t_{MO,DA}) + t_{MO,SO} (t_{MO,DA} + D * t_{FA,DA}) + r_{C,OS}) + \\ & 0.5 * a_{SO} * c_{DA} * (a_{FA} * t_{FA,DA} + a_{FA} * D * t_{MO,DA} + a_{MO} * t_{MO,DA} + a_{MO} * D * t_{MO,DA}) + \\ & 0.5 * c_{SO} * a_{DA} * (a_{FA} * t_{FA,SO} + a_{MO} * D * t_{FA,SO} + a_{MO} * t_{MO,SO} + a_{FA} * D * t_{MO,SO}) \end{aligned}$$

G16 – Genetic variance constraint Son=Daughter

$$\text{var}(A_{SO}) = \text{var}(A_{DA}) = 0.5 * (1 + a_{FA} * D * a_{MO}) + 0.5$$

G17 – A-C covariance constraint Son

$$\text{cov}(A_{SO}, C_{SO}) = 0.5 * a_{FA} * (t_{FA,SO} + D * t_{MO,SO}) + 0.5 * a_{MO} * (t_{MO,SO} + D * t_{FA,SO})$$

G18 – A-C covariance constraint Daughter

$$\text{cov}(A_{DA}, C_{DA}) = 0.5 * a_{FA} * (t_{FA,DA} + D * t_{MO,DA}) + 0.5 * a_{MO} * (t_{MO,DA} + D * t_{FA,DA})$$

G19 – Phenotypic variance constraint Son

$$\text{var}(P_{\text{SO}}) = a_{\text{SO}}^2 * \text{var}(A_{\text{SO}}) + c_{\text{SO}}^2 * \text{var}(C_{\text{SO}}) + e_{\text{SO}}^2 + 2 * a_{\text{SO}} * c_{\text{SO}} * \text{cov}(A_{\text{SO}}, C_{\text{SO}})$$

G20 – Phenotypic variance constraint Daughter

$$\text{var}(P_{\text{DA}}) = a_{\text{DA}}^2 * \text{var}(A_{\text{DA}}) + c_{\text{DA}}^2 * \text{var}(C_{\text{DA}}) + e_{\text{DA}}^2 + 2 * a_{\text{DA}} * c_{\text{DA}} * \text{cov}(A_{\text{DA}}, C_{\text{DA}})$$

G21 – Phenotypic variance constraint Father

$$\text{var}(P_{\text{FA}}) = a_{\text{FA}}^2 + e_{\text{FA}}^2$$

G22 – Phenotypic variance constraint Mother

$$\text{var}(P_{\text{MO}}) = a_{\text{MO}}^2 + e_{\text{MO}}^2$$

G23 – Common environment variance constraint Son

$$\text{var}(C_{\text{SO}}) = t_{\text{MO,SO}}^2 + t_{\text{FA,SO}}^2 + 2 * t_{\text{MO,SO}} * D * t_{\text{FA,SO}} + 1$$

G24 – Common environment variance constraint Daughter

$$\text{var}(C_{\text{DA}}) = t_{\text{MO,DA}}^2 + t_{\text{FA,DA}}^2 + 2 * t_{\text{MO,DA}} * D * t_{\text{FA,DA}} + 1$$



List of publications



Published articles

- De Moor, M.H.M.,** Beem, A.L., Stubbe, J.H., Boomsma, D.I., & De Geus, E.J.C. (2006). Regular exercise, anxiety, depression and personality: A population-based study. *Preventive Medicine*, 42 (4), 273-279.
- Rebollo, I., **De Moor, M.H.M.,** Dolan, C.V., & Boomsma, D.I. (2006). Phenotypic factor analysis of family data: Correction of the bias due to dependency. *Twin Research and Human Genetics*, 9 (3), 367-376.
- De Moor, M.H.M.,** Stubbe, J.H., Boomsma, D.I., & De Geus, E.J.C. (2007). Exercise participation and self-rated health: Do common genes explain the association? *European Journal of Epidemiology*, 22 (1), 27-32.
- Stubbe, J.H., **De Moor, M.H.M.,** Boomsma, D.I., & De Geus, E.J.C. (2007). The association between exercise participation and well-being: A co-twin study. *Preventive Medicine*, 44 (2), 148-152.
- De Moor, M.H.M.,** Posthuma, D., Hottenga, J.J., Willemsen, G., Boomsma, D.I., & De Geus, E.J.C. (2007). Genome-wide linkage scan for exercise participation in Dutch sibling pairs. *European Journal of Human Genetics*, 15 (12), 1252-1259.
- De Moor, M.H.M.,** Spector, T.D., Cherkas, L.F., Falchi, M., Hottenga, J.J., Boomsma, D.I., & De Geus, E.J.C. (2007). Genome-wide linkage scan for athlete status in 700 British female DZ twin pairs. *Twin Research and Human Genetics*, 10 (6), 812-820.
- De Moor, M.H.M.,** Boomsma, D.I., Stubbe, J.H., Willemsen, G., & De Geus, E.J.C. (2008). Testing causality in the association between regular exercise and symptoms of anxiety and depression. *Archives of General Psychiatry*, 65 (8), 897-905.
- De Geus, E.J.C. & **De Moor, M.H.M.** (2008). A genetic perspective on the association between exercise and mental health. *Mental Health and Physical Activity*, in press.
- De Moor, M.H.M.,** Distel, M.A., Trull, T.J. & Boomsma, D.I. (2009). Assessment of Borderline Personality features in population samples: Is the PAI-BOR scale measurement invariant across gender and age? *Psychological Assessment*, in press.
- Distel, M.A., **De Moor, M.H.M.** & Boomsma, D.I. Factorstructuur en betrouwbaarheid van de Nederlandse vertaling van de Personality Assessment Inventory-Borderline schaal (PAI-BOR) in een niet-klinische steekproef. *Psychologie en Gezondheid*, in press.
- De Moor, M.H.M.,** Boomsma, D.I. & De Geus, E.J.C. (2008). Sportgedrag en depressie: De rol van erfelijke factoren. *Tijdschrift voor Neuropsychiatrie & Gedragsneurologie*, in press.

- De Moor, M.H.M.**, Liu, Y-J., Boomsma, D.I., Li, J., Hamilton, J.J., Hottenga, J.J., Levy, S., Liu, X-G., Pei, J-F., Posthuma, D., Recker, R.R., Sullivan, P.F., Wang, L., Willemsen, G., Yan, H., De Geus, E.J.C. & Deng, H-W. Genome-wide association study of leisure-time exercise behavior in Dutch and American adults, *Medicine and Science in Sports and Exercise*, under review.
- De Moor, M.H.M.**, Beem, A.L., Rebollo-Mesa, I., Stubbe, J.H., Willemsen, G., De Geus, E.J.C. & Boomsma, D.I. Exercise participation in adolescents and their parents: Evidence for genetic and generation specific environmental effects, *Behavior Genetics*, under review.

Published book chapters

- De Moor, M.H.M.** & Bartels, M. Geluk en welbevinden (2008). In: *Tweelingonderzoek – Wat meerlingen vertellen over de mens* (ed. D. Boomsma). Amsterdam: VU Uitgeverij.

Published abstracts

- De Moor, M.H.M.**, Beem, A.L., Stubbe, J.H., Boomsma, D.I., & De Geus, E.J.C. (2005). The association between regular exercise and negative affect in twins, siblings and other family members: A study from the Netherlands Twin Registry. *Twin Research and Human Genetics*, 8 (4), 417.
- De Moor, M.H.M.**, Stubbe, J.H., Boomsma, D.I. & De Geus, E.J.C. (2007). Genetics of exercise behavior: Findings from the Netherlands Twin Register. *Behavior Genetics*, 37 (6), 747.
- De Moor, M.H.M.**, Boomsma, D.I. & De Geus, E.J.C. (2007). Association between regular exercise and anxious depression: A genetic analysis. *Behavior Genetics*, 37 (6), 746.
- De Moor, M.H.M.**, Boomsma, D.I., Martin, N.M., Spector, T.D. & De Geus, E.J.C. (2007). Genome-wide linkage scan for exercise participation across seven countries participating in the Genomewin project. *Twin Research and Human Genetics*, 10 (suppl), 28.
- De Moor, M.H.M.**, Posthuma, D., Sullivan, P.F., Boomsma, D.I. & De Geus, E.J.C. (2008). Genome-wide association study of leisure-time exercise behavior. *Behavior Genetics*, 38 (6), 612.
- De Moor, M.H.M.**, Boomsma, D.I., Stubbe, J.H., Willemsen, G. & De Geus, E.J.C. (2008). Causality in genetic epidemiology: An application to exercise and depression. *Behavior Genetics*, 38 (6), 612.
- Bartels, M., **De Moor, M.H.M.**, Boomsma, D.I. (2008). Genome-wide linkage scan for well-being. *Behavior Genetics*, 38 (6), 612.





Dankwoord



Het schrijven van dit proefschrift was niet mogelijk geweest zonder de hulp en steun van een heleboel mensen. De volgende mensen wil ik bedanken:

Mijn promotoren. Dorret en Eco, bedankt voor de inhoudelijke bijdragen, de snelle correcties, de verhelderende inzichten en de inspirerende discussies! Dank ook voor alle tijd en energie die jullie in dit proefschrift hebben gestoken.

De leden van de promotiecommissie. Dank voor het lezen van mijn proefschrift en voor jullie bijdrage aan de verdediging. Tim Spector, thank you for our collaboration on chapter 6, and for reading and approving my thesis. Hermine Maes and David Evans, thank you for traveling to the Netherlands and for your contributions to my defense.

Alle aio's en andere collega's. Bedankt voor de wetenschappelijke uitwisseling en de niet altijd wetenschappelijke maar wel heel gezellige kletspraatjes! In het bijzonder wil ik mijn kamergenoten van de afgelopen 4 jaar bedanken. Janine, veel dank voor de dataverzameling van lijst 6, je prachtige proefschrift waar ik op voort kon bouwen en de gezelligheid als kamergenoot! Ik ben erg blij dat je als opponent bij mijn verdediging aanwezig bent. Irene, I really enjoyed sharing scientific ideas and knowledge with you, and of course also discussing our personal lives. I am already missing you as a great roommate and colleague!

Sylvana, ook al was je relatief kort mijn kamergenoot, ik denk met plezier terug aan de gezelligheid die je mee bracht en de goede gesprekken. Marijn, dank voor je reisgezelschap tijdens het bezoeken van congressen en workshops en voor de fijne samenwerking bij het schrijven van onze gezamenlijke papers. Ik hoop dat je nog een tijdje mijn kamergenoot blijft!

Al mijn vriendinnen en vrienden. In het bijzonder wil ik Anne, Froukje, Janneke, Maartje en Noortje bedanken. Met 5 vriendinnen die zelf ook in het onderzoek zitten kon ik altijd bij iemand terecht om de pieken en dalen van het promotieproces te delen. Bedankt voor jullie steun en vriendschap! Laura, bedankt voor de mooie cover en lay-out.

Mijn familie. Mijn ouders en zussen wil ik bedanken voor hun onvoorwaardelijke steun en interesse in mijn onderzoek. Anouk, ik leerde je kennen toen ik net als aio was begonnen. Het is voor mij erg bijzonder dat je het hele promotieproces van zo dichtbij hebt meegemaakt. Ik hoop dat wij samen nog heel veel mooie dingen gaan meemaken!

Marleen, 15 december 2008

