Molecular and behavioral genetic investigation of voluntary exercise behaviors

van der Zee, Matthijs Daniël

2021

Matthijs Daniël van der Zee

Molecular and behavioral genetic investigation of voluntary exercise behaviors

The studies described in this thesis were carried out at the Department of Biological Psychology, Vrije Universiteit Amsterdam, and the Netherlands Twin Register (NTR). Funding for the NTR was provided by ZonMw (904-61-090, 985-10-002, 912-10-020, 904-61-193, 480-04-004, 463-06-001, 451-04-034, 400-05-717, 016-115-035, 481-08-011 and 056-32-010), Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Addiction-31160008, NWO-Middelgroot-911-09-032, OCW_NWO Gravity program–024.001.003, NWO-Groot 480-15-001/674, NWO-56-464-14192), Centre for Medical Systems Biology (CSMB, NWO Genomics), Biobanking and Biomolecular Resources Research Infrastructure (184.021.007, 184.033.111), Koninklijke Nederlandse Akademie van Wetenschappen (NL) (PAH/6635), European Science Foundation (EU/QLRT-2001-01254), FP7 Health (01413: ENGAGE, 602768: ACTION), H2020 European Research Council (ERC AG 230374, ERC SG 284167, ERC CG 771057), National Institutes of Health (NIH R01 DK092127-04) and Avera Institute for Human Genetics.

ISBN/EAN 978-90-831893-4-5

Paranymphs Tamara J. Pelgrim-de Haan
Margot P. van de Weijer

Design Rowdy Pelgrim

Print Print Service Ede

Printing of this thesis has been accomplished with gratefully acknowledged financial support provided by Vrije Universiteit Amsterdam.

VRIJE UNIVERSITEIT

MOLECULAR AND BEHAVIORAL GENETIC INVESTIGATION OF VOLUNTARY EXERCISE BEHAVIORS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. C.M. van Praag,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op vrijdag 19 november 2021 om 11.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Matthijs Daniël van der Zee

geboren te Cappelle aan den IJssel

promotor:	prof.dr. J.C.N. de Geus
copromotoren:	dr. J. van Dongen dr. E. Ehli
promotiecommissieleden:	prof.dr. P.J. Beek prof.dr. Y. Pitsiladis prof.dr. J. Kaprio dr. H. Van Der Ploeg dr. N. Schutte

Contents

CHAPTER 1

Introduction

Throughout our evolutionary history, humans lived in a physically demanding environment. Hunting, gathering and the constant threat of predators caused genes in humans to be specifically selected to cope with this environment causing most of our biological systems to function optimally in high physically active environments (Booth et al., 2008). Since the industrial era many technological advancements for both home- and work-environments have continuously decreased the necessity of physical activity. This has had the benefit of reducing physical hardships on workers, and substantially reducing the amount of disabilities caused by heavy labor demanding jobs. Unfortunately, however, these beneficial aspects are mirrored by detrimental effects. Nowadays, physical activity is considered a well-established major contributor to both physical and mental health (Bauman et al., 2016; Berlin & Colditz, 1990; de Geus, 2020; Ekelund et al., 2015; Fishman et al., 2016; Garber et al., 2011; Janssen & Leblanc, 2010; Lee et al., 2012; Morris et al., 1980; Samitz et al., 2011; Zhu et al., 2016) with an estimated economic burden of $53.8 billion worldwide (Ding et al., 2016). Despite many public health recommendations (Haskell et al., 2007; Kohl et al., 2012; "Physical Activity Guidelines for Americans," ; World Health Organization, 2013) and calls for worldwide governmental policy changes (Kohl et al., 2012), a large proportion of the population still does not engage in enough physical activity for optimal health-benefits, regardless of which guidelines are used (Ekelund et al., 2016; Hallal et al., 2012; Lee et al., 2012; World Health Organization, 2014). Moreover, it has recently been suggested that the population of high-income countries spends the majority of awake time being physically inactive (Hansen et al., 2012; Matthews et al., 2008).

The broadly defined physical activity has proven difficult to measure reliably through self-report (Prince et al., 2008). More reliable objective alternatives, such as accelerometers, on the other hand are

less viable options for large-scale population-based studies. An important component of physical activity, for both research and public health in general, is voluntary regular exercise behavior, such as jogging, swimming or going to a gym. Measuring voluntary regular exercise through self-report has proven to be more reliable than overall physical activity (de Moor et al., 2008; Haase et al., 2004; Stubbe et al., 2006; Stubbe et al., 2007). Additionally, voluntary regular exercise behavior has been shown to be a highly stable trait (van der Zee, van der Mee, et al., 2019). On top of that, whilst overall physical activity can be highly dependent on environmental factors that may be difficult to change (e.g. physical labor, walkability, and the distance from home to work), there is less dependence of regular exercise behavior on these factors, thus making regular exercise behavior a more feasible target for behavioral intervention.

In order to develop (more) effective interventions more knowledge of the determinants of the behavior are required (Dishman et al., 1985). These determinants can have different sources. For instance, many studies have assessed environmental determinants such as access to facilities (Matson-Koffman et al., 2005), socio-economic status (Haase et al., 2004; Varo et al., 2003), job strain (Payne et al., 2005), or social support (Sherwood & Jeffery, 2000). Apart from, and in addition to the environment, determinants of exercise behavior can also be genetic, either influencing exercise behavior directly such as through endurance capacity (de Geus, 2020; de Geus & de Moor, 2008; Schutte et al., 2019), or indirectly through intrapersonal traits such as personality (H. J. Eysenck et al., 1982) or subjective benefits and barriers to exercise (Schutte et al., 2019). Another possible biological determinant would be epigenetic effects (non-permanent changes to DNA that can alter gene expression), which, similar to genetic effects, can influence exercise behaviour directly or indirectly. Unlike genetic effects, the causal direction between epigenetic changes and exercise

behaviour is more complex, as epigenetic change can causally influence exercise behaviour (either directly or indirectly), but exercise behaviour may also induce epigenetic changes (either directly or indirectly).

With a growing number of scientific studies into the contribution of genetic factors to variation in regular exercise behavior it has become increasingly clear that genetic factors are an important determinant for this behavior (Bartels et al., 2012; de Geus et al., 2003; de Moor, Stubbe, et al., 2007; den Hoed et al., 2013; Huppertz et al., 2012; Mustelin et al., 2009; Stubbe et al., 2006; van der Zee et al., 2020). Despite this large contribution of genetic factors to the variance in regular exercise behavior, finding molecular genetic variants that have a significant effect has proven to be a difficult endeavor. Some genetic variants have been identified through candidate gene studies (Berentzen et al., 2008; Cole et al., 2010; Fuentes et al., 2002; Gielen et al., 2014; Hubáček et al., 2011; Huppertz, Bartels, Groen-Blokhuis, et al., 2014; Jozkow et al., 2013; Kostrzewa et al., 2015; Loos et al., 2005; Lorentzon et al., 2001; Mäestu et al., 2013; Murakami et al., 2011; Richert et al., 2007; Simonen, Rankinen, Pérusse, et al., 2003; Stefan et al., 2002; van der Mee et al., 2017; Wilkinson et al., 2013; Winnicki et al., 2004) but only a handful of these have been independently replicated, and only a small number of genome-wide association studies have been performed (de Moor et al., 2009; Doherty et al., 2018; Kim et al., 2014; Klimentidis et al., 2018; Lin et al., 2018), with even fewer producing significant results. In this thesis several methods will be described and applied to further our understanding of the genetic etiology of voluntary regular exercise behavior.

Voluntary regular exercise behavior in the Netherland Twin Register

The Netherlands Twin Register (NTR) is an ongoing research initiative where twins, their relatives, and their partners volunteer to complete extensive surveys every two to three years. A detailed description of methods and procedures used for data collection in the NTR is provided elsewhere (Ligthart et al., 2019). Throughout the years the same questions on regular exercise behavior were included in most of these surveys, which has led to a total of 162,286 observations of voluntary regular exercise behavior by 78,525 unique individuals. To assess voluntary regular exercise behavior participants were first asked whether they regularly exercise ('yes' or 'no'), if affirmative they were asked a number of follow up questions. They were asked to list all exercise activities they regularly did, and per activity they were asked (1) for how many years they have been doing so, (2) how many months a year (on average) they do this activity, (3) the weekly frequency, and (4) the average duration of each exercise bout. The definition of voluntary regular exercise behavior adhered to throughout this thesis does not include the following activities: non-physical sports such as chess, seasonal activities such as skiing during a winter holiday (i.e. any activity that is performed less than 6 months a year), irregular activities performed less than once a week, non-voluntary activities such as physical education, biking as a form of transportation, and any activity as a result of physical labor. Throughout my thesis these activities were treated the same as no exercise.

The most common way to summarize this information into a single variable is by first assigning each exercise activity the age-appropriate Metabolic Equivalent of Task (MET) score (Ainsworth et al., 2011; Ridley et al., 2008). This MET value is the ratio of energy expenditure rate during an exercise activity compared to sitting at rest (1 MET). An activity with a higher MET score indicates greater rate of

energy expenditure. Then by multiplying this MET score with the weekly frequency and average duration per exercise activity a single value of MET-minutes (or MET-hours) per week is obtained per activity. Finally these values are summed across all activities performed by an individual to obtain a single total MET-minutes per week (METmin/wk) value for that individual.

This total value, however, is still a broad definition of exercise behavior, as it includes all activities, which as described previously may yield less reliable results compared to a more homogenous and narrow definition, especially in genetic association studies (N. Cai et al., 2020). Ideally, instead of taking the total value, we would study each different type of exercise as a separate outcome. However, doing so would dilute any sample size too much, as even the most endorsed activities only occur in about a quarter of the regular exercisers at best. Alternatively, exercise can be separated into several different dimensions. One of these dimensions is team-based exercise, which includes the activities where a group of individuals have to cooperate towards common objectives; the other end of this dimension is solitary exercise where the activity is performed by a single exerciser. Another is competitive exercise, which includes activities where an exerciser or a group of exercisers compete against another exerciser or group of exercisers for victory. Lastly externally paced exercise includes activities where the rate of performing that activity is determined by external factors (such as other exercisers in basketball, or a changing environment in windsurfing), whereas internally paced exercise includes activities where the rate of performing the exercise is solely determined by the exerciser.

In **Chapter 2** of this thesis I leveraged the unique dataset provided by the NTR to assess the longitudinal stability of total exercise, and the exercise domains separately. This chapter also includes a more detailed description of the definitions of exercise and

its domains in the NTR, as well as extensive descriptive statistics of total exercise and the exercise domains. Furthermore, the exercise domains were also used in later chapters where different genetically informed designs were used.

Twin-, family and extended family-study designs

Exercise behavior, like many other behavioral traits 'runs in the family'. Within any given family the odds of an individual being an exerciser increases if other family members are also exercisers. This resemblance can, as alluded to earlier, be due to environmental factors (sometimes referred to as 'nurture') or genetic factors (sometimes referred to as 'nature'). By analyzing monozygotic (MZ) and dizygotic (DZ) twins we can disentangle the relative influence of environmental and genetic variation on a given trait. In the classical twin model we leverage the knowledge that MZ twins share (close to) all their inherited genetic variants, whereas dizygotic twins share half of their inherited genotypes on average (Falconer & Mackay, 1996; Plomin et al., 2013). In the classical twin model the relative impact of three of four possible sources of variance can be estimated. First, additive genetic factors (A) that represent the sum of linear effects of genetic variants that influence a given trait. Second, unique environmental factors (E) that represent environmental factors unique to any single individual, such as different lifestyles, or life events, and includes measurement error. Third shared environmental factors (C) which represent environmental factors shared within a family, such as parenting, socio-economic status, and in the case of twins the prenatal environment. Lastly, non-additive genetic factors (D), which represent non-linear effects of genetic variants, such as dominance effects, or interaction effects (Plomin et al., 2013). Since MZ twins share all factors included in A, all factors in C, and all factors in D, any difference between MZ twins can only be ascribed to unique environmental factors. However, since

DZ twins only share half of the factors included in A on average, if MZ resemblance is higher when compared to DZ resemblance, this is indicative of additive genetic factors being a determinant for the trait.

Due to the limited number of available correlations and covariances in a classical twin study, only three of these four factors can be tested simultaneously. But this issue can be overcome by including multiple different family-members. Parents of twins for example, share half of their additive genetic factors with their offspring, as well as their shared environment, but none of their non-additive genetic factors. Biological aunts and uncles share a quarter of their additive-genetic factors with their nieces and nephews, but none of their shared environmental factors. Including such extended pedigree-data with more variation in shared A, C, D, and E factors makes it possible to take all four factors into account simultaneously (Keller et al., 2009).

In **Chapters 3 and 4** of this thesis I describe twin and family-models in more detail and how these designs can elucidate the genetic effects of voluntary regular exercise behavior. Additionally **Chapter 3** of this thesis provides an overview and meta-analysis of twin-studies on voluntary regular exercise behavior, as well as other related physical activity phenotypes such as moderate-to-vigorous physical activity. **Chapter 4** of this thesis not only describes twin-studies, but also includes an overview of family-studies on regular exercise behavior. In **Chapter 5**, I present the results of our extended-family study on regular exercise behavior, and the six exercise domains. In **Chapter 5** I also explore the various sources of spousal resemblance in exercise behavior.

Genome-wide association studies (GWAS)

As mentioned previously, finding genetic variants related to exercise behavior has proven to be challenging. A reason for this might

well be that, like many other complex behavioral traits, genetic variants contributing to exercise behavior are likely widespread across the genome, all with very small effect sizes (Buniello et al., 2019). Moreover, the pathways through which these variants affect exercise behavior may also vary wildly. Some may affect exercise through the brain, some may affect exercise through pathways related to muscle- or bone-development, others through metabolism, or blood-composition. Some genetic variants have been identified through candidate-gene studies (Berentzen et al., 2008; Cole et al., 2010; Fuentes et al., 2002; Gielen et al., 2014; Hubáček et al., 2011; Huppertz, Bartels, Groen-Blokhuis, et al., 2014; Jozkow et al., 2013; Kostrzewa et al., 2015; Loos et al., 2005; Lorentzon et al., 2001; Mäestu et al., 2013; Murakami et al., 2011; Richert et al., 2007; Simonen, Rankinen, Pérusse, et al., 2003; Stefan et al., 2002; van der Mee et al., 2017; Wilkinson et al., 2013; Winnicki et al., 2004), where exercise behavior is tested for association with one (or a few) genetic variants. Unfortunately, replication for these variants that show significant association is minimal. A better alternative to detect the many genetic variants likely influencing exercise compared to these candidate-gene approach would be a genome-wide association study (GWAS). In a GWAS millions of single nucleotide polymorphisms (SNPs) spread across the human genome are tested for their effect on a given trait simultaneously (Claussnitzer et al., 2020; Visscher et al., 2012; Visscher et al., 2017). By testing a wide range of SNPs known to vary between humans, GWAS studies are free from selection of genetic variants based on previous knowledge, and often yield results related to genes that would not have been selected based on prior knowledge. The downside of GWAS is that stringent statistical corrections are applied due to the millions of tests performed simultaneously. Due to these corrections, and the expected small effect sizes of individual variants, the sample sizes required for a GWAS are much larger than conventional studies.

Two GWASs have been published before the beginning of my PhD project (de Moor et al., 2009; Kim et al., 2014), both of which did not yield results that withstand the scrutiny of the multiple testing correction (*p*-value below 5×10^{-8}). During the creation of my thesis three more GWAS studies have been published, one of which did not yield significant results (Lin et al., 2018), whereas the other two produced the first genome-wide significant genetic variants for exercise behavior (Hara et al., 2018; Klimentidis et al., 2018).

In **Chapter 6** of this thesis I present the results of a GWAS on total volume of exercise behavior, and the six exercise domains described earlier. This chapter also includes a more detailed description of previously published GWAS and candidate-gene studies.

Post-GWAS: Polygenic Scores and Mendelian Randomization

The results of GWA studies can in turn be used for a number of different methods, the most common of which is the polygenic score (PGS), sometimes also referred to as polygenic risk score (L. E. Duncan et al., 2019; Slunecka et al., 2021). The results of a GWAS include effect-sizes, often as regression coefficients, for each genetic variant, related to a specific allele of that genetic variant. These estimates indicate a predicted change in the outcome trait per copy of the associated allele an individual has. In other words, a positive estimate results in a predicted increase in volume of exercise behavior for each copy of the associated allele an individual has. By multiplying the estimates from a GWAS by the number of effect alleles per variant, and summing these estimates for all variants a polygenic score is obtained for each individual, which represents the sum of the predicted genetic effects on that individual's exercise behavior. It is important to note that when these scores are intended to be used for any kind of prediction the

sample used for the GWAS should be a different sample than the sample in which the PGS is generated in order to prevent bias.

By leveraging the information provided by genetic information from GWAS and PGS, Mendelian Randomization (MR) can be used to test causality in an observational epidemiological setting (Davey Smith & Ebrahim, 2003). Briefly, MR is based on Mendel's second law (Bateson & Mendel, 2013), the law of random assortment, which states that the inheritance for one trait is independent of the inheritance of another trait. Because of this, if a subject is genetically more liable for trait A, and trait A causes trait B, in an epidemiological study the group with higher susceptibility for trait A will also have an increase in trait B. Note this does assume that at the very least trait A has some genetic basis, and if a genetic proxy is used for trait A it needs to explain a significant portion of the variance in the trait, put otherwise it must be a strong genetic instrument. For most complex behavioral traits, influenced by many genes of small effect, only a PGS may qualify as a strong instrument. However, one of the key assumptions of MR is that of no pleiotropy, i.e. genes that cause trait A exclusively influence trait B through their causal effect on trait A. This assumption is not likely to hold for polygenic scores (Solovieff et al., 2013; Visscher & Yang, 2016). By using twin-data in conjunction with MR this assumption can be relaxed and pleiotropy directly estimated (Minică et al., 2018), also controlling for assortment (i.e. choosing a partner based on their similar exercise behavior) and dynastic effects (i.e. direct effects of parental genotype on the offspring phenotype) (Minică et al., 2020).

In **Chapter 7** of this thesis I use Mendelian Randomization with a PGS, in conjunction with a number of other methods to assess possible causal effects of the 'Big Five' personality dimensions (Digman, 1990) on voluntary exercise behavior, both total volume and the two main dimensions of competitive team-based exercise activities and non-competitive solitary exercise activities.

Finally, in **Chapter 8** the main results and conclusions of the chapters contained in this thesis are summarized. Also in this chapter I will discuss my perspective on the future development of genetic etiology of regular exercise behavior and possible implications.

Tracking of Voluntary Exercise Behaviour over the Lifespan

Published as van der Zee MD, van der Mee D, Bartels M, de Geus EJC (2019). Tracking of voluntary exercise behaviour over the lifespan. *Int J Behav Nutr Phys Act. 4;16(1):17. doi: 10.1186/s12966-019-0779-4.*

Abstract

Background

The aim of many physical activity interventions is to develop life-long habits of regular exercise and sports activities in leisure time. Previous studies that assessed tracking (i.e., the stability of a trait or behavior over the lifespan) of leisure time exercise behavior across various parts of the life span have treated it as a uniform construct by summing all types of leisure time exercise activities into a single summary score for the total volume of exercise. This study provides new insight by additionally determining tracking across leisure time exercise activities in six different domains: (1) team-based versus solitary activities, (2) competitive versus non-competitive activities, and (3) externally paced versus internally paced activities. We also assessed which of the domains of exercise activities best predicted total volume of exercise at follow-up.

Methods

A large dataset (*N* = 43,889) from the Netherlands Twin Register (NTR) was used to analyze the tracking of exercise behavior over time. Using this dataset, we were able to examine tracking as a function of baseline age (8 to 80 years) and tracking duration (2 to 22-year follow-up), taking into account sex differences, using generalized estimating equations.

Results

Two-year tracking coefficients are moderate to high for total volume of exercise across ages at baseline, ranging from .38 to .77 with a median of .57. Tracking coefficients tend to decrease as the distance to follow-up increases, down to a median of .38 for the 22-year tracking coefficients. The patterns of tracking were largely domain-

independent and were largely similar for solitary, competitive, non-competitive, externally and internally paced activities. With the exception of team-based activities, tracking was seen to increase as a function of baseline age. Cross-domain tracking did not favor any specific domain of exercise activity as the best predictor for total volume of exercise behavior and this was true at all baseline ages.

Conclusion

We conclude that exercise behavior is moderately to highly stable across the life span. In particular in adulthood, where the tracking of exercise mimics that of a classical behavioral trait like personality. This stability reinforces existing evidence that exercise habits are hard to change, but at the same time suggests that successful intervention leading to the adoption of exercise habits will tend to last.

Background

In our evolutionary history, the physically demanding environment caused genes in humans to be specifically selected to cope with high levels of physical activity. This resulted in human physiology in which most systems (e.g. metabolic and cardio-vascular) do not function optimally unless regularly engaged by sufficient physical activity (Booth et al., 2008). However, technological advancements, made since the industrial era, have continually decreased the necessity of physical activity, both in work- and home-environments. This has drastically reduced physical hardships of workers as well as the amount of disabilities caused by jobs that demand heavy labor, but unfortunately these beneficial aspects of lower levels of physical activity are mirrored by detrimental effects on the risk for chronic non-communicable diseases. Physical inactivity is considered a well-established risk factor for cardiovascular disease and cancer (Aune et al., 2015; Kyu et al., 2016) as well as for chronic psychiatric disorders (Schuch et al., 2016). The importance of intervention on physical activity has been stressed for many decades (Lewis et al., 2017; Morris et al., 1980).

Such interventions can be aimed at increasing total physical activity including means of transportation and work-related physical activity, or they can specifically target leisure time physical activity. A disadvantage of targeting total physical activity is that it can be highly dependent upon factors that are outside of a person's immediate control, such as the type of employment, and the mode of travel to and from work allowed by the distance between home and work locations. Workplace based exercise facilities are offered by some but not by all employers, and not all allow this to be done in office hours. From a public health perspective, increasing voluntary physical activity during leisure time may be a more viable target for behavioral intervention. Such interventions might be well served by building on the natural

tendency to participate in sports and exercise activities that is still evident in many people. Such tendencies are actively nursed by physical education classes in primary and secondary education and by public health policies supporting sports and exercise in youngsters by co-financing sports clubs and facilities. However, the extent to which this leads to the development of life-long habits of regular exercise behavior in leisure time remains to be established. This is unfortunate because the physiological and psychological benefits of regular physical activity do not seem to last if such activity is discontinued (Telama et al., 1996; Twisk et al., 2002).

A number of 'tracking' studies have been performed to identify patterns in the development of leisure time physical activity over the lifespan. Tracking is typically defined as the tendency of individuals to maintain their rank or position within a group over time (Telama, 2009) and tracking coefficients of sufficient magnitude allow a reliable prediction of future exercise behavior from the current behavior. Not surprisingly, most research on tracking of leisure time exercise behavior has been focused on tracking from childhood to adolescence (Aarnio et al., 2002; Francis et al., 2013; Pahkala et al., 2013; R. Tammelin et al., 2014), or on tracking in the transition from childhood/adolescence to adulthood (Kjønniksen et al., 2008; Parsons et al., 2006; Richards et al., 2007; T. Tammelin et al., 2003; Telama et al., 1996; Telama et al., 2005); although some studies have also looked at tracking in middle-aged and older cohorts (Anderssen et al., 1996; Barnett et al., 2008; Fortier et al., 2001; Parsons et al., 2006; Picavet et al., 2011; Tudor-Locke et al., 2008). A number of replicated results can be extracted from these studies (Malina, 2001; Telama, 2009). First, tracking of leisure time physical activity decreases with the duration of follow-up with tracking coefficients from baseline to follow-up of 0.37 to 0.71 on shorter intervals of three to nine years, but lower tracking when time intervals are longer than nine years, as low as 0.07 in some

cases of 31-year follow-ups (Parsons et al., 2006). Second, tracking from childhood to adolescence or from adolescence to adulthood is lower than the tracking across the longer phase of adulthood itself. Third, in keeping with the first two findings, lowest tracking is found from childhood to adulthood. Fourth, tracking seems to be influenced by sex, with lower coefficients found in women than in men (Telama, 2009).

A major shortcoming of all these previous studies is that they have treated leisure time exercise behavior as a uniform construct by summing all exercise and sports activities that the person engages in into a summary score (McMurray et al., 2003; Parsons et al., 2006) that either reflects the time spent on exercising, or an estimation of the total energy expended in exercise. The latter is done by multiplying frequency, duration, and an estimate of the intensity of the exercise activities usually expressed as a multiple of the basal metabolic rate (i.e., in MET hours weekly). Although the use of such a summary score is very valuable, it cannot detect more detailed patterns in the tracking of different types of exercise behavior. Ideally tracking coefficients would be analyzed for each sport or exercise activity individually (fitness, soccer, jogging, tennis, karate, etc.) but doing so would significantly reduce the number of available subjects per cell even in very large databases. Therefore, as a compromise, exercise activities could be divided into domains based on three dimensions (1) a team-based or individual nature, (2) a competitive or non-competitive nature, and (3) an externally paced or an internally paced nature. The third of these dimensions is based on the nature of the skills required in the exercise activities (Galligan, 2000; van der Mee et al., 2017). These domains were chosen because exercise behavior in these different domains of exercise behavior is likely to be differently sensitive to the many transitions and life-changing events experienced during the course of life. Their contribution to the overall volume of exercise and

sports behavior may likewise differ across time. Team-based exercise behavior for example is expected to track less well into adulthood when compared to solitary activities (Sallis et al., 1996; Telama et al., 2006), likewise participation in competitive exercise is expected to decline (da Silva, 1980). We believe personality and cognitive flexibility to be the major determinants of a preference for internally, or externally paced exercise (van der Mee et al., 2017) and thus, given the stability of personality, produce larger tracking in these domains. A better understanding of the development of life-long habits of regular exercise behavior in these various domains, and particularly the identification of 'vulnerable periods' with lowered tracking is directly needed. Supporting such a favorable development still proves a major challenge in many societies.

Using a large nationwide dataset (longitudinal N = 43,889), available at the Netherlands Twin Register (NTR), our aim is to assess the shortcomings mentioned above, allowing us to add significantly to the existing knowledge of the longitudinal tracking of exercise behavior in various domains. The NTR dataset spans a wide age range and includes children, adolescents and adults (8-82 years) in which exercise behavior has been assessed longitudinally in a very similar manner in biennial NTR survey waves. This unique resource allows us to analyze tracking-coefficients of up to a 22-year follow-up, to assess this long-term tracking with starting baseline ages in childhood, adolescence or adulthood, and to split all analyses by sex to compare tracking coefficients in males and females. Importantly, the information provided by the subjects allowed us to compute the tracking coefficients across total MET-minutes weekly but also separately for MET-minutes in the domains of team vs individual, competitive vs non-competitive, and externally paced vs internally paced exercise activities. We expected to find significant stability for the leisure time exercise behaviors in all domains with higher tracking

coefficients across shorter time spans, lower values in females than in males, and a gradual increase in tracking from childhood to adulthood.

Methods

Participants

The NTR is an ongoing research initiative that includes twins and their relatives. Subjects (both adults and children) are assessed every two to three years using extensive surveys. Subjects aged 14 and older fill out self-report surveys, and parents of children under 14 are asked to fill out surveys on the behavior of their children. In all cases the parent-rating of the mother is preferred and if this rating is not available, rating of the father is used (inter-rater correlations ranged from .74 to .85 depending on the age of the subject). Methods and procedures used for data collection in the NTR are described elsewhere (van Beijsterveldt et al., 2013; Willemsen et al., 2013). Subjects were included only if any measure of voluntary exercise behavior was available in at least two surveys (N = 43,889).

Voluntary Exercise Behavior

Overall physical activity has proven to be difficult to measure reliably through self-report (Prince et al., 2008), but the more salient regular exercise activities voluntarily done in leisure time are much more reliably recalled (de Moor et al., 2008; Stubbe et al., 2006). The current study focusses on voluntary regular exercise behavior in leisure time, collected by virtually the same survey methods in large numbers of twins, their parents, siblings, spouses and the children of twins and siblings (van Beijsterveldt et al., 2013; Willemsen et al., 2013). To assess the volume of voluntary exercise behavior we ask whether the participant exercised regularly in their leisure time ('Yes' or 'No'). If the participants (or their parental informant) responded affirmative, they were asked to indicate all of the exercise activities done regularly.

For each exercise activity reported, they noted the number of months per year, weekly frequency and average duration in minutes of the activity. Depending on the age of the participants (≤16 vs >16) Ainsworth's Compendium of Physical Activity (Ainsworth et al., 2011) or the adaptation for children by Ridley et al. (2008) was used to assign a Metabolic equivalent of task (MET) value to each exercise activity, reflecting its energy expenditure as a multiple of the basal energy expenditure (approximately 1 kcal/kg/hour) in an average subject engaged in that activity. For each participant, a total weekly MET value (METmin) was computed across all exercise activities by summing the products of the number of minutes spent weekly on each exercise activity and its MET value. Activities were only considered if that participant had engaged in them for at least three months during the past year, thereby excluding short-lived, non-regular activities, such as outdoor skiing in winter or beach sports in summer holidays. For children and adolescents, exercise during obligatory physical education classes at school was not included in the weekly MET value for voluntary exercise behavior.

Six additional variables were computed based on the specific type of exercise and sports activities. The first was the total volume spent on competitive exercise activities, i.e., activities where an individual, or team, has to compete against another individual, or team, for victory. If a subject's only voluntary exercise behavior was spent in non-competitive sports (e.g., cardio, muscle training) this was counted as zero METmin (as was done for all non-exercisers). Conversely all METmin spent on non-competitive activities were also summed into a second variable for total volume spent on non-competitive exercise activities. The third variable was the sum of all METmin spent on team-related activities, in which exercisers work together to achieve a common goal. Conversely, all of the METmin spent on solitary exercise (e.g., horse sports, aiming based sports or fighting sports) was summed

to a fourth variable. The fifth and sixth variable were based on the exercise being either externally paced (exercise activities where the environment controls the rate of performing the activity) or internally paced (exercise activities where the performer controls the rate at which the activity is performed).

Statistical Analyses

Sex-differences, age effects, age-by-sex interaction, and the quadratic age effect on exercise behavior were analyzed in a cross-sectional dataset containing the last non-missing observation of each subject. These effects were fitted using generalized equation estimation (GEE) models, in order to control for the nested (family) structure of the data. In the GEE model the correlation structure was set to be exchangeable as suggested for family data by previous work (Minică et al., 2015).

Because surveys were generally sent out every two years, ages were transformed to always be even for the tracking analyses. Distance from age at baseline to age at follow-up were therefore always multiples of 2 years, the longest follow-up duration being 22 years. The tracking of exercise behavior was computed for every combination of baseline age and follow-up duration (i.e. 2-year, 4-year, 6-year, .., 20-year, 22-year) where at least 50 complete combinations were available. GEE models were used to correct for family structure that was present in the NTR, as mentioned earlier. A GEE model was fitted with family as a random factor, and exercise at follow-up as dependent variable. Before the GEE slope-estimates were estimated, MET-scores at baseline and follow-up were normalized. This normalization allows us to consider the slope-estimates as tracking-coefficients, enabling a comparison of our results to results from previous publications where a correlation was used. The standardized slope estimates were plotted

in a heatmap that represents tracking as a function of the age at baseline and the follow-up duration.

To analyze the changes in exercise tracking across the lifespan, N-weighted linear regression was performed, regressing the tracking-coefficient on age at baseline for each follow-up duration. This regression was repeated for team-based, individual, competitive, non-competitive, and externally paced or internally paced MET-scores for all subjects together, as well as separately for males and females. To analyze the effects of sex on the tracking-coefficients this linear regression was also run including sex as a predictor in this regression, as well as an interaction effect between sex and age. Because this analysis is repeated over seven exercise domains a *p*-value threshold of .05/7 (0.007) will be considered for statistical significance.

Results

Means and standard deviations for METmin scores in 2-year age-bins are presented in an additional table (see Supplementary Table 2.1), for the entire sample and graphically displayed separately for males and females in Figure 2.1. An increase is seen from age 8 to age 16 for total METmin after which a rapid decline sets in that gradually slows down in mid adulthood. The major source of this biphasic pattern is participation in competitive, team-based, and often externally paced activities that is characteristic of Dutch children and adolescents and includes the most endorsed sports soccer and field hockey (Ministry of Sports Wellbeing and Health, 2016). In particular team participation drops sharply after age 16-18, reaching low levels (< 40 METmin weekly) after age 50. A contrasting pattern is seen for solitary non- competitive and internally paced activities which gradually increase to a peak level in early adulthood staying reasonably constant till age 50, after which they also decline. The most endorsed solitary exercise activities were running and weight/muscle training in fitness centers.

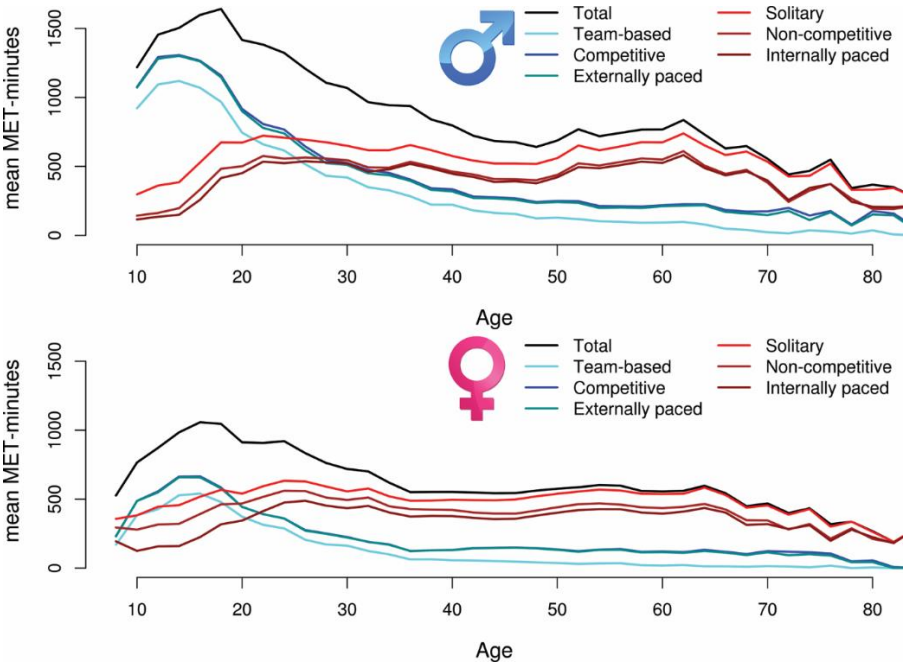


Figure 2.1. Mean volume of exercise in males (top panel), and females (bottom panel) over time, total volume and volume of exercise in all three dimensions (competitive & non-competitive, team & solitary, and externally & internally paced).

Formal testing of the effects of age and sex on METmin in a GEE model revealed significant linear decrease (β = -26.8, $p < 10^{-90}$) of exercise over age, a significant quadratic effect (β = 0.04, p = 0.0006), as well as a significant ($p < 10^{-90}$) effect of sex with males showing higher levels of exercise. Additionally, the GEE model revealed a significant ($p = 5.31 \times 10^{-65}$) interaction effect between age and sex demonstrating that the decrease of exercise over time is less pronounced in females when compared to males. The age and sex effects are consistent across total volume, competitive METmin, team-based METmin and externally paced METmin. For solitary, and non-competitive METmin females had higher values than males ($p < 10^{-90}$, and $p = 2.55 \times 10^{-31}$ respectively), while for internally paced METmin the effect of sex was not significant (p =

0.86). For solitary (β = 19.22, $p < 10^{-90}$), non-competitive (β = 22.67, $p < 10^{-90}$), and internally paced (β = 22.78, $p < 10^{-90}$) METmin a positive linear effect of age is observed, a negative quadratic effect ($p < 10^{-90}$ in all cases), and a negative age-by-sex interaction term in these domains ($p_{solitary} = 1.73 \times 10^{-22}$, $p_{non\text{-}competitive} = 4.12 \times 10^{-39}$, $p_{internally\ paced} = 1.98 \times 10^{-11}$) shows that the increase in exercise over time is less pronounced in females compared to males.

Correlations at various ages (8 to 80) of exercise activities in the various domains are presented in Figure 2.2. It is clear that internally paced, non-competitive, and solitary exercise correlate highly at all ages, and that competitive and externally paced exercise also correlate highly at all ages. For team-based exercise a more complex pattern is seen. It correlates highly with externally paced and competitive exercise in childhood and adolescence, but this correlation decreases strongly in adulthood. In parallel, the correlation of externally paced and competitive exercise with solitary exercise increases across adulthood.

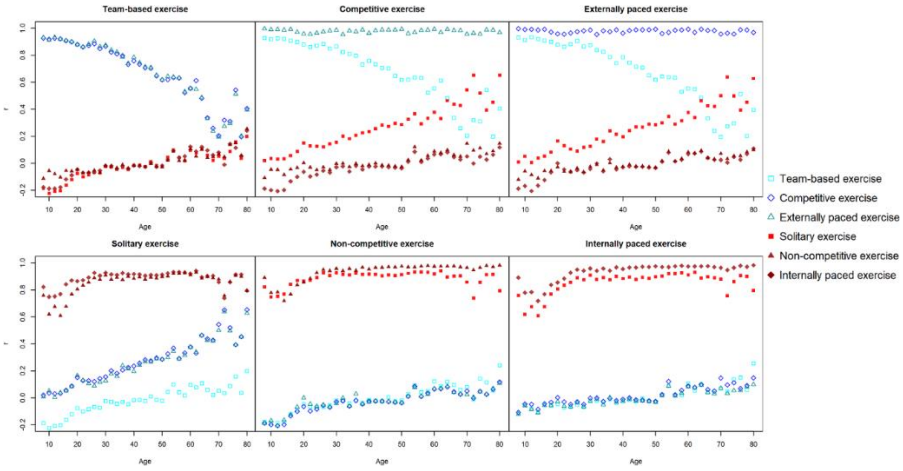


Figure 2.2. Correlations (r) between exercise domains as a function of age.

Tracking

All available tracking coefficients split by males and females for total METmin are displayed in the heatmap in Figure 2.3 and tracking coefficients for the various domains are displayed in an additional Figure (see Supplementary Figure 2.1). Insets show the sample size for each of the heatmap entries (all coefficients were based on N>50). A summary of 2-, 8-, 14-, and 20-year tracking coefficients of all domains is provided in an additional table (see Supplementary Table 2.2).

Two-year tracking coefficients are moderate to high for total METmin across all ages, ranging from .38 to .77 with a median of .57. Tracking coefficients tend to decrease as the distance to follow-up increases, ranging from a median of .57 at 2-year tracking down to a median of .38 for 22-year tracking coefficients. A similar pattern of decreased tracking with increased follow-up duration is seen for all other exercise domains (see Supplementary Figure 2.1).

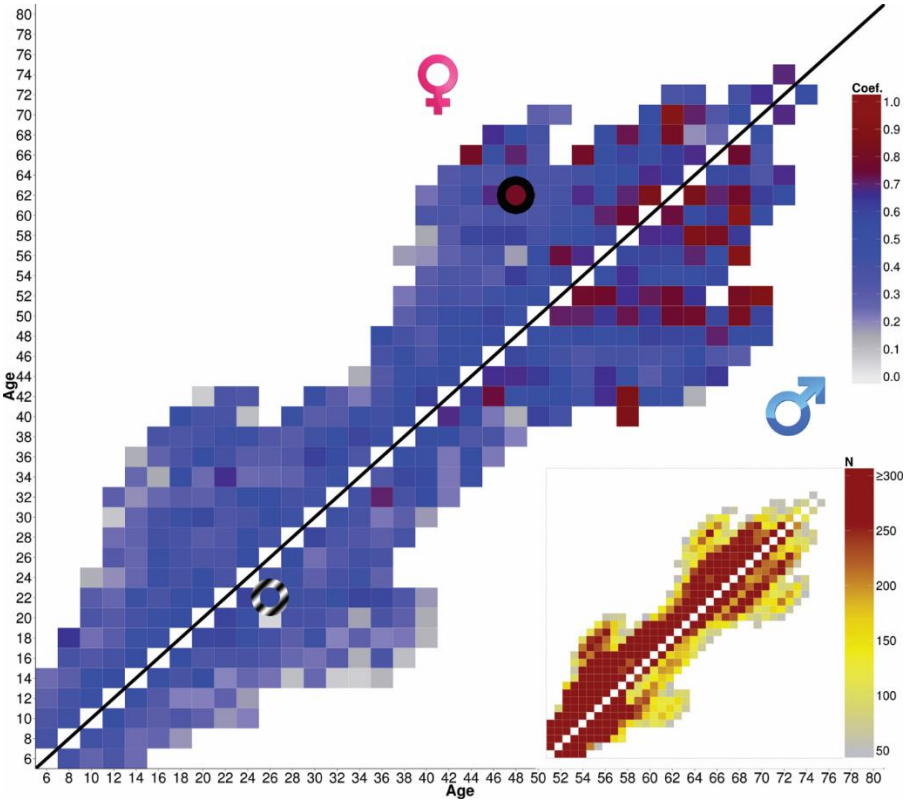


Figure 2.3. Tracking coefficient heatmap of total volume of exercise. The inset in the bottom right corner represents sample size (N) in each cell, cut off at N=300 for visual purposes. Colors in the main panel represent the value of the tracking coefficient (grey to blue to red from low to high), colors in the inset represent number of samples (grey to yellow to red from low to high). To demonstrate; the cell in the full circle represents the tracking coefficient (0.81) of age 48 at baseline, and age 62 at follow-up in females. The cell in the striped circle represents the tracking coefficient (0.57) of age 22 at baseline, and age 26 at follow-up in males.

Changes in exercise tracking across the life span

The tracking of exercise as expressed by total METmin shows a clear increase with age at baseline regardless of distance to follow-up. Figure 2.4 fits regression lines to the tracking-coefficients for follow-up durations 2, 8, 14 and 20 years as a function of the age at

baseline. The increase in tracking across the life span held for all other exercise subdomains (for plots see Supplementary Figures 2.2-2.6) with the single exception of team-based METmin where tracking coefficients often decrease with age (see Figure 2.4).

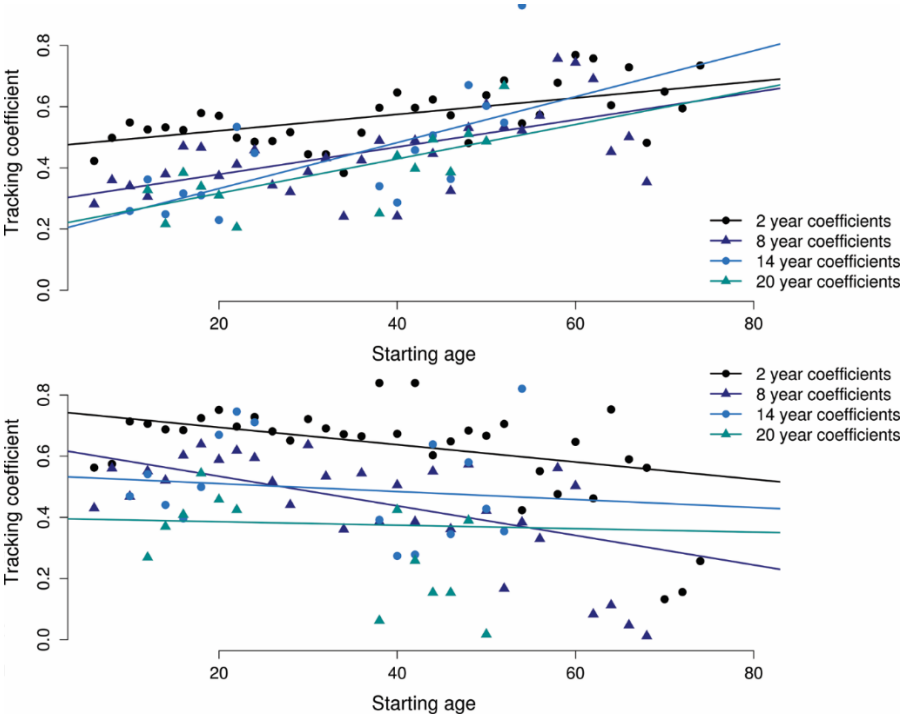


Figure 2.4. Tracking coefficients for total volume of exercise behavior (top panel), and team-based exercise (bottom panel) as a function of age at baseline.

It is important to observe that, regardless of which type of exercise is analyzed, there is no noticeable change in the patterns of tracking coefficients when the informant changes, that is when we change from parent-rated to self-rated voluntary exercise behavior (from 12 to 14 years) with the same instrument.

Sex differences in tracking

To assess sex-differences in the tracking-coefficients the previous linear regression model was expanded to include an effect of sex, and the interaction of sex and age. Significant main effects of sex were present in total METmin and competitive METmin where tracking was slightly higher in females compared to males. In total METmin, as well as solitary, non-competitive and internally paced METmin the interaction was also significant, suggesting that in these domains the increase of the tracking coefficient over time is less pronounced in females compared to males.

Cross-domain tracking

In an exploratory analysis we tested whether team-based, individual, competitive, non-competitive, externally paced or internally paced activities were the best predictor of the total volume of exercise at follow-up. For this, the above analyses were repeated with METmin in any of the domains at baseline, and the total amount of MET-minutes spent exercising at follow-up in the GEE models described earlier.

Heatmaps of tracking coefficients of any exercise domain at baseline are displayed in an additional figure (see Supplementary Figure 2.7), and the summary of 2-, 8-, 14-, and 20-year tracking coefficients of all domains is included in an additional table (see Supplementary Table 2.2). Overall, the cross-domain tracking coefficients were lower than the within-domain tracking coefficients. For team-based the median two-year tracking coefficients was 0.24, compared a median of 0.67 for within-domain tracking coefficients. For competitive, and non-competitive exercise the median two-year cross-domain coefficients were 0.35 and 0.41 compared to 0.67 and 0.50 within-domain respectively. For externally paced and internally paced exercise the median two-year cross-domain coefficients were 0.35 and 0.41 compared to 0.66 and 0.52 within-domain respectively.

Only for solitary based exercise were the median two-year cross-domain tracking coefficients (0.49) very comparable to that of within-domain tracking (0.52). The magnitude of between-domain tracking decreased as the distance to follow-up increased, just as was seen for within-domain tracking. This led to a decrease in the differences between within-domain and between domain tracking, because the 'surviving' exercise activity had become the main source of the total volume of exercise behavior. For team-based exercise, the median 22-year cross-domain tracking coefficient was 0.21 compared to 0.27 in within-domain coefficients. For solitary based exercise the median 22-year between-domain tracking coefficient was 0.17 compared to 0.21 within-domain. For competitive and non-competitive the median 22-year cross-domain coefficients were 0.24 and 0.19 compared to 0.28 and 0.20 within-domain respectively. For externally paced and internally paced exercise the median 22-year cross-domain tracking coefficients were 0.24 and 0.21 compared to 0.27 and 0.29 within-domain respectively

Similar to what was done for the within-domain coefficients, linear regression of the tracking coefficients on age at baseline was performed across different follow-up durations. Regressions for 2-, 8-, 14-, and 20-year coefficients for each domain are given in additional figures (see Supplementary Figures 2.8-2.13). These slopes show the pattern that was already obvious from Figure 2.1: the total volume of exercise behavior in METmin is best predicted by previous levels of competitive, team-based activities in childhood and adolescence, but much better by non-competitive, solitary, and internally paced exercise after age 30.

Discussion

The aim of this paper was to examine tracking of voluntary exercise behavior in leisure time over the entire lifespan. Exercise

activities were clustered in different domains reflecting the nature of the activity to be (1) competitive or non-competitive, (2) team or solitary, and (3) externally or internally paced. The prevalence of engaging in voluntary exercise behavior, especially competitive, team-based, and externally paced sports increases during childhood, to reach its peak during mid-adolescence, around age 16. After mid-adolescence the total volume of exercise behavior starts a decelerating downward trend which is largely explained by the decline in participation in competitive, team-based sports. The participation in non-competitive, solitary sports increases from childhood to late adolescence to remain relatively stable between 18 and 64 years (51% to 60% exercisers). After the age of 64 the participation in non-competitive solitary sports starts to decline as well, causing the total participation in regular exercise to drop down to 23.1% at age 82, at which point 100% of the exercise activities engaged in are solitary.

The results of the analysis of tracking-coefficients suggest that tracking of voluntary exercise behavior is moderate to high (.38 to .57) regardless of the exact domain of exercise activities analyzed. In concordance with the expectations from the extant literature, as briefly reviewed in the introduction, tracking coefficients decreased with increased duration to follow-up and were generally lower from childhood or adolescence into early adulthood compared to tracking across early to late adulthood. In deviation of the latter, the tracking coefficients of team-based exercise decreases with age. These findings are in line with previous findings (Sallis et al., 1996; Telama et al., 2006). Contrary to our expectation, competitive exercise did not demonstrate a decline in tracking as substantial as the decline observed in team-based exercise. Given the strong decrease in team sports after adolescence it becomes increasingly difficult to form teams of similar skill and motivation, and (2) may cause some teams to

fall apart and exercisers to quit due to a lack of likeminded team-members.

In spite of decreased tracking of team sports, our results show that the tracking coefficients of competitive and externally paced exercise increase, and that their correlation to solitary exercise also increases over time. This suggests that subjects switch from a team-based, externally paced, competitive exercise (such as football or field-hockey) to a solitary, externally paced competitive exercise (such as tennis or squash) later in life. In accordance with previous reports, we observed higher tracking coefficients in males for total volume of exercise, however our results suggest that when baseline age, and an age-sex interaction are taken into account it appears that in total volume, as well as solitary, non-competitive and internally paced exercise the effect of sex is dependent on age at baseline.

We show that long-term tracking from childhood to adulthood for leisure time exercise activities in all domains, albeit significant, is at most modest. This may not be surprising because this period contains some of the major transitional phases and life-changing events that can impact on the time available for leisure exercise, including entering high school, leaving home, taking up a job, and starting a family. Yet, the idea that childhood and adolescence should be specifically targeted to ensure life-long exercise habits is widespread, and motivates physical education classes in primary and secondary education and national, regional, and local policies to enable and encourage sports and exercise in youngsters. Telama (2009) formulated a number of mechanisms that could underlie tracking from childhood to adulthood, the carry-over value hypothesis, the ability and readiness hypothesis, the habit formation hypothesis, and the self-selection hypothesis.

The carry-over value hypothesis suggests that early adoption of specific exercise/motor skills acquired in childhood will lead adults

to participate in that same exercise. Splitting the analyses over the domains and comparing the results of within-domain tracking coefficients with cross-domain tracking coefficients across the lifespan allows us to address this hypothesis. The carry-over value would predict higher within-domain tracking compared to cross-domain tracking. The ability and readiness hypothesis suggests that subjects with high level of exercise in youth will also be among the higher exercisers in adulthood, regardless of the type of sport. The ability and readiness hypothesis would predict high tracking coefficients from any domain at baseline to total exercise at follow-up. In our data, the ability and carry-over value hypothesis finds stronger support than the ability and readiness hypothesis. Within-domain coefficients were higher than cross-domain tracking coefficients. However, some evidence for the ability and readiness hypothesis is found in late adulthood where the tracking coefficients of individual exercise at baseline and total exercise at follow-up approach those of within-domain total exercise. We caution that this may be a result of (nearly) all exercisers partaking in individual exercise at this stage of life.

The habit formation hypothesis suggests that simply repeating a behavior many times (independent of skill level) will form a habit. This hypothesis would, more strongly than the carry-over value hypothesis, predict higher within-domain than cross-domain tracking across the entire life span. Our results only partly support the habit formation hypothesis, because substantial exchange between exercise activities across the different domains is seen, although in late adulthood tracking of domain-specific exercise is very high. It may be that habits are more easily formed at later ages due to circumstances allowing for easier habit formation such as a steady job, whereas in childhood, adolescence and young adulthood skill level is more important and

there are many other factors impacting on exercise choice (puberty, education, career change, etc.).

Finally, the self-selection hypothesis states that subjects with a genetic predisposition for high exercise ability and/or exercise motivation will more often partake in exercise (Lightfoot et al., 2018). Using the twin-family structure of part of the data presented here, we have previously provided evidence for this hypothesis by showing substantial heritability of voluntary exercise behavior across the life span (de Moor et al., 2011; Huppertz et al., 2016). However, heritability has been established only for total volume of exercise. For exercise activities in the team-based/solitary, competitive/non-competitive and external/internal paced domains the self-selection hypothesis has not been addressed although a candidate gene study provides some indication of possible different genetic architectures for these domains (van der Mee et al., 2017).

A limitation of this study is the use of self-report which may suffer from both recall and social desirability biases. Self-report of participation in sports and leisure time exercise behavior has shown to be more reliable than self-report on other leisure time physical activities (Telama, 2009). Whether leisure time exercise behavior could be better quantified by using objective methods like accelerometers is not an easy question. When the focus is on energy expenditure in leisure time, more reliable estimates will likely derive from accelerometers than from self-report. Accelerometer-derived activity counts can, however, not be used to discriminate between competitive and non-competitive or team-based versus solitary activities, at least not without additional self-report on the content of the activities. Accelerometry, typically with a 7-day recording time, may also miss sports activities performed with a frequency that is less than weekly. So far, direct comparison between objective and recall methods have not revealed clear differences in tracking coefficients for leisure time

physical activity (Telama, 2009) and self-reported leisure time physical activity robustly predicted objectively measured moderate to vigorous activity across a 25 year follow-up period (Waller et al., 2018). Without discrediting recall and social desirability bias, the use of self-report instead of objective measurements may not be overly problematic for these specific behaviors.

Conclusions

In conclusion, voluntary exercise behavior in leisure time decreases over age, although this depends on sex, and the type of exercise. Team-based competitive exercise activities appear more prone to decay than solitary, non-competitive activities. Tracking coefficients show voluntary exercise behavior to be a moderate to highly stable behavioral tendency, although this too depends on the age and domain of exercise. Comparing domains over the lifespan non-competitive and internally paced exercise proved to be the most stable domains, in particular in late adulthood. Stability decreases as the distance to follow-up increases, resulting in less carry-over from exercise at young ages to late adulthood. Team-based exercise in particular appears to be a poor predictor of the total volume of exercise behavior in late adulthood. The varying results we found for age-effects, sex-effects, and the longitudinal tracking between the exercise domains substantiates the need to examine voluntary exercise behavior as a function of age and sex, and split the activities across the various domains in favor of using a single score summarizing across all domains. For exercise interventionists, our results signal that the glass is half full. The substantial stability of this important health behavior reinforces the existing evidence that exercise habits are hard to change, but at the same time suggests that any successful intervention that leads to the adoption of exercise habits stands a good chance to last.

Is Physical Activity Regulated by Genetics? Evidence From Studies in Humans

Published as van der Zee MD, Schutte N & de Geus, EJC (2019) Is physical activity regulated by genetics? Evidence From Studies in Humans. In: *Lightfoot et al (Ed) Routledge Handbook of Sport and Exercise Systems Genetics, Routledge, London, UK, pp 67-79 doi.org/10.4324/9781315146287*

Introduction

Physical activity (PA) is a broad concept containing a variety of different human behaviors that share the common denominator of expected beneficial effects on mental and somatic health (World Health Organization, 2013). Typically daily physical activities take up about 30% of total energy expenditure and reflect a mixture of obligatory activities related to transportation, work, or household chores and self-chosen physical active behaviors including sports and exercise activities in leisure time. Objective measurements of daily PA in studies conducted in Europe and the US have confirmed the existence of large individual differences in activity levels (Hagströmer et al., 2010; Hansen et al., 2012). These differences are observed in men and women alike, and persist after stratifying for age which is known to have a strong effect on average PA level (Rowland, 2016).

To increase the success of intervention on this important health behavior, much research has been devoted to the causes of the individual differences seen in PA in the general population. In this chapter, we focus on the role played by genetic factors. Broadly, the genetic research on PA phenotypes can be classified into (a) twin studies that partition the observed variance in PA phenotypes into environmental and genetic components based on well-established biometric models of human inheritance (Plomin et al., 2013), and (b) gene-finding studies that use either candidate gene based approaches based on the known biological role of proteins (e.g. in neurotransmission) or agnostic whole-genome searches using genome-wide microsatellite (linkage) or SNP markers to detect genomic loci harboring variants associated with PA. We will summarize the current evidence from each of these types of studies to address the question of whether PA is regulated by genetics.

Twin studies

Physical active behaviors appear to 'run in the family', for example, the chance of one family member being a regular exerciser increases the chance of all other family members to be, or to become, an exerciser. Familial aggregation of PA can be investigated by computing correlations among relatives such as siblings, parents and their offspring. However, siblings among each other, and parents and their offspring, share not just half of their genes, they also share a household, socio-economic status, the neighborhood, and various other aspects of belonging to the same family or living in the same neighborhood (the so-called shared environment), including parenting behaviors, family functioning, shared peers, etc.

Twin studies can separate the two mechanisms of familial aggregation by comparing the resemblance in monozygotic (MZ) or identical twins to the resemblance in dizygotic (DZ) or fraternal twins (Plomin et al., 2013). When twins are reared together the amount of this sharing of the family environment is the same for MZ and DZ twins. The important difference between MZ and DZ twins is that the former share most, if not all, of their genotypes, whereas the latter share on average only half of the genotypes segregating in that family. If the resemblance in PA within MZ pairs is larger than that in DZ pairs, this suggests that additive genetic factors (A) influence PA. If MZ resemblance is more than double as large it suggests the influence of non-additive (D) genetic factors. Additive genetic factors represent the sum of all linear effects of the genetic loci that influence the trait of interest. Non-additive factors include dominance and epistatic interaction effects. If, however, the resemblance in PA in DZ twins is more than half as large as it is in MZ twins, this points to shared environmental factors (C) as the cause of twin resemblance. Furthermore, the extent to which MZ twins do not resemble each other is ascribed to the unique environmental factors (E). These include all

person-specific experiences like differential jobs or lifestyles, accidents or other life events, and in childhood, differential treatment by the parents, going to different schools and having non-shared friends and peers. Measurement error will also be subsumed by the unique environmental factor.

We have previously reviewed all twin and family studies on the heritability of total PA and regular sports and exercise behaviors in childhood or adolescent samples (Schutte et al., 2018). In younger children, the shared environmental factors (C) explain the largest part of the variation in physical activity. However, the importance of these shared environmental factors decreases in adolescence and young adulthood, where genetic effects become the dominant factor explaining individual differences in both total PA and regular sports and exercise behaviors. A sample size weighted meta-analysis showed PA heritability estimates of 20% (95% CI: 13, 27) in children, 35% (95% CI: 17, 52) in early adolescents, and 53% (95% CI: 47, 59) in late-adolescents. This increase in heritability ran in parallel to a gradual decrease in the importance of shared environmental factors, which are the main cause of individual differences in physical activity and exercise behavior in young childhood, but then gradually become overwhelmed, by the importance of genetic factors during adolescence. Using longitudinal follow-up of twins form childhood to adolescence, we show that the increase in heritability, at least for voluntary exercise behaviors, was due to a strong increase in the genetic variance in the course of adolescence that was not paired to a similar increase in environmental variance (Huppertz et al., 2016). Whether the trend of increasing heritability is continued or curbed in young, middle or late adulthood remains to be established.

Sufficient studies in adult twins (> 18 years of age) have now accrued to allow a similar meta-analytic approach and to arrive at solid estimates of additive genetic and shared environmental variance for

PA phenotypes in adulthood. We searched publications in the English language on human subjects in PubMed and Web-of-Science from January 1980 to December 2017 using the keywords ('Physical Activity' OR Exercise OR Sports OR Lifestyle) AND (Genetic OR Genes OR Linkage OR QTL OR Twin OR Family OR Familial OR Heritability) AND "Humans"[MeSH terms]. From the 850 putative papers, title and abstract analysis was used to select only those publications reporting MZ and DZ/sib correlations (in at least 30 complete pairs) and/or estimates of (non-)additive genetic, shared and/or unique environmental variance components. Reference sections of selected papers were used to identify additional papers missed by these search terms. We then removed samples reporting on twins with a mean age lower than 18. Next, we removed partly overlapping datasets from the same twin cohorts. When the same phenotype was used in largely the same age group, we only used the study reporting the largest dataset, which typically would be the most recent study. A final set of 27 adult twin studies (Aaltonen et al., 2010, 2013; Carlsson et al., 2006; de Geus et al., 2003; de Moor, Posthuma, et al., 2007; de Moor, Stubbe, et al., 2007; de Moor et al., 2011; den Hoed et al., 2013; G. E. Duncan et al., 2015; Eriksson et al., 2006; Frederiksen & Christensen, 2003; Gielen et al., 2014; Heller et al., 1988; Huppertz et al., 2017; Huppertz et al., 2016; Huppertz, Bartels, Jansen, et al., 2014; Joosen et al., 2005; Karvinen et al., 2015; Kujala et al., 2002; McCaffery et al., 2009; Mustelin et al., 2012; Mustelin et al., 2011; Nelson et al., 2006; Simonen et al., 2004; Spinath et al., 2002; Stubbe et al., 2005; Vink et al., 2011) were selected by the above search criteria.

Physical activity phenotypes

In the final set of 27 adult twin studies, we encountered a large variation in measurement instruments and PA measures used. By far the largest common denominator was the use of survey-based

methods using subjective PA reporting on self or family members. In large population-based twin registries, surveys are often considered the only feasible strategy. However, subjective reporting of PA is vulnerable to distortion due to recall error and reporting biases (Hagstromer et al., 2010). It is particularly difficult to estimate both the duration and frequency of physical activities that are light to moderate in intensity, including common activities like walking and standing, or household activities (Bassett Jr et al., 2000). For total PA and light to moderate PA, objective measurement strategies using indirect calorimetry, the double labeled water method or movement sensors (Westerterp, 2013) are therefore superior.

When the focus is not on the detection of total PA, but when people are asked to report moderate to vigorous PA, specifically when confined to structured activities in leisure time, self-reporting seems to fare much better (Craig et al., 2003). The cognitive salience of such intensive physical activities is higher than light to moderate intense activities occurring as part of daily routine. Reliability of self-reporting further increases when a restriction is made to the reporting of regular exercise and sports behaviors in leisure time. In our own research, for instance, we have shown a high short-term test-retest reliability (de Moor & de Geus, 2012; Stubbe et al., 2007) as well as substantial tracking over longer periods of time for the weekly volume of voluntary exercise behavior (Huppertz et al., 2016).

Our meta-analysis was organized into the main domains of PA that we encountered in the literature: Total Physical Activity (TPA), Moderate to Vigorous Physical Activity (MVPA, including separate measures for moderate or vigorous PA where applicable), Leisure Time Physical Activity (LTPA), and Voluntary Exercise Behavior (VEB). The first two categories (TPA and MVPA) cover activities that are only partly under the individual's control, whereas the latter two (LTPA and VEB) cover PA that is largely voluntary in nature. We excluded PA measures

that deliberately excluded sports and exercise activities (e.g., Baecke's non-exercise related PA index), measures exclusively reporting on PA in the occupational setting (e.g. Baecke's Work/School index), and measures of daily activities of low intensity (e.g. accelerometer derived time spent in low LPA).

To convert the measured PA behaviors into an actual summary metric for use in the genetic analyses, again a breadth of different strategies has been used. Most studies adopt an estimation of the total energy expenditure by PA as their focus. Estimation of energy expenditure is usually based on published compendia (Ainsworth et al., 2011) that convert each activity into equivalents of the resting energy expenditure (1 MET, approximately 1 kcal/kg/hour). For VEB, most studies use a strategy akin to what we do in the Netherlands Twin Register, where we use an open format that allows reporting all sports and exercise activities that are performed for at least three months a year and then record duration, frequency and intensity of each reported activity. Typically, activities are censored that do not reach a minimal threshold of intensity like fishing or chess. Exercise related to swimming, sailing, or skiing that are restricted to the annual holidays is discarded. Physical Education classes are also discarded as they are poorly standardized and the activities are not voluntary. Energy expenditure in all leisure time VEB is then summed across all valid exercise activities in a weekly METhour score by taking the sum of the products of their weekly frequency, average duration, and MET score.

Another major difference across studies is the type of scale used to quantify PA. Continuous interval scales seem most optimal in terms of statistical power but quite often either ordinal categories (e.g., tertiles of METhours weekly; frequency of LTPA as once per month, once per week, every day) or even nominal dichotomies (YES/NO regular exerciser; YES/NO adheres to PA guidelines) were used. This use of dichotomies and categories is mostly inspired by the frequency

distribution of some of the LTPA and VEB phenotypes. For instance, moderate to vigorous PA in leisure time is very skewed. Likewise, many individuals do not engage in voluntary exercise behavior (zero score) and even in those who do the amount of exercise can still show a skewed distribution. Often no transformation is available that converts such mixed distributions to a normal distribution. For these phenotypes, therefore, the use of ordinal categories or a dichotomy can be meaningful. A liability threshold model can then be used in the genetic analysis to recapture the normal distribution of the latent 'liability to be a (vigorous) exerciser'.

Meta-analyses

To assess heritability and influence of common environmental factors in the four PA categories mentioned previously meta-analysis on A and C estimates was conducted across twin studies. Within the 27 study database, the A, C estimates and their standard errors were often reported separately for males and females. However, if the A, C estimates were reported only for males and females combined, we assigned the same estimates to the male and female part of the sample, adjusting the N to reflect the number of male or female twins. Most studies used age and sex as covariates, but additional covariates were sometimes used too, including socioeconomic status, body-mass index, or fitness. When multiple estimates were generated for different covariate compositions, we opted to use the estimates only correcting for age and sex effects on mean PA level. In these studies, structural equation based variance decomposition modeling was by far the most used analytic strategy. Unless stated otherwise, the results from the most parsimonious structural equation model were used in the meta-analysis. This model was most often an AE or ACE model. If the AE model was used, C was set to zero. No studies reported a model with non-additivity (D).

For each of the four PA phenotypes, the estimates for A, C were computed in a sample-size weighted meta-analysis across all available studies for males and females separately. Inverse variance weighting was not possible because not all studies provided either standard errors of the estimates or 95% confidence intervals from which the standard errors could be approximated. Forest plots in Figure 3.1 and Figure 3.2 present the characteristics and the A and C estimates per study, and the meta-analytic results per PA phenotype. Study characteristics are: country, whether males and females were combined in A and C estimation, whether PA measurement was through surveys (SUBJ) or experimental (OBJ), type of scale used (DIchotomy, CATegorical, or CONtinuous), mean age, and sex-specific sample size. The estimates for A and C are listed in the Forest plots as a function of mean sample age.

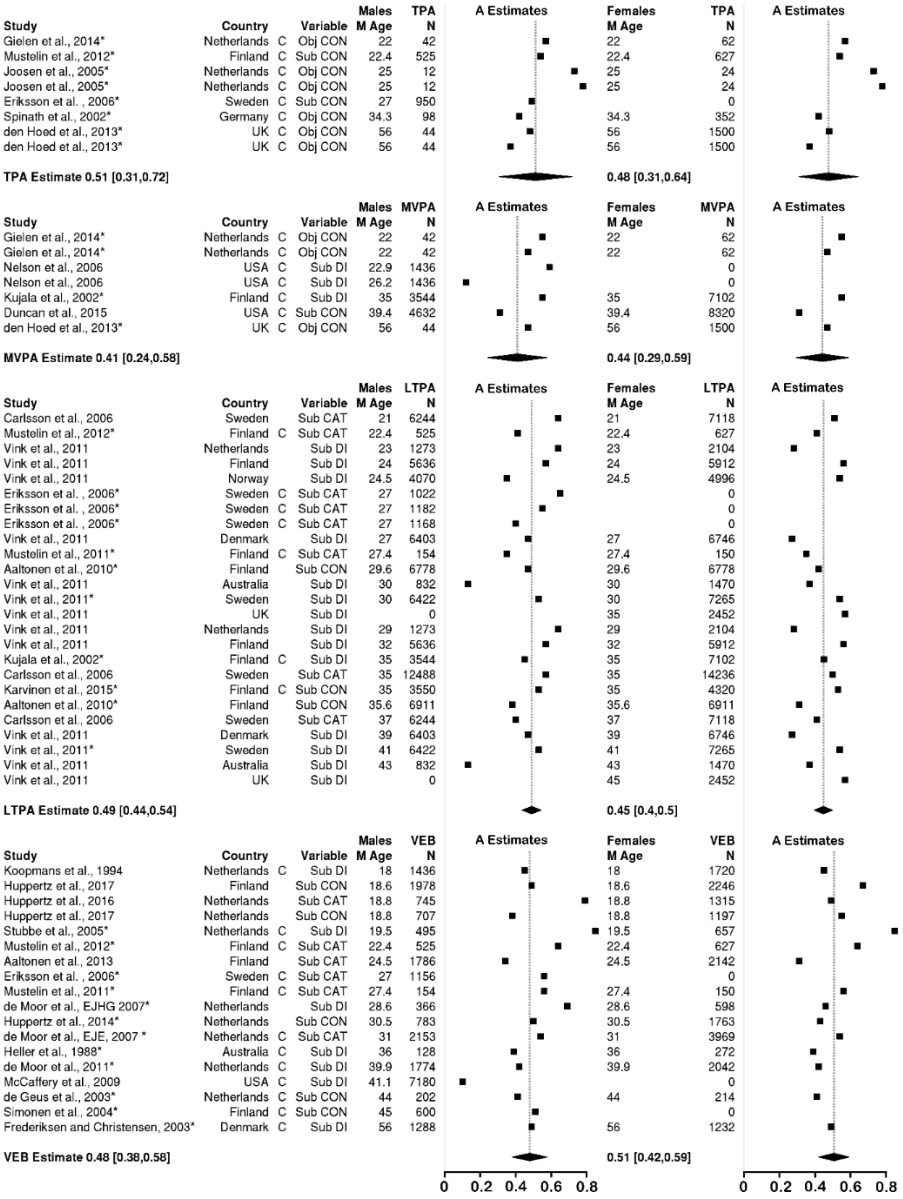


*Figure 3.1. Study-specific and meta-analytic estimates of heritability (A) of total physical activity (TPA), Moderate to vigorous activity (MVPA), Leisure Time Physical activity (LTPA), and voluntary exercise behavior (VEB), for males and females. Estimates include results from subjective (Sub), and objective (Obj), continuous (CON), categorical (CAT), and dichotomous (DI) variables. *; Studies where an AE model was fitted. C; Estimates from a combined male and female sample.*

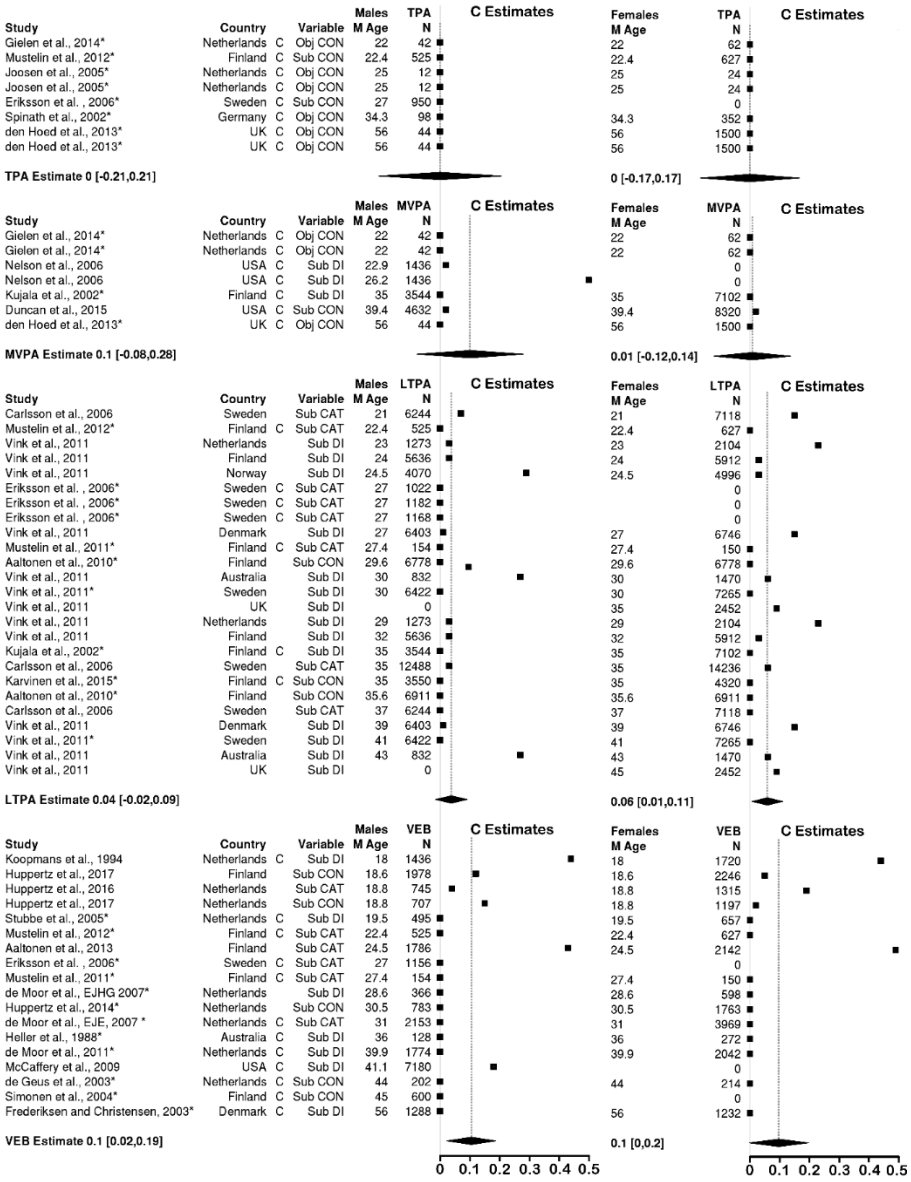


*Figure 3.2. Study-specific and meta-analytic estimates of shared environmental contribution (C) to the variance in total physical activity (TPA), Moderate to vigorous activity (MVPA), Leisure Time Physical activity (LTPA), and voluntary exercise behavior (VEB), for males and females. Estimates include results from subjective (Sub), and objective (Obj), continuous (CON), categorical (CAT), and dichotomous (DI) variables. *; Studies where an AE model was fitted. C; Estimates from a combined male and female sample.*

Results for Total Physical Activity

For the meta-analysis of TPA (see Figure 3.1 and Figure 3.2) a total of 6 studies using 8 phenotypes was available, most using an objective measure to quantify PA, and all combining males and females. The meta-analytic heritability estimate was 48% for females and 51% for males. Largest heritability estimates were found by Joosen et al. (2005) who used the most reliable methods, namely the doubly labeled water method and more than 14 days of accelerometer recording, albeit in the smallest sample. The two survey methods using the Baecke questionnaire yielded results comparable to the objective methods. No evidence for common environmental effects was found.

Results for Moderate to Vigorous Physical Activity

For the meta-analysis of MVPA (see Figure 3.1 and Figure 3.2) a total of 5 studies using 7 phenotypes was available, all combining males and females. The meta-analytic heritability estimate was 44% for females and 41% for males. Comparable heritability was found for objective and subjective measures. Somewhat deviant results were found in the 26 year old Add Health study participants (Nelson et al., 2006). Using a 7-day recall questionnaire to obtain a dichotomy of MVPA frequency that indicated whether participants met national PA recommendations (> 5 bouts/wk), lower heritability was paired to the only evidence for a shared household effect on MVPA (50%).

Results for Leisure Time Physical Activity

We used 8 twin studies for the meta-analysis of LTPA (see Figure 3.1 and Figure 3.2) reporting on more than 200,000 twins from 7 countries. The meta-analytic heritability estimate was 49% for males and 45% for females with narrow confidence intervals (95% CI upper – lower ≈ 10%) reflecting little heterogeneity. LTPA was only available

from self-report measures in these very large epidemiological samples. In two countries, Norway and Australia, evidence for common environmental impact on LTPA was found (~28%). Comparing sample sizes, the absence of large C in other twin studies cannot be attributed to low power.

Results for Voluntary Exercise Behavior.

VEB was also measured exclusively by survey or interview measures. A total of 15 different studies using 14 unique samples/measures yielded a meta-analytic heritability estimate of 48% for males and 51% for females (see Figure 3.1) with mild heterogeneity (95% CI upper – lower ≈ 20%). Some evidence for common environmental effects were found, mostly driven by the younger samples (~10%, see Figure 3.2) with a particularly strong C effect in the young adult Finnish twins (43% males, 49% females).

Gene-finding studies

The above meta-analyses show that genetic variants play a key role in adult PA, whether it is total daily PA as measured with objective means (e.g., accelerometers), or self-reported voluntary sports and exercise activities in leisure time. We now turn to the gene finding studies aiming to detect the actual 'PA genes'. For this, we have to use a narrative approach because not enough replication is currently available for any single locus to consider meta-analysis. The majority of gene finding studies on TPA, MVPA, LTPA, and VEB have used a candidate gene approach. The use of this approach was predicated on the expectation that individual genetic variants would be associated with complex behavioral traits at a magnitude that would be small but detectable with hundreds, or a few thousand individuals. The advent of large international consortia performing meta-analyses on genome-wide association results for many complex behavioral traits in samples

as large as hundreds of thousands of participants has found this expectation to be untenable. The risk contributed by any single variant is tiny rather than just small, with only an increase of ~0.05 standard deviation per risk allele at best, necessitating much larger sample sizes to be able to detect them (Visscher et al., 2017). Based on these types of concerns regarding candidate gene studies using small samples, we focus here mainly on the data-driven genome-wide approach in linkage and genome-wide association (GWA) studies.

Linkage

The three genome-wide linkage studies (G. Cai et al., 2006; de Moor, Posthuma, et al., 2007; Simonen, Rankinen, Perusse, et al., 2003) performed to date have only produced suggestive hits, in keeping with their modest sample sizes (767 < N < 1120). In the Quebec family study, Simonen, Rankinen, Perusse, et al. (2003) reported linkage for TPA (13q22-q31), MVPA (4q28.2, 7p11.2, 9q31.1, 13q22-q31), and VEB (11p15 and 15q13.3). Interestingly, the latter 15q linkage region contains the *GABRG3* gene in which a SNP (rs8036270) was found to be significantly associated with VEB in GWA study on a combined sample of 2,622 Dutch and European American middle-aged adults (de Moor et al., 2009). Significant association with *GABRG3* was replicated for LTPA in older adults in a GWA study in 10684 European and 11093 African Americans (Lin et al., 2018) although with different SNPs (*rs72707657, rs12438610, rs12902711, rs12595253*). The *GABRG3* gene may be involved in the aversive effects of exercise-induced fatigue because high expression levels of the *GABRG3* gene was found after a bout of exhaustive exercise (Kawai et al., 2007). However, a GWA study in 13,980 Japanese older adults did not find significant association with any SNPs in the *GABGR3* gene region and the gene also did not surface in the other two GWA studies.

In the Viva La Familia study (G. Cai et al., 2006) physical inactivity, recorded as the percentage time in sedentary activity, significantly mapped to markers *D18S1102-D18S64* in the chromosome 18q21 region where the *MC4R* gene resides. An additional suggestive linkage signal for TPA was detected in the same region. *MCR4* had been implicated in smaller scaled candidate gene studies (Cole et al., 2010; Loos et al., 2005) but it failed to replicate in any of the five GWA studies, two of which explicitly tested it as a candidate gene (i.e., with a more liberal p-value). In the Viva La Familia study a further linkage signal was found for MVPA in the region at 9q31.1, which harbors the *RN7SK* gene. A SNP (*rs7023003*) near this gene produced a suggestive ($p < 10^{-5}$) association with TPA in 8,454 Korean participants (Kim et al., 2014). However, the same *RN7SK* SNP was explicitly but unsuccessfully tested for replication in Japanese participants (Hara et al., 2018) and the gene also did not surface in the other two GWA studies.

Finally, in a sample from the Netherlands Twin Register (de Moor, Posthuma, et al., 2007), a suggestive linkage with exercise participation was found in all subjects on chromosome 19p13.3 (LOD 2.18). A SNP (*rs12462609*) in this region, located in the *CACNA1A* gene, was associated with vigorous PA in 261,055 older adult participants in the UK Biobank (Klimentidis et al., 2018), and with TPA in 8,454 older adult Koreans (Kim et al., 2014) whereas another SNP (*rs111901094*) in *GATAD2A* in the 19p13 linkage region was associated with VEB in the UKB participants (Klimentidis et al., 2018). The 19p13 region did not generate a significant or suggestive signal in the other two GWA studies (Hara et al., 2018; Lin et al., 2018).

Genome wide association

Five genome-wide association (GWA) studies (de Moor et al., 2009; Hara et al., 2018; Kim et al., 2014; Klimentidis et al., 2018; Lin et al., 2018) conducted for exercise behavior or related PA phenotypes were

available at the time of writing. The three smallest GWA studies (2632 < N < 11093) did not detect significant associations after the required stringent correction for the multiple testing burden that is inherent in the agnostic genome-wide approach (de Moor et al., 2009; Kim et al., 2014; Lin et al., 2018). However, some genes listed as receiving suggestive evidence based on more lenient thresholds ($p < 10^{-5}$) showed interest-generating patterns of replication across multiple studies. For example, a suggestive association of VEB was found in Dutch and American participants of European descent (de Moor et al., 2009) for an eQTL (*rs12612420*) that reflects the expression of the *DNAPTP6* gene. This association was replicated ($p = 0.0092$) in 16026 Japanese participants (Hara et al., 2018).

Suggestive evidence for an association between VEB and the *PAPSS2* gene in Dutch and American participants (de Moor et al., 2009) was replicated in 11,903 African-American participants (Lin et al., 2018) for LTPA, although for different SNPs (*rs10887741* vs *rs1819162*). The *PAPSS2* gene has been linked to maximal exercise capacity (Rico-Sanz et al., 2004) which may be a factor influencing the motivation to engage in voluntary exercise (de Geus & de Moor, 2011). Explicit testing of an association between VEB and *PAPSS2* in 13,980 Japanese participants, however, did not produce a significant replication (Hara et al., 2018). Suggestive evidence for an association with LTPA was found for *rs116550874* at 1p36.23, and *rs3792874, rs3792877, rs3792878, rs79173796* at 5q31.1 ($p < 8.61\times10^{-7}$) in 11093 African Americans and for *rs28524846* at 14q23.3 ($p = 1.30\times10^{-6}$) in 10684 European Americans (Lin et al., 2018). The authors point to the *ENO1, SLC22A5, PDLIM4* genes as potential sources of the association signal at 1p36. *ENO1* encodes a glycolytic enzyme and the integral membrane protein SLC22A5 is associated with skeletal myopathy, whereas the PDLIM4 protein is involved in the pathway of actin cytoskeleton remodeling, bone and skeletal muscle development. The *rs28524846* SNP at 14q23 is an eQTL for *MPP5* and

APT6V1D in nerve tissue, with the MPP5 protein known to regulate myelinating Schwann cells and *ATP6V1D* related to synaptic vesicle cycle and ATPase activity.

Encouragingly, the two largest GWA studies to date did successfully detect genome-wide significant associations with PA. In a Japanese sample (N=16016), Hara et al. (2018) identified an association (p_{meta} = 2.2×10^{-9}) between leisure-time exercise behavior and rs10252228, a SNP located in the intergenic region between *NPSR1* and *DPY19L1* genes. Hara et al. (2018) point out that the product of the *NPSR1* and that of the *DNAPTP6* gene previously found by the GWA of de Moor et al. (2009) are both involved in pulmonary function. Variants in *NPSR1* have been shown to be associated with asthma-related phenotypes, and *DNAPTP6* could be involved in bronchodilator response via the down-regulation of ß2-adrenergic receptors. Impairments of pulmonary function could be a barrier to the adoption of (vigorous) PA.

In the UKB (N = 380,492), Klimentidis et al. (2018) detected genome-wide significant genetic associations with touchscreen-survey based measures of MVPA. After applying corrections for work-related PA and an indicator of socioeconomic status, associations with MVPA were found at *rs429358* (*APOE*), *rs169504* (*PBX2*), *rs4129572* (*EXOC4*), *rs3094622* (*RPP21*), *rs181220614* (*ARHGEF26-AS1*), *rs149943* (*ZNF165*), and *rs2988004* (*PAX5*). Intriguingly, the Alzheimer disease risk allele (E4) of the *APOE* gene was associated with higher levels of MVPA. That *APOE* e4 carriers have more MVPA could be in keeping with the idea that physiological determinants of the ability to perform (intense and/or prolonged) PA lead to higher levels of PA. As noted by Klimentidis et al. (2018), *APOE* e4 carriers had a more favorable response to exercise, for instance as suggested by a study showing larger aerobic fitness in response to training (Thompson et al., 2004). This is in keeping with the theoretical notion that individuals who have

higher exercise ability and/or trainability will find it easier to adopt regular exercise as a lifestyle (de Geus & de Moor, 2008, 2011; Lightfoot et al., 2018).

From the UKB self-reports (Klimentidis et al., 2018) two dichotomous measures of more vigorous PA were defined: VPA by classifying participants as engaged in vigorous PA if they spend more than 25 minutes on activities "that make you sweat or breathe hard" for three or more days a week and VEB by classifying participants as regular vigorous exercisers if they spend more than 15 minutes on two or more days a week doing strenuous sports or other exercises. Significant genetic association for VPA was found at *rs1248860* (*CADM2*), *rs2764261* (*FOXO3*), *rs3781411* (*CTBP2*), *rs12707131* (*EXOC4*), and *rs328919* (*DPY19L1*) and for VEB at *rs62253088* (*CADM2*), *rs166840* (*AKAP10*), *rs10946808* (*HIST1H1D*), *rs75930676* (*SIPA1L1*), and *rs4865656* (*LOC642366*). In addition, accelerometer data were available in just over 91,000 participants. From up to seven days of accelerometer wear overall acceleration was obtained as a measure of TPA and the fraction of accelerations > 425 m*g* as a measure of VPA, yielding significant associations at *rs55657917* (*CRHR1*) and *rs185829646* (*ANKRD22*) for TPA and *rs743580* (*PML*) and *rs6433478* (*CIR1*) for VPA.

The *CADM2* gene, primarily expressed in the brain, surfaced in the UKB study as a gene influencing vigorous sports and exercise related PA. Previously this gene had been linked to risk-taking behavior and extraversion (Boutwell et al., 2017) as well as executive function (Ibrahim-Verbaas et al., 2016). This shows a remarkable parallel to results reported in the largest candidate study to date by van der Mee et al. (2017) where a polygenic dopaminergic risk score that summed the increaser alleles in *COMT* and *DAT1* for *both* executive function and reward sensitivity was associated with the volume of externally paced sports and exercise activities. Possibly, *CADM2*, like

these dopaminergic genes operates through a double whammy of increased reward value of exercise and increased sports skills.

Conclusions

Although the gradual rise in heritability of PA seen in adolescence is not continued in adulthood, it is clear that genetic factors remain a major contributor to individual differences in adult physical activity behaviors. In the reports on over 283,904 adult twins we find that about half of the variance in the four PA phenotypes can be explained by genetic factors (Males 48% CI 44%-52%; Females 47% CI 44%-50%). The substantial contribution of the shared environment (C) to childhood PA seems to have largely dissipated in adulthood (Males & Females 6% CI: 3%-10%). There is generally a good correspondence in estimates across individual studies even when they use samples from different countries or measures. For TPA and MVPA, objective measurement showed slightly higher genetic estimates than subjective measurement, but the differences are not striking. Major sex differences in genetic architecture of LTPA and VEB seemed to be absent; for TPA and MVPA estimates the verdict is still out as analyses were mostly based on combined male/female samples.

The extant studies using a candidate gene, linkage or whole-genome association approach have yielded a number of genes and variants worth careful monitoring in future gene finding efforts. Previously, three biological themes have emerged as a potential source for "Physical Activity Genes" from theory: (1) the brain circuitry related to motivational and affective aspects of PA, (2) the brain circuitry involved in the maintenance of energy intake/expenditure balance, and (3) the physiological determinants of the ability to perform (intense and/or prolonged) PA, ideally at an above average level (de Geus & de Moor, 2008, 2011; Lightfoot et al., 2018). Many of the variants reviewed above do seem to fit these theoretical notions, but we issue

a note of caution that this may partly reflect our deep desire (and uncanny ability) to make sense of data by 'reasoning towards' our theoretical models. Rigorous replication followed by experimental validation (in animal models) is direly needed. However, the experience from many other complex behavioral traits allows us to end upbeat: once very large samples are amassed in international meta-analytic consortia, we can confidently expect large progress in our understanding of how physical activity is regulated by genetics.

Genetic Influences on Regular Exercise Behavior

Published as van der Zee MD, Schutte N, De Moor, MHM & de Geus, EJC (2019). Genetic influences on regular exercise behavior. In: *Rippe, JM. (Ed) Lifestyle Medicine 3rd edition.* Taylor & Francis, LLC, Boca Raton, FL, USA, pp 1235-1249 doi.org/10.1201/9781315201108

Introduction

Regular engagement in physical activity lowers the risk for morbidity and mortality through cardiovascular disease (CVD), type 2 diabetes, and cancer (Aune et al., 2015; Kyu et al., 2016; L. Liu et al., 2016; Pandey et al., 2015). Despite these well documented benefits of physical activity, a large proportion of adults worldwide does not engage in such activities on a regular basis (Troiano et al., 2008). As a consequence, a physically inactive lifestyle remains a major threat to health in modern-day society, with substantial economic costs (Ding et al., 2016). This is reflected in public health recommendations worldwide, which unanimously include encouragement to a more active lifestyle (Brown et al., 2012; "Physical Activity Guidelines for Americans," ; World Health Organization, 2013).

Correlates and determinants of a physically active lifestyle have been studied intensively over the past three decades in an attempt to define optimal targets for intervention. Over 30 years ago, Dishman and colleagues stated that "one barrier to developing effective methods to encourage physical activity (...) is the lack of knowledge of the determinants of regular physical activity" (Dishman et al., 1985). Since then, researchers have identified, in cross-sectional or longitudinal studies, numerous potential determinants that contribute to the adoption and maintenance of physical activity. Physical activity encompasses a broad domain of activities at work, at home, during transportation, and in leisure time. In this chapter, the emphasis is on voluntary exercise activities in leisure time such as jogging, swimming, exercising at fitness clubs, and participation in recreational or competitive team and individual sports. Such voluntary exercise activities are often specifically aimed at improving skills, fitness, and health and are rapidly becoming the major source of physical activity of moderate-to-vigorous intensity in many industrialized countries.

The bulk of studies devoted to the determinants of voluntary exercise behavior have attempted to explain low exercise prevalence in terms of social and environmental barriers. These include poor access to facilities (Matson-Koffman et al., 2005), low socioeconomic status (Haase et al., 2004; Varo et al., 2003), non-Caucasian race (Hasson et al., 2017), high job strain (Payne et al., 2005), subjective "lack of time" (Sherwood & Jeffery, 2000), low belief in health benefits (Haase et al., 2004), and low social support by family, peers, or colleagues (Sherwood & Jeffery, 2000). Although these social factors are important, they explain only a small part of the individual differences in voluntary exercise behavior. In this chapter, we review the evidence from twin and family studies that genetic factors play an important role too.

Definitions of voluntary exercise behavior

Heritability estimates of behavioral phenotypes are sensitive to the exact definition of the phenotype. We will therefore first give an overview of how voluntary exercise behavior has been defined in the studies used in this review.

The four important characteristics of the studies presented in Tables 4.1 through 4.3 are that the reported exercise activities were performed regularly, voluntarily, in leisure time, and in the context of fitness or sports. We stress that this definition of voluntary exercise behavior is different from total physical activity, even when a restriction is made to leisure-time physical activity (LTPA) or more narrowly to LTPA that is of moderate-to-vigorous intensity. Such LTPA measures will encompass almost all voluntary exercise behavior but may additionally include gardening, dancing, or active transportation by walking or bicycling to school/work, none of which are considered to fall under voluntary exercise behavior in our definition. We further exclude obligatory exercise activities like physical education (PE)

classes or job-related sports and exercise (army boot camps; fireman training) and activities that are done infrequently like sailing, surfing, or skiing during a limited number of days per year. We note that our definition of voluntary exercise behavior also excludes accelerometer/step counter based physical activity, unless an ancillary diary or recall method is used. Without additional self-reporting, these activity counts cannot distinguish between voluntary exercise activities or other forms of moderate-to-vigorous activity (e.g., vacuum cleaning, bicycling, stair climbing).

Almost all studies on regular voluntary leisure-time sports and exercise activities as defined above have used self-reports, but the methods of assessment of these activities have differed across studies. Some studies used surveys, whereas others used interviews. Exercise behavior was assessed with only a single "YES/NO participation" question (Boomsma et al., 1989) or by presenting subjects with a list of specific exercise activities (Martinez-Gonzalez et al., 2001; van der Aa et al., 2010), whereas still others allowed all exercise activities to be reported in a free format (de Moor et al., 2006). The quantitative metrics for exercise behavior have differed accordingly. Some studies use a categorical scale classifying participants in a dichotomy of "regular exercisers" and non-exercisers, sometimes adding an additional level of moderate exercisers. These studies have used different thresholds for the minimum frequency and the minimum intensity of the exercise activities to separate regular exercisers from non-exercisers, making it somewhat difficult to either pool or compare the prevalence of regular exercisers across studies.

To avoid confusion by these differing thresholds and more optimally use the variation in the volume of exercise behavior among exercisers themselves, voluntary exercise behavior should preferably be expressed as the energy expenditure in exercise activities across a fixed time interval (e.g., weekly). This can be done by multiplying

frequency, estimated intensity and duration for each activity, and summing across all reported exercise activities. Unfortunately, the type, duration, frequency, and intensity of exercise activities have not always been assessed in great detail in older studies.

Prevalence of voluntary exercise behavior

In the Netherlands Twin Register, data on exercise behavior has been collected by virtually the same methods through surveys in large numbers of twins, their parents, siblings, spouses, and the children of twins and siblings (van Beijsterveldt et al., 2013; Willemsen et al., 2013). To assess the volume of voluntary exercise behavior, we ask whether the participant exercised regularly in their leisure time ("Yes" or "No"). If the participants (or their parental informant) responded affirmative, they were asked to indicate all of the exercise activities done regularly. For each exercise activity reported, they noted down the number of months per year, weekly frequency, and average duration in minutes of the activity. Depending on the age of the participants (≤16 vs >16), Ainsworth's Compendium of Physical Activity (Ainsworth et al., 2011) or the adaptation for children by Ridley et al. (2008) was used to assign a Metabolic equivalent of task (MET) value to each exercise activity, reflecting its energy expenditure as a multiple of the basal energy expenditure (approximately 1 kcal/kg/ hour) in an average subject engaged in that activity. For each participant, a total weekly MET value (METhr) was computed across all exercise activities by summing the products of the number of hours spent weekly on each exercise activity and its MET value. Activities were only scored if that participant had engaged in them for at least three months during the past year. For children and adolescents, exercise during obligatory PE classes at school was not included in the weekly MET value for voluntary exercise behavior. The six-months test–retest reliability of the weekly MET value was found to be .91 (Stubbe et al., 2007), and twins proved fully

comparable to non-twins with regard to this definition of exercise behavior (de Moor et al., 2006).

Up until 2016, exercise data was collected in 78,524 participants (54.4% females) with a mean age of 31.5 (SD =17.9). A total of 67.0% of these participants engaged in regular voluntary exercise activities (N = 52,603), whereas 33.0% were non-exercisers. When taking into account the frequency and intensity of the exercise activities in METs, little over half of the population (62.6%) engages in moderate to vigorous exercise activities (4 MET or more) for at least 60 minutes per week, and only 12.9% can be considered vigorous exercisers, that is, engaging in vigorous exercise activities (7 MET or more) for at least 300 minutes per week. In Table 4.1 we illustrate the effects of two of the best-known determinants, age and sex, on the prevalence of voluntary exercise behavior. We furthermore report exercise behavior separately for the twins, their singleton siblings, and, in the older age groups, for the parents and spouses.

		MALES					FEMALES				
		Weekly METhours					Weekly METhours				
Age		Mean	SD	IQR	%SED	%VIG	Mean	SD	IQR	%SED	%VIG
≤9*	N=6425	17,3	13,7	22,4	21,2%	8,9%	10,6	9,9	12,0	25,2%	2,4%
	Twins (99.92%)	17,3	13,7	22,4	21,2%	8,9%	10,6	10,0	12,0	25,1%	2,4%
10-12	N=6683	25,3	18,6	24,9	15,0%	27,2%	16,1	14,4	19,2	16,9%	11,0%
	Twins (97.71%)	25,2	18,5	24,9	15,0%	27,2%	16,1	14,4	19,2	16,9%	11,0%
	Singletons (2.29%)	26,8	24,6	21,0	18,9%	27,0%	16,7	14,9	18,5	17,7%	12,7%
12-15	N=6181	26,6	22,7	32,3	20,5%	34,3%	18,4	18,7	23,9	21,8%	17,5%
	Twins (91.98%)	26,7	22,4	32,3	20,1%	34,4%	18,4	18,8	23,5	21,7%	17,5%
	Singletons (7.99%)	24,7	25,4	34,2	24,7%	32,5%	18,6	18,2	25,1	22,3%	17,9%
15-18	N=6948	27,3	26,0	41,0	27,2%	37,3%	17,5	21,1	27,4	33,0%	18,2%
	Twins (86.07%)	27,3	25,6	41,0	26,5%	37,6%	17,6	21,5	27,5	32,8%	18,7%
	Singletons (13.46%)	27,3	28,3	42,1	31,0%	35,6%	16,5	18,4	26,4	34,6%	15,3%
	Parents/spouses (0.47%)	16,6	16,0	26,1	36,4%	18,2%	21,0	16,6	25,9	22,7%	18,2%
18-21	N=4534	23,5	25,4	36,0	30,9%	28,5%	15,5	17,7	23,4	32,7%	13,1%
	Twins (76.93%)	23,4	24,8	36,0	30,2%	28,7%	15,9	17,8	24,0	30,9%	13,9%
	Singletons (21.72%)	23,3	26,4	36,0	33,9%	26,8%	13,6	17,3	21,9	39,6%	9,6%
	Parents/spouses (1.35%)	39,3	42,6	27,0	18,8%	50,0%	17,3	18,8	23,9	26,7%	17,8%
21-25	N=4055	21,8	23,0	36,0	29,6%	25,9%	15,2	17,6	22,0	30,6%	10,8%
	Twins (73.93%)	22,2	22,7	36,0	27,8%	26,4%	15,8	17,7	22,8	29,8%	12,2%
	Singletons (21.11%)	21,3	23,4	36,0	32,8%	25,6%	14,3	18,6	19,6	31,2%	7,5%
	Parents/spouses (4.96%)	18,1	25,9	30,6	42,3%	21,1%	9,8	11,2	14,9	39,2%	3,9%
26-35	N=6101	16,6	20,0	24,8	36,2%	15,7%	10,5	14,2	15,6	40,4%	5,8%
	Twins (56.71%)	17,7	20,5	26,6	32,6%	17,0%	11,4	15,1	16,5	38,2%	6,7%
	Singletons (16.82%)	17,4	20,2	26,4	34,7%	17,0%	11,5	14,1	16,5	35,7%	6,2%
	Parents/spouses (26.47%)	14,0	18,8	22,0	44,0%	12,4%	8,1	12,0	12,0	48,4%	3,6%
36-45	N=13668	12,1	17,8	18,0	46,0%	9,0%	8,8	11,7	13,6	43,7%	3,4%
	Twins (17.28%)	15,5	19,3	22,0	32,9%	10,9%	11,1	13,6	16,0	35,0%	5,3%
	Singletons (4.78%)	15,9	22,7	24,0	36,4%	12,3%	10,3	11,9	15,8	35,6%	3,8%
	Parents/spouses (77.9%)	11,3	17,1	17,0	48,8%	8,5%	8,1	11,1	12,5	46,5%	2,8%
46-55	N=10838	11,8	18,5	17,0	49,3%	9,0%	9,4	13,0	13,8	45,3%	4,7%
	Twins (8.72%)	14,2	17,7	19,3	37,8%	12,7%	10,6	12,8	16,5	40,1%	5,7%
	Singletons (3.03%)	10,1	14,2	16,2	48,6%	5,6%	11,5	15,0	16,0	37,1%	6,3%
	Parents/spouses (88.24%)	11,7	18,6	17,0	49,9%	8,9%	9,1	12,9	13,5	46,5%	4,5%
≥56	N=7338	11,4	19,6	16,0	53,0%	9,3%	8,0	13,2	11,1	52,1%	3,7%
	Twins (16.29%)	11,4	20,3	14,4	53,8%	9,7%	8,2	12,9	11,0	53,0%	4,7%
	Singletons (4.74%)	10,2	20,2	14,6	52,1%	6,3%	9,2	15,5	12,0	52,4%	6,3%
	Parents/spouses (78.96%)	11,5	19,6	16,3	52,9%	9,4%	7,9	13,1	11,1	51,8%	3,3%

*Table 4.1. Prevalence of voluntary exercise behavior in a large population-based sample (N=78,524) from the Netherlands. SED: Sedentary, VIG: Vigorous, *: Almost no siblings participate in this age range. N: Number of participants with valid data, SD: standard deviation, IQR: Inter-quantile range, vigorous exercisers are defined as subjects with a METhr score over 35 (i.e., at least five hours per week vigorous (MET=7) activity).*

We find a lower volume of exercise behavior in young children (< 9) compared to older children and adolescents. This contrasts with findings for total physical activity, which is known to decrease from a highest level in early childhood to lower levels in late childhood and adolescence (Rowland, 2016). We interpret this to reflect the predominance of play in determining the total physical activity of young children (Pellegrini & Smith, 1998). In keeping with previous reports, the volume of voluntary exercise behavior in leisure time is lower in females than in males and shows a general decline after adolescence in both sexes. For example, 72.8% of the young people between 15 and 18 years old engages in some form of regular exercise and 37.3% engages in regular vigorous exercise, whereas in adults over 55 years, these numbers have dropped to 47.0% and 9.3%. Formal testing of the linear and quadratic effects of age (in years) and sex (females coded 1) and their interaction on weekly METhr in general equation estimation (GEE) models, accounting for the nested (family) structure of the data, revealed effects of sex, with males showing higher levels of exercise (β = –0.53, SE = 0.02, $p < 10^{-99}$) and linear decrease in exercise with age (β = –0.09, SE = 0.004, $p = 1.8\times10^{-88}$) as well as an additional quadratic age effect ($\beta = -1.03\times10^{-4}$, SE = 4.84×10^{-6}, $p < 10^{-99}$). A significant interaction between age and sex (β = 0.04, SE = 0.002, p = 5.91×10^{-55}) shows that the decrease in exercise after adolescence is less pronounced in females compared to males, which is in line with previous findings (de Moor et al., 2006).

To reconfirm that exercise behaviors do not differ between twins and non-twins, we selected families with same-sex twins where at least one non-twin sibling of the same sex had also participated. We then compared the voluntary exercise behavior of twins with their non-twin siblings. No differences between twins and non-twin siblings within the same family were found for either males (N families = 5141, β_{MZ} = –0.02, SE_{MZ} = 0.06, p_{MZ} = 0.73; β_{DZ} = -0.02, SE_{DZ} = 0.06, p_{DZ} = 0.73) or

females (N families = 6028, β_{MZ} = 0.05, SE_{MZ} = 0.03, p_{MZ} = 0.12; β_{DZ} = 0.05, SE_{DZ} = 0.03, p_{DZ} = 0.07).

The prevalence of regular voluntary exercise behavior in the Dutch twin families is very consistent with the prevalence of voluntary leisure-time physical activity reported by five large-scale studies (Caspersen et al., 2000; Haase et al., 2004; Martinez-Gonzalez et al., 2001; Steptoe et al., 2002; Steptoe et al., 1997). Three of these studies were conducted in 10,000–20,000 young adults (mostly students) aged 18 to 30 years old from the European Union and elsewhere (Haase et al., 2004; Steptoe et al., 2002; Steptoe et al., 1997). All three studies found comparable prevalences of leisure-time exercise: about 70% of the students exercised regularly, meaning they engaged in exercise at least one time per two weeks (e.g., sports activities, physically active pastime). The fourth study was a pan-European study (PAN) of exercise participation in more than 15,000 adults aged 15 up to more than 65 years old from 15 member states of the European Union (Martinez-Gonzalez et al., 2001). They found that 76% of the male and 71% of the female EU population participated in some kind of physical activity during leisure time. In a fifth study, physical activity levels were assessed in 43,732 men and women from the United States aged 18 years and older (Caspersen et al., 2000). The average prevalence across adulthood for any form of activity (five or more times a week and 30 minutes or more per occasion of any activity, but no intensity specified) varied between 73% for women and 79% for men. These prevalences are somewhat higher than our 67%, but this reflects the use of total leisure-time physical activity compared to our restriction to exercise and sports activities.

Overall, a clear picture arises from the six studies described above. Despite its well-documented benefits, a large group of people does not engage in voluntary exercise on a regular basis, particularly in adulthood. Across Europe and the United States, the prevalence for

a complete lack of regular voluntary exercise behavior in adults varies between 21% and 53% depending on the exact age group and the assessment methods used. What factors cause exercisers to exercise and, more importantly, what keeps non-exercisers from doing the same? The remainder of this chapter will review evidence from twin and family studies for a significant genetic contribution to voluntary exercise behavior.

Twin studies on voluntary exercise behavior

As with many other traits, voluntary exercise behavior appears to run in the family. The familial resemblance observed for exercise can represent genetic influences ("nature") or they can represent environmental influences that are shared within a family ("nurture"). Twin studies can separate these two mechanisms by comparing the resemblance in monozygotic (MZ) or identical twins to the resemblance in dizygotic (DZ) or fraternal twins. When twins are reared together they share part of their environment and this sharing of the family environment is the same for MZ and DZ twins. The important difference between MZ and DZ twins is that the former share (close to) all of their genotypes, whereas the latter share on average only half of the genotypes segregating in that family (Falconer & Mackay, 1996; Plomin et al., 2013).

If the resemblance in voluntary exercise behavior within MZ pairs is larger than in DZ pairs, this suggests that additive genetic factors (A) influence exercise behavior. Additive genetic factors represent the sum of all linear effects of the genetic loci that influence the trait of interest. If the resemblance in voluntary exercise behavior is as large in DZ twins as it is in MZ twins, this points to shared environmental factors (C) as the cause of family resemblance (Boomsma et al., 2002). Shared environmental factors are defined as those environmental factors that are shared within a family, such as

parenting behaviors, family functioning, neighborhood or socioeconomic status. The extent to which MZ twins do not resemble each other is ascribed to the unique (or non-shared) environmental factors (E). These include all unique experiences like differential jobs or lifestyles, accidents or other life events, and in childhood, differential treatment by the parents, and nonshared friends and peers. Measurement error will also be subsumed by the unique environmental factor.

The effect of genetic and environmental factors on voluntary exercise behavior in a sample of MZ and DZ pairs can be graphically depicted in a path diagram (see Figure 4.1).

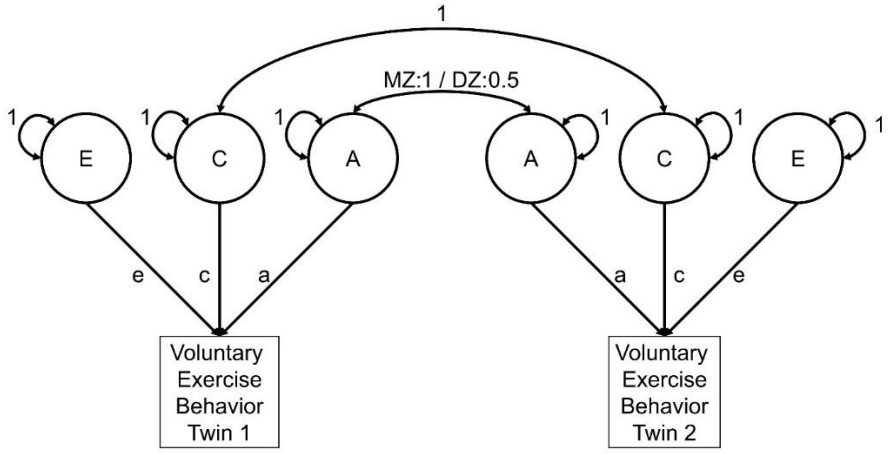


Figure 4.1. Path diagram depicting correlated latent additive genetic (A) and shared environmental (C) factors that can cause twin resemblance in voluntary exercise behavior, as well as unique environmental (E) factors that are uncorrelated in the twins. We set the correlation of the latent genetic factors for MZ twins to 1, and for DZ twins to 0.5. The correlation of shared environmental factors is set to 1 for both types of twins. The latent unique environmental factors are (per definition) uncorrelated. Unobserved latent variables have no scale: their variance is arbitrarily set to 1 (the double arrows that start and end in the latent factors). Path coefficients a, c, and e represent the factor loadings of exercise on the latent factors.

The heritability of voluntary exercise behavior is defined as the relative proportion of the total variance explained by genetic factors. In terms of the parameter estimates obtained from fitting the twin model to the data, the heritability can be expressed as $a^2/(a^2 + c^2 + e^2)$. The heritability can also be represented as a percentage by multiplying this ratio by 100. The proportions of variance ascribed to shared environmental and unique environmental factors can be computed in a similar manner and are given by, respectively, $c^2/(a^2 + c^2 + e^2)$ and $e^2/(a^2 + c^2 + e^2)$.

An overview of twin studies conducted on voluntary exercise behavior is given in Table 4.2.

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
Heller et al., 1988, Australia	adolescent and adult twins, 106 MZ pairs (36.8% male), 94 DZ pairs (26.6% male), ages 17 ~ 66 (M = 36, SD = 12)	Vigorous Exercise in the past two weeks	Males: a^2 = 39% Females: a^2 = 39%
Boomsma et al., 1989, Netherlands	adolescent twins, 44 MZ pairs (36.4% male), 46 DZ pairs (32.6% male), ages 14 ~ 20 (M = 17, SD = 2)	Sports participation variable is dichotomous: "Have you been involved in sports activities during the last three months?" YES/NO	Males: a^2 = 77%, e^2 = 23% Females: a^2 = 35%, e^2 = 65%
Koopmans et al., 1994, Netherlands	adolescent and adult twins, 578 MZ pairs (43.1% male), 1000 DZ pairs (24.1% male), ages 13 ~ 22 (M = 18, SD = 2)	Sports participation variable is dichotomous: "Do you participate in sports?" YES/NO	Males: a^2 = 45%, c^2 = 44%, e^2 = 11% Females: a^2 = 45%, c^2 = 44%, e^2 = 11%
Lauderdale et al., 1997, USA	adult male Vietnam Era Twin Registry, 1006 MZ pairs, 530 DZ pairs, ages 33 ~ 51 (M = 41)	Vigorous Exercise (YES/NO): jogging > 10 miles/week)	Males: a^2 = 53%, e^2 = 47%
		Vigorous Exercise (YES/NO): racquet sports > 5 hr/week	Males: a^2 = 48%, c^2 = 4%, e^2 = 48%
		Vigorous Exercise (YES/NO): strenuous sports	Males: a^2 = 30%, c^2 = 17%, e^2 = 53%
		Vigorous Exercise (YES/NO): bicycling > 50 miles/week	Males: a^2 = 58%, e^2 = 42%
		Vigorous Exercise (YES/NO): swimming > 2 miles/week	Males: a^2 = 8%, c^2 = 31%, e^2 = 61%
Beunen & Thomis, 1999, Belgium	adolescent twins, 43 MZ pairs (48.8% male), 61 DZ pairs (34.4% male), M age = 15	Number of hours spent on sports each week	Males: a^2 = 83%, e^2 = 17% Females: a^2 = 44%, c^2 = 54%, e^2 = 02%
Kujala et al., 2002, Finland	adult same sex twins, 1772 MZ pairs, 3551 DZ pairs, ages 24 ~ 60	Participation in vigorous physical activity: either “alternatively walking and jogging”, “jogging”, or “running”	Males: a^2 = 45% Females: a^2 = 45%
Frederiksen and Christensen, 2003, Denmark	elderly twins, 616 MZ pairs, 642 DZ pairs, ages 45 ~ 68	Engage in any of: jogging, gym, swimming, tennis, badminton, football, handball, aerobics, rowing, table tennis, volleyball, in leisure time	Males: a^2 = 49%, e^2 = 51% Females: a^2 = 49, e^2 = 51%

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
de Geus et al., 2003, Netherlands	adolescent twins, 69 MZ pairs (50.7% male), 88 DZ pairs (34.1% male), ages 13 ~ 22 (M = 16.7, SD = 2)	Average weekly METh spent on sports or other moderate to vigorous voluntary exercise activities (with intensity of 4 METh/wk or higher) in the last three months	Males: a^2 = 79%, e^2 = 21% Females: a^2 = 79%, e^2 = 21%
	adult twins, 93 MZ pairs (48.4% male), 115 DZ pairs (32.2% male), ages 35 ~ 62 (M = 44, SD = 7)	Average weekly METh spent on sports or other moderate to vigorous voluntary exercise activities (with intensity of 4 METh/wk or higher) in the last three months	Males: a^2 = 41%, e^2 = 59% Females: a^2 = 41%, e^2 = 59%
Simonen et al., 2004, Finland	adult twins, 147 MZ pairs, 153 DZ pairs, M age = 50 (SD = 8)	Weekly hours spent on leisure time exercise at ages 12 ~ 18	Males: a^2 = 18%, c^2 = 37%, e^2 = 45% Females: a^2 = 18%, c^2 = 37%, e^2 = 45%
		Weekly hours spent on leisure time exercise at ages 18 ~ 70	Males: a^2 = 51%, e^2 = 49% Females: a^2 = 51%, e^2 = 49%
Stubbe et al., 2005, Netherlands	adolescent twins NTR, 276 MZ pairs (41.7% male), 370 DZ pairs (23.5% male), ages 13 ~ 14	The METh/week score was dichotomized into YES/NO at a 4 METh/wk threshold, excluded PE	Males: a^2 = 0%, c^2 = 84%, e^2 = 16% Females: a^2 = 0%, c^2 = 84%, e^2 = 16%
	adolescent twins NTR, 321 MZ pairs (42.4% male), 442 DZ pairs (25.3% male), ages 15 ~ 16	The METh/week score was dichotomized into YES/NO at a 4 METh/wk threshold, excluded PE	Males: a^2 = 0%, c^2 = 78%, e^2 = 22% Females: a^2 = 0%, c^2 = 78%, e^2 = 22%
	adolescent twins NTR, 248 MZ pairs (40.3% male), 395 DZ pairs (24.3% male), ages 17 ~ 18	The METh/week score was dichotomized into YES/NO at a 4 METh/wk threshold, excluded PE	Males: a^2 = 36%, c^2 = 47%, e^2 = 17% Females: a^2 = 36%, c^2 = 47%, e^2 = 17%
	adolescent twins NTR, 250 MZ pairs (36.8% male), 326 DZ pairs (25.2% male), ages 19 ~ 20	The METh/week score was dichotomized into YES/NO at a 4 METh/wk threshold, excluded PE	Males: a^2 = 85%, e^2 = 15% Females: a^2 = 85%, e^2 = 15%
Eriksson et al., 2006, Sweden	young adult male twins, N = 1234	Leisure time sports from a modified version of the Baecke questionnaire.	Males: a^2 = 56%, e^2 = 44%
Stubbe et al., 2006, Australia	adult twins, 1260 MZ pairs (32.6% male), 1468	Spending at least 60 minutes each week on	Males: a^2 = 48%, e^2 = 52% Females: a^2 = 48%, e^2 = 52%

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
	DZ pairs (18.3% male), ages 19 ~40 (M = 31.8)	exercise activities with intensity > 4 MET	
Stubbe et al., 2006, Denmark	adult twins, 3046 MZ pairs (43.3% male), 6410 DZ pairs (25.6% male), ages 19 ~40 (M = 31.1)	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Males: a^2 = 52%, e^2 = 48% Females: a^2 = 52%, e^2 = 48%
Stubbe et al., 2006, Finland	adult twins, 2841 MZ pairs (43.8% male), 6001 DZ pairs (44.5% male), ages 19 ~40 (M = 26.9)	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Males: a^2 = 62%, e^2 = 38% Females: a^2 = 62%, e^2 = 38%
Stubbe et al., 2006, Netherlands	adult twins NTR, 1266 MZ pairs (33.4% male), 1415 DZ pairs (20.8% male), ages 19 ~40 (M = 25.7)	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Males: a^2 = 67%, e^2 = 33% Females: a^2 = 67%, e^2 = 33%
Stubbe et al., 2006, Norway	adult twins, 1502 MZ pairs (42.5% male), 2493 DZ pairs (21.8% male), ages 19 ~40 (M = 24.7)	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Males: a^2 = 26%, c^2 = 37%, e^2 = 37% Females: a^2 = 56%, e^2 = 44%
Stubbe et al., 2006, Sweden	adult twins, 3598 MZ pairs (45.4% male), 5329 DZ pairs (47.3% male), ages 19 ~40 (M = 28.6)	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Males: a^2 = 62%, e^2 = 38% Females: a^2 = 62%, e^2 = 38%
Stubbe et al., 2006, UK	adult female twins, 163 MZ pairs (0% male), 259 DZ pairs (0% male), ages 19 ~40 (M = 32.4)	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Females: a^2 = 71%, e^2 = 29%
de Moor, Posthuma et al., 2007, Netherlands	Adult NTR twins 156 MZ pairs (37.8% male), 326 DZ pairs (20.2% male), ages 18 ~ 50	Spending at least 60 minutes each week on exercise activities with intensity > 4 MET	Males: a^2 = 69%, e2 = 31% Females: a^2 = 46%, e^2 = 54%
de Moor, Stubbe et al., 2007, Netherlands	adult twins NTR, 755 MZ pairs (26.9% male), 2306 DZ pairs (15.3% male), M age = 31 (SD = 7)	Three groups based on two MET-thresholds: 0 sedentary; 0–6METh/wk moderate; 6+ METh/wk vigorous	Males: a^2 = 54%, e^2 = 46% Females: a^2 = 54%, e^2 = 46%
McCaffery et al., 2009, USA	Vietnam Era Twin Registry, 2024 MZ pairs (100% male), 1566 DZ pairs (100% male), M ages = 41.07 (SD = 3.15)	Jog or run ≥10 miles/wk OR play strenuous racquet sports ≥ 5 h/wk OR ride a bicycle ≥ 50 miles/wk OR swim ≥ 2 miles/wk OR play other strenuous sports	Males: a^2 = 10%, c^2 = 18%, e^2 = 72%
Aaltonen et al., 2010, Finland	Baseline young adult twin data, N = 13556, ages 18 ~ 54	MET hours/day.	Males: a^2 = 47%, e2 = 53%

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
			Females: a^2 = 42%, e^2 = 58%
	Follow-up young adult twin data, N = 13822, ages 24 ~ 60	MET hours/day.	Males: a^2 = 38%, e^2 = 62% Females: a^2 = 31%, e^2 = 69%
van der Aa et al., 2010, Netherlands	young twins NTR, 554 MZ pairs (38.1% male), 948 DZ pairs (21.2% male), ages 13 ~14 (M = 14.5, SD = 0.31)	Three groups based on two thresholds :0-5 METh/wk; 5-20 METh/wk; 20+ METh/wk	Males: a^2 = 85%, e^2 = 15 % Females: a^2 = 38%, c^2 = 46%, e^2 = 16%
	young twins NTR, 662 MZ pairs (42.6% male), 969 DZ pairs (21.7% male), ages 15 ~16 (M = 16.2, SD = 0.61)	Three groups based on two thresholds :0-5 METh/wk; 5-20 METh/wk; 20+ METh/wk	Males: a^2 = 80%, e^2 = 20% Females: a^2 = 80%, e^2 = 20%
	young twins NTR, 488 MZ pairs (34.6% male), 747 DZ pairs (20.9% male), ages 17 ~19 (M = 18.1, SD = 0.7)	Three groups based on two thresholds :0-5 METh/wk; 5-20 METh/wk; 20+ METh/wk	Males: a^2 = 72%, e^2 = 28% Females: a^2 = 72%, e^2 = 28%
Mustelin et al., 2011, Finland	young adult twins FinnTwin13, 59 MZ pairs, 92 DZ pairs, ages 23 ~ 31 (M = 27.4, SD = 2)	Baecke questionnaire: sport index (sports activities during leisure time)	Males: a^2 = 56% Females: a^2 = 56%
de Moor et al., 2011, Netherlands	young twins NTR, 656 MZ pairs (23.3% male), 1628 DZ pairs (13.9% male), ages 13 ~18 (M = 16.4, SD = 1.1)	Spending at least 60 minutes each week on exercise activities > 4 MET	Males: a^2 = 42%, c^2 = 44%, e^2 = 14% Females: a^2 = 36%, c^2 = 52%, e^2 = 12%
	adult twins NTR, 685 MZ pairs (42.6% male), 1223 DZ pairs (24.9% male), ages 30 ~65 (M = 39.9, SD = 9.4)	Spending at least 60 minutes each week on exercise activities > 4 MET	Males: a^2 = 42%, e^2 = 58% Females: a^2 = 42%, e^2 = 58%
Vink et al., 2011, Australia	(young) adult twins, N = 4604	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 22 ~ 36	Males: a^2 = 13%, c^2 = 27, e^2 = 60 % Females: a^2 = 37%, c^2 = 6%, e^2 = 57%
		Spending at least 60 minutes each week on exercise activities > 4 MET at ages 37 ~ 50	Males: a^2 = 13%, c^2 = 27%, e^2 = 60% Females: a^2 = 37%, c^2 = 6%, e^2 = 57%

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
Vink et al., 2011, Denmark	(young) adult twins, N = 26298	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 19 ~ 33	Males: a^2 = 47%, c^2 = 1%, e^2 = 52% Females: a^2 = 27%, c^2 = 15%, e^2 = 58%
		Spending at least 60 minutes each week on exercise activities > 4 MET at ages 34 ~ 50	Males: a^2 = 47%, c^2 = 1%, e^2 = 52% Females: a^2 = 27%, c^2 = 15%, e^2 = 58%
Vink et al., 2011, Finland	(young) adult twins, N = 23095	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 19 ~ 27	Males: a^2 = 57%, c^2 = 3%, e^2 = 40% Females: a^2 = 56%, c^2 = 3%, e^2 = 42%
		Spending at least 60 minutes each week on exercise activities > 4 MET at ages 28 ~ 50	Males: a^2 = 57%, c^2 = 3%, e^2 = 40% Females: a^2 = 56%, c^2 = 3%, e^2 = 42%
Vink et al., 2011, Netherlands	(young) adult twins, N = 6753	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 19 ~ 25	Males: a^2 = 64%, c^2 = 3%, e^2 = 32 % Females: a^2 = 28%, c^2 = 23%, e^2 = 48%
		Spending at least 60 minutes each week on exercise activities > 4 MET at ages 26 ~ 50	Males: a^2 = 64%, c^2 = 3%, e^2 = 32% Females: a^2 = 28%, c^2 = 23%, e^2 = 48%
Vink et al., 2011, Norway	young adult twins, N = 9066	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 19 ~ 31	Males: a^2 = 35%, c^2 = 29%, e^2 = 36% Females: a^2 = 54%, c^2 = 3%, e^2 = 44%
Vink et al., 2011, Sweden	(young) adult twins, N = 27414	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 19 ~ 35	Males: a^2 = 53%, e^2 = 46% Females: a2 = 54%, e^2 = 46%
		Spending at least 60 minutes each week on exercise activities > 4 MET at ages 36 ~ 50	Males: a^2 = 53%, e^2 = 46% Females: a^2 = 54%, e^2 = 46%
Vink et al., 2011, UK	(young) adult female twins, N = 2451	Spending at least 60 minutes each week on exercise activities > 4 MET at ages 19 ~ 40	Females: a^2 = 57%, c^2 = 9%, e^2 = 42%
		Spending at least 60 minutes each week on exercise activities > 4 MET at ages 41 ~ 50	Females: a^2 = 57%, c^2 = 9%, e^2 = 42%

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
Mustelin et al., 2012, Finland	young adult twins FinnTwin12, 229 MZ pairs (42.4% male), 347 DZ pairs (25.4% male), ages 20 ~26 (M = 22.4, SD = 0.7)	Sport activities are scored as 1,3 or 5 according to their intensity	Males: a^2 = 64%, e^2 = 36% Females: a^2 = 64%, e^2 = 36%
Huppertz et al., 2012, Netherlands	young twins NTR, 648 MZ pairs (45.8% male), 1320 DZ pairs (26.1% male), M age = 7.45 (SD = 0.32)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 24%, c^2 = 71%, e^2 = 6% Females: a^2 = 22%, c^2 = 67%, e^2 = 11%
	young twins NTR, 620 MZ pairs (45.8% male), 1141 DZ pairs (26.1% male), M age = 10.1 (SD = 0.33)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 66%, c^2 = 25%, e^2 = 10% Females: a^2 = 16%, c^2 = 72%, e^2 = 11%
	young twins NTR, 1540 MZ pairs (46.7% male), 2746 DZ pairs (24.3% male), M age = 12.3 (SD = 0.4)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 38%, c^2 = 50%, e^2 = 11% Females: a^2 = 36%, c^2 = 50%, e^2 = 11%
Aaltonen et al., 2013, Finland	adolescent twins FinnTwin16, 769 MZ pairs (38.1% male), 1743 DZ pairs (24.6% male), M age = 16.2 (SD = 0.1)	1: exercising less than once a week; 2: exercising one to three times per week; 3: exercising four or more times per week	Males: a^2 = 52%, c^2 = 19%, e^2 = 29 % Females: a^2 = 52%, c^2 = 24%, e^2 = 24%
	adolescent twins FinnTwin16, 724 MZ pairs (36.3% male), 1614 DZ pairs (24.7% male), M age = 17.1 (SD = 0.1)	1: exercising less than once a week; 2: exercising one to three times per week; 3: exercising four or more times per week	Males: a^2 = 44%, c^2 = 24%, e^2 = 32% Females: a^2 = 50%, c^2 = 26%, e^2 = 24%
	young adult twins FinnTwin16, 715 MZ pairs (35.9% male), 1603 DZ pairs (24.2% male), M age = 18.6 (SD = 0.2)	1: exercising less than once a week; 2: exercising one to three times per week; 3: exercising four or more times per week	Males: a^2 = 46%, c^2 = 23%, e^2 = 31% Females: a^2 = 51%, c^2 = 21%, e^2 = 28%
	young adult twins FinnTwin16, 613 MZ pairs (37.8% male), 1351 DZ pairs (23.2% male), M age = 24.5 (SD = 0.9)	1: exercising less than once a week; 2: exercising one to three times per week; 3: exercising four or more times per week	Males: a^2 = 34%, c^2 = 43%, e^2 = 23% Females: a^2 = 31%, c^2 = 49%, e^2 = 20%
Huppertz, Bartels, Jansen et al., 2014, Netherlands	adult twins NTR, 701 MZ pairs (27% male), 572 DZ pairs (14.2% male), ages 18 ~50 (M = 30.5, SD = 7)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 50%, e^2 = 50% Females: a^2 = 43%, e^2 = 57%

Study	Sample	Exercise Behavior Phenotype	Heritability estimates
Huppertz, Bartels, Groen-Blokhuis et al., 2014, Netherlands	adult twins NTR, 701 MZ pairs (27% male), 572 DZ pairs (14.2% male), M age = 30.7 (SD = 7)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 50%, e^2 = 50% Females: a^2 = 43%, e^2 = 57%
Huppertz et al., 2016, Netherlands	young twins NTR, 1262 MZ pairs (47.9% male), 2384 DZ pairs (27.1% male), M age = 7.52 (SD = 0.34)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 14%, c^2 = 80%, e^2 = 6% Females: a^2 = 12%, c^2 = 80%, e^2 = 8%
	young twins NTR, 1384 MZ pairs (48.3% male), 2582 DZ pairs (26.1% male), M age = 9.84 (SD = 0.43)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 26%, c^2 = 69%, e^2 = 7% Females: a^2 = 26%, c^2 = 65%, e^2 = 8%
	young twins NTR, 2615 MZ pairs (46.4% male), 4589 DZ pairs (24.9% male), M age = 12.25 (SD = 0.4)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 31%, c^2 = 62%, e^2 = 7% Females: a^2 = 27%, c^2 = 65%, e^2 = 8%
	young twins NTR, 1451 MZ pairs (39.4% male), 2333 DZ pairs (21.4% male), M age = 14.61 (SD = 0.6)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 43%, c^2 = 36%, e^2 = 21% Females: a^2 = 40%, c^2 = 43%, e^2 = 17%
	young twins NTR, 959 MZ pairs (39.9% male), 1305 DZ pairs (21.2% male), M age = 16.87 (SD = 0.45)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 56%, c^2 = 27%, e^2 = 17% Females: a^2 = 49%, c^2 = 31%, e^2 = 20%
	young twins NTR, 458 MZ pairs (30.6% male), 572 DZ pairs (18.5% male), M age = 18.77 (SD = 0.51)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 79%, c^2 = 4%, e^2 = 17% Females: a^2 = 49%, c^2 = 19%, e^2 = 33%
Nederend et al., 2016, Netherlands	young twins NTR, 114 MZ pairs (50.9% male), 111 DZ pairs (32.4% male), M age = 17.1 (SD = 1.1)	The METh/week score based on the summed MET×time/wk score over all activities performed	Males: a^2 = 80%, e^2 = 20% Females: a^2 = 80%, e^2 = 20%

Table 4.2. Twin studies on voluntary exercise behavior. NB: A number of studies reported on overlapping samples, most notably Vink et al. (2011) with Stubbe et al. (2005), and van der Aa (2010) and Huppertz et al. (2014) with Huppertz et al. (2016) although the exact exercise phenotype and age categories used differed.

Despite the slightly different measures used (i.e., Vigorous Exercise, Sports Participation, Moderate-to-Vigorous Exercise, Leisure-Time Sports, Vigorous Exercise, Exercise Participation, Sports Physical Activity) and the use of either ordinal or interval scales, these studies converge on the finding that voluntary exercise behavior is heritable across the life span. There are, however, remarkable age-related changes in the relative contribution of genetic factors to the total population variance in this behavior. The largest changes occur during childhood and adolescence (Huppertz et al., 2016; Huppertz et al., 2012; van der Aa et al., 2010). Huppertz et al. (2016) used a combination of univariate modeling on a very large cross-sectional set of Dutch twins (birth year ranged between 1986 and 2004) and a genetic simplex model on the subset of twins that had longitudinal data across multiple ages. Heritability was found to be low in seven-year-olds (14% in males and 12% in females) but gradually increased up to age 18 (79% in males and 49% in females). In contrast, the initially very high relative contribution of the shared environment to the variance in childhood exercise behavior rapidly waned when children grew into adolescence and young adulthood (from 80% to 4% in males and from 80% to 19% in females). This decrease in the shared environmental variance, which is stronger in males than in females, occurred in parallel to a large increase in the genetic variance. The genetic effects in males remained largely the same from childhood to late adolescence (although their relative importance increased), whereas in females, the contribution of novel genetic effects as a function of age was more profound.

Twin studies in Finland, Sweden, and Belgium have shown that the age-moderation of the genetic and shared environmental effects on exercise behaviors hold in these countries too (Aaltonen et al., 2013; Beunen & Thomis, 1999; Carlsson et al., 2006), and further confirmed

that shared environmental factors seem more important in young adolescent girls than in boys. The clear age moderation seen by Huppertz et al. (2016) in twins born between 1986 and 2004 was already observed by Stubbe et al. (2005) in an earlier birth study of Dutch adolescent twins born between 1971 and 1987. Taken together, these results imply that family-based interventions are useful to increase this health behavior in children, whereas individually based interventions might be better suited for adolescents.

Studies conducted in adult twins between 19 and 68 years (Aaltonen et al., 2010, 2013; Boomsma et al., 1989; de Geus et al., 2003; de Moor, Posthuma, et al., 2007; de Moor, Stubbe, et al., 2007; de Moor et al., 2011; Eriksson et al., 2006; Frederiksen & Christensen, 2003; Heller et al., 1988; Huppertz, Bartels, Jansen, et al., 2014; Koopmans et al., 1994; Kujala et al., 2002; Lauderdale et al., 1997; McCaffery et al., 2009; Mustelin et al., 2012; Mustelin et al., 2011; Nederend et al., 2016; Simonen et al., 2004; Stubbe et al., 2005; Stubbe et al., 2006; Vink et al., 2011) show that variation in exercise behavior is influenced by genetic factors, with heritability estimates ranging between 10% and 85% of the variance in exercise. Most of these studies included twins of a wide age range, for example, 19 to 40 years (Stubbe et al., 2006) or 18 to 60 years (Aaltonen et al., 2010). A large study by Vink et al. (2011) investigated in seven different countries whether the heritability of exercise behavior during adulthood (19 to 50 years) changes with age. In four of the seven countries (Denmark, Finland, Sweden, and the Netherlands), a significant decrease in heritability was observed with age (from 60–90% at age 19 to 13–40% at age 50). In the other three countries (Australia, Norway, and United Kingdom), the heritability of exercise was found to be stable during adulthood.

The results from the twin studies in Table 4.2 demonstrate convincingly that heritable factors are present at all ages. The

heritability of voluntary exercise behavior peaks during late adolescence and young adulthood.

Family studies on voluntary exercise behavior

The heritability estimates obtained from twin studies are valid only if the assumptions underlying the twin method hold. An important assumption is the equal environment assumption, which states that environmentally caused similarity in the trait or behavior of interest (i.e., voluntary exercise behavior) is the same in MZ and DZ twin pairs (Plomin et al., 2013). Only one twin study has explicitly tested the equal environment assumption for a set of relevant traits (Eriksson et al., 2006). It was tested whether twin pairs with higher contact frequency were more similar in their total physical activity, physical activity at work or in leisure-time separately, and whether this effect depended on zygosity status. No evidence for violation of the equal environment assumption was found for any of the physical activity measures. Other assumptions of the twin method include absence of assortative mating and the independence of genes and environment. If phenotypic assortative mating on exercise behavior is present, the correlation between genetic factors influencing exercise behavior is higher than the theoretical 50% in DZ twins and full siblings, which can cause underestimation of heritability and overestimation of the shared environmental effects. If there is cultural transmission, that is, the parents' exercise behavior is a part of the relevant shared environment by the twins and acts as a "role-model" for the exercise behavior of the children, then the genetic and the shared environmental effects on exercise behavior are no longer independent. Correlation of an exercise-inducing family environment with a genetic propensity to exercise will inflate estimates of C in the twin model. The presence of assortative mating and cultural transmission and how they impact heritability estimates can be tested by adding data from more types of

relatives to the twin design, such as data from parents, grandparents, spouses, and siblings. Spouse correlations inform on the presence of assortative mating. Parent offspring correlations enable testing for the effects of cultural transmission.

Table 4.3 lists the five studies that employed extended family designs to report on the familial resemblance of voluntary exercise behavior among adolescent offspring (age range 13 to 20 years) and their parents (Choh et al., 2009; de Moor et al., 2011; Koopmans et al., 1994; Seabra et al., 2008, 2014).

Study	Sample	Phenotype	Spousal, sibling and parent-offspring correlations	Estimates
Koopmans et al., 1994, Netherlands	N = 6470 Fathers: M age = 48 SD age = 5.7 Mothers: M age = 46 SD age = 5.2 Siblings: M age = 18 SD age = 2.3	Regular sports participation (YES/NO)	Spousal: 0.49 Father-son 0.37, Mother-son 0.32, Father-Daughter: 0.29, Mother-Daughter-0.30	a2 = 45% c2 = 44% e2 = 11%
Seabra et al., 2008, Portugal	N = 9500, Fathers (2375): M age = 45.4 SD = 5.8 Mothers(2375): M age = 42.9 SD = 5.5 Sons (2425): M age = 16.1 SD = 4 Daughters (2325) M age = 16 SD = 4	Baecke et al. questionnaire: sport index	Spousal: 0.29 Female siblings: 0.25, Male siblings: 0.23 Female-male siblings: 0.24 Father-son: 0.19, Mother-son: 0.15 Father-daughter: 0.18, Mother-daughter: 0.18	a2 = 25% e2 = 75%
Choh et al., 2009, USA	N = 518, aged 18 ~ 86 Fathers (219), Mothers(302)	Baecke et al. questionnaire: sport index	Not reported	a2 = 17% c2 = 25% e2 = 58%
de Moor et al, 2011, Netherlands	N = 3663 Fathers (1488): M age = 45.5 SD age = 4.6 Mothers (1650): M age = 45.5 SD age = 4.6 Sons (1636): M age = 16.4 SD age = 1.2 Daughters (1889): M age = 16.4 SD age = 1.2	The METh/wk score was dichotomized into YES/NO regular exercise participation at a > 4 METh/wk threshold	Spousal: 0.41, Father-son: 0.36, Mother-son: 0.18 Father-daughter: 0.21, Mother-daughter: 0.21	a2 = 64% c2 = 0% e2 = 36%

Seabra et al., 2014, Portugal	N = 9500 Fathers (2375): M age = 45.4 SD age = 5.8 Mothers (2375): M age = 42.9 SD age = 5.5 Sons (2425): M age = 16.1 SD age = 4 Daughters (2325): M age = 16 SD age = 4	From Baecke et al. questionnaire: Sports Participation	Spousal: 0.23, Female siblings: 0.31, Male siblings: 0.29 Female-male siblings: 0.23 Father-son: 0.16, Mother-son: 0.11 Father-daughter: 0.14, Mother-daughter: 0.14	a2 = 50% e2 = 50%
		From Baecke et al. questionnaire: Intensity of exercise	Spousal: 0.12, Female siblings: 0.39, Male siblings: 0.51 Female-male siblings: 0.18 Father-son: 0.15, Mother-son: 0.15 Father-daughter: 0.1Mother-daughter: 0.1	a2 = 40% e2 = 60%
		From Baecke et al. questionnaire: Weekly amount of exercise	Spousal: 0.46 Female siblings: 0.52, Male siblings: 0.37 Female-male siblings: 0.22 Father-son: 0.14, Mother-son: 0.21 Father-daughter: 0.23, Mother-daughter: 0.23	a2 = 46% e2 = 54%
		From Baecke et al. questionnaire: Proportion of the year exercising	Spousal: 0.48 Female siblings: 0.52, Male siblings: 0.22 Female-male siblings: 0.27 Father-son: 0.22, Mother-son: 0.08 Father-daughter: 0.4, Mother-daughter: 0.4	a2 = 49% e2 = 51%

Table 4.3. Family studies on voluntary exercise behavior.

The spouse correlation for exercise behavior was significant in all these studies and ranged from 0.16 to 0.48. de Moor et al. (2011) analyzed twin-spouse pairs and demonstrated that the significant spouse correlation for exercise was best explained by phenotypic

assortment (i.e., that partners choose each other because they are similar in their exercise behavior) rather than by social homogamy (i.e., that partners similar in exercise behavior meet and marry because they come from similar social backgrounds) or social interaction processes (i.e., that partners resemble each other because they spend time together and mutually influence each other). Importantly, an extended twin-family model that took assortative mating into account led to nearly the same heritability estimates as the twin-only model.

Parent-offspring correlations have ranged from 0.13 to 0.25 and are generally somewhat lower than the full sibling correlations (0.24 to 0.31). de Moor et al. (2011) estimated in a sample of 3,525 adolescent twins and their siblings (13 to 18 years) and 3,138 parents from 1,736 families that 42% of variation in adolescent exercise behavior in boys could be explained by genetic factors, and 52% could be explained by environmental factors shared by family members of the same generation but not with the parents. In girls, 36% of the variance in exercise was explained by genetic factors and 41% by generation-specific environmental effects. Cultural transmission effects from parents to offspring were not significant in adolescence, that is, adolescents do not simply "copy" the exercise behaviors they see in their parents.

Taken together, the results of extended twin family studies suggest that heritability estimates for exercise behavior obtained from studies using only twins are not overly biased.

Gene-finding studies

The heritability of exercise behavior has been well established, but less is known about the genetic variants that are associated with this trait. Candidate gene studies in single small-sized cohort studies are now widely distrusted as a reliable source of replicable association, because most complex behavioral traits are highly polygenic, with only

a very low percentage of the variance in the trait explained by a single genetic variant. The first genome-wide association (GWA) study on voluntary exercise behavior that we conducted in 2009 was, in retrospect, also heavily underpowered (de Moor et al., 2009). It tested 1,607,535 observed and imputed SNP markers that passed stringent quality controls for their association with leisure-time exercise behavior in two independent samples comprising 1,644 Dutch and 978 American subjects. The most promising finding was in the *PAPSS2* gene (37 SNPs with pooled p-values $<1.0\times10^{-5}$) and in two intergenic regions on chromosomes 2q33.1 and 18p11.32. The *PAPSS2* gene encodes a protein that is involved in the sulfation of compounds such as lipids, carbohydrates, proteins, and exogenous drugs. Interestingly, the 10q23 region that harbors the *PAPSS2* gene has been linked to maximal exercise ability in a genome-wide linkage study of 453 sib pairs (Rico-Sanz et al., 2004). This is in keeping with the theoretical notion that exercise ability may be an important determinant of leisure time exercise behavior (de Geus & de Moor, 2008), which was supported by the finding that regular exercise behavior is correlated with various fitness traits, particularly maximal endurance capacity, not only phenotypically but also genetically (Schutte, Nederend, Hudziak, Bartels, et al., 2016; Schutte, Nederend, Hudziak, de Geus, et al., 2016).

More large-scale collaborative consortia are now direly needed to conduct a meta-analysis on GWA studies on voluntary exercise behavior. These could repeat the successes of similar studies for other health behaviors, like smoking and alcohol use or BMI, that have already discovered many genetic variants (L. S. Chen et al., 2012; Locke et al., 2015; Schumann et al., 2011). Detection of the genetic variants underlying voluntary exercise behavior is a route to increased future understanding of the actual biological pathways underlying the individual differences in the choice for regular exercise behavior, that

is, the biology that makes exercisers exercise but keeps non-exercisers from doing the same. Knowledge of these pathways may further help resolve causality in the well-known association of exercise with other lifestyle behaviors such as overeating and smoking, or with physical health problems like cardiovascular and metabolic disease (Kujala et al., 2002) and mental health problems, including depression (de Moor et al., 2008). More importantly, such understanding may be exploited in stratified or personalized interventions that take innate biological differences into account (de Geus & de Moor, 2008, 2011; Schutte et al., 2017). Future large-scale meta-analytic GWA efforts can therefore be expected to render important clues to improve the success of interventions on this crucial health behavior.

An Extended Twin-Pedigree Study of Different Classes of Voluntary Exercise Behavior

Published as van der Zee MD, Helmer Q, Boomsma DI, Dolan CV, de Geus EJC (2020). An extended twin-pedigree study of different classes of voluntary exercise behavior. *Behavior Genetics 50(2):94-104. doi: 10.1007/s10519-019-09990-7.*

Abstract

We investigated the familial clustering of different classes of voluntary regular exercise behavior in extended twin-family pedigrees. In contrast to the earlier work based on twin data only, this allowed us to estimate the contributions of shared household effects (C), additive (A), and non-additive (D) genetic effects on voluntary exercise behavior. To test whether shared household effects were inflated by assortative mating we examined the causes of spousal resemblance. For adolescent and adult participants (aged 16 to 65) in the Netherlands Twin Register we constructed 19,543 pedigrees which specified all relations among nuclear family members and larger families in the register (N = 50,690 individuals). Data were available on total weekly MET minutes spent on leisure time exercise, and on total weekly MET minutes spent on exercise activities in team-based, solitary, competitive, non-competitive, externally paced and internally paced exercise. We analyzed the data in the Mendel software package (Lange et al., 2013) under multiple definitions of household sharing and used data from spouses of twins to test phenotypic assortment, social homogamy, and marital interaction as potential sources of spousal resemblance. Results confirmed the influence of genetic factors on the total volume of weekly exercise behavior throughout the life span. Broad sense heritability ranged from 34% to 41% (19-26% A, 12-21% D), and did not depend on the definition for household sharing. Engaging in team-based, competitive, externally paced activities (e.g., soccer) was ~13% more heritable than engaging in non-competitive, solitary activities (e.g., jogging). Having shared a household as siblings explained 4-8% of the variance in adult exercise behavior, whereas sharing a household by spouses yielded higher C estimates (20-24%), as it incorporates spousal resemblance. Spousal resemblance was explained by both social homogamy and marital interaction, with little evidence for phenotypic assortment. We conclude that both the

amount of voluntary exercise behavior and the preference for specific classes of exercise activities in adults is explained by additive and non-additive genetic factors and unique environmental influences that include correlated exercise behavior of spouses.

Introduction

Regular voluntary exercise activities in leisure time, such as jogging, swimming, exercising at fitness clubs, and participation in recreational or competitive team and individual sports, are rapidly becoming the major source of physical activity of moderate-to-vigorous intensity in many industrialized countries. Recent meta-analyses of twin studies revealed broad sense heritability estimates for voluntary exercise behavior in adolescence and adulthood of 48% for males and 51% for females (van der Zee & de Geus, 2019), with inconsistent evidence for an effect of shared environmental factors.

Most of the studies included in these meta-analyses had shortcomings that need to be addressed to increase our understanding of the causes of individual differences in this crucial health behavior. First, the vast majority of the studies base their conclusions on the classical twin model comparing monozygotic (MZ) and dizygotic (DZ) twin correlations. While producing robust estimates for broad sense heritability, this model cannot simultaneously estimate all four variance components potentially influencing exercise behavior: additive (A) and non-additive (D) genetic factors and shared (C) and unique (E) environmental factors. Choosing to set either non-additive or shared environmental variance components to zero can yield biased estimates, "sometimes wildly so" (page 377, Keller et al. 2009). Second, previous studies have combined all recorded voluntary exercise activities into a single summary score, lumping together rather different activities like football, jogging, hockey, weight, strength training in fitness centers, tennis, swimming, boxing etc. Preference for these activities may, however, depend on rather different dispositional traits, both in terms of personality and in the physical, cognitive and motor skill demands placed on the individual.

In the current paper we assessed the sources of variation in voluntary regular exercise behavior using an extended twin-pedigree

study design using a rich dataset from the Netherlands Twin Register (NTR). The NTR repository contains exercise data collected with virtually the same methods in twins, their parents, siblings, spouses and the children of twins and siblings. This dataset is therefore ideally suited for an extended twin family design. Such a design allows for the simultaneously estimation of additive and non-additive genetic factors and shared and unique environmental factors (Keller et al., 2009). The often-used analytical approach for this estimation is structural equation modeling (SEM) software such as Mx, OpenMx, Mplus, or LISREL (Jöreskog, 1970; Muthén et al., 2006; Neale et al., 1994). These programs offer great flexibility in model specification but are less well suited for large and irregular pedigrees where the specification of the complex multidimensional parameter space, as well as iteratively fitting the model to actual data, becomes quite challenging (Boomsma et al., 2018).

We previously used the Mendel software package (Lange et al., 2013) as an alternative approach to the analysis of data from large and complex twin-pedigrees (Boomsma et al., 2018). Mendel constructs the genetic relations among all pedigree members, e.g., MZ twin, sibling, cousins, aunt–niece, grandparent–grandchild, etc., and estimates components of genetic covariance for any pair of relatives from the weighted combination of additive and dominance effects (Weiss, 1993). Using Mendel yielded estimates of additive genetic and dominance variance components for neuroticism of 24% and 19%, respectively (Boomsma et al., 2018). These results are consistent with those produced by SEM modeling in extended twin kinships among 45,850 family members from Australia and the United States (Lake et al., 2000).

In an extended twin-pedigree study design, the presence of spouses of MZ and DZ twins further allows us to address the sources of spousal resemblance (Eaves, 1979; Heath & Eaves, 1985; van Grootheest

et al., 2008). Spousal resemblance has been reported for voluntary exercise behavior with a median spouse correlation of 0.28, varying between 0.05 and 0.49 depending on the exercise trait considered (de Moor et al., 2011; Horimoto et al., 2011; Koopmans et al., 1994; Maia et al., 2014; Pérusse et al., 1988; Pérusse et al., 1989; Seabra et al., 2008, 2014). Spousal resemblance can be caused by the preferential mating of regular exercisers with other regular exercisers. This phenotypic assortment is based on a direct preference for similar traits and behaviors in a potential mate. Two well-known alternative mechanisms can cause spousal resemblance with respect to exercise behaviors. Social homogamy refers to mate selection based on sharing the same social milieu. Exercise behaviors scale positively with socioeconomic status, which in turn scales with neighborhood safety, higher mixed land use, and more exercise facilities. All of these may increase exercise behavior, although the effect sizes are rather modest (D'Haese et al., 2014; Gidlow et al., 2006). A third cause of spousal resemblance in exercise behavior is marital interactions. In this case, the longer spouses influence each other, the more similar they become. Note this can go both ways. Non-exercising partners may cause exercisers to swap out exercise activities for other social activities. Alternatively, the regular exercise activities of the exercising partner may become contagious for the non-exercisers. A telltale sign of marital interaction is that longer relationships are associated with greater similarity in exercise behaviors of the spouses.

The causes of spousal resemblance have consequences for the interpretation of genetic and unique and shared environmental variance estimates in an extended twin design. If spousal resemblance is caused by the preferential mating of exercisers with other exercisers it will increase parent-offspring and (non MZ twin) sibling similarity. In the classical twin design, this will cause inflated estimates for shared environmental effects (Falconer & Mackay, 1996). Under social

homogamy, the genetic resemblance between parents and offspring or between siblings is not expected to increase (Eaves et al., 1989). Marital interaction does not have consequences for genetic similarity in the next generation with the caveat that the increase in exercise similarity with marriage duration can be correlated with resemblance at mating: those who are more similar to each other at the start of the relationship may be more likely to remain together (Caspi & Herbener, 1993).

To test the presence and the causes of spousal resemblance, van Grootheest et al. (2008) employed a design first coined by Reynolds et al. (2006) that is illustrated in Figure 5.1. Given data in MZ and DZ twin pairs and their spouses, this method allows one to test for the presence of phenotypic assortment, social homogamy, and marital interaction by comparing the twin-spouse correlations (r_1), co-twin spouse-correlations (r_2), spouse1-spouse2 correlations (r_3), and parent-parent correlations (r_4) in pedigrees with data for twins and their spouses. If phenotypic assortment drives spousal resemblance then the expected pattern of correlations is $r_1 > r_2 > r_3$, and –provided the trait is heritable– the resemblance between co-twin and spouse (r_2) and between the twins' spouses (r_3) should be larger in MZ twins than in DZ twins. If, on the other hand, these correlations are equal ($r_1 = r_2 = r_3$) or conform to ($r_1 = r_2 > r_3$) and the comparison of r_2 with r_3 does not differ for MZ and DZ families this suggests non-random assortment driven by social homogamy (Reynolds et al., 2000). Marital interaction would show as a larger resemblance in the older parent-of-twins couples compared to the younger twins-spouse couples, i.e., r_4 should be greater than r_1, assuming no other generational effects.

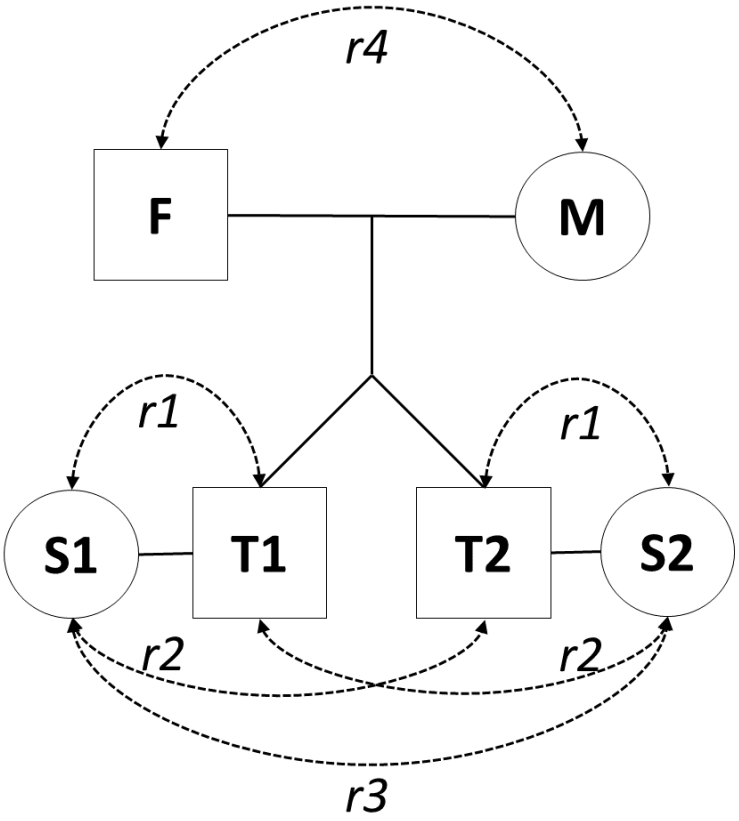


Figure 5.1. Model to analyze sources of spousal resemblance (van Grootheest et al., 2008).

In the NTR, participants list the exact type of individual sports and exercise activities they engage in, in addition to the weekly frequency and duration. Performing the analyses per exercise activity, while arguably desirable, comes with the disadvantage of reducing the number of informative pedigrees. This holds even for the activities that are most frequently endorsed (jogging, fitness, soccer, tennis, swimming). However, exercise activities can be classified in six general domains using the three axes of solitary vs team-based activities, competitive vs non-competitive activities, and activities that are internally paced vs those that are externally paced (van der Mee et al.,

2017). In competitive exercise, only activities with a competitive nature are included, i.e., a type of sport where a team or individual competes against another team or individual to win (e.g., soccer, tennis), whereas in non-competitive exercise only sports without a direct competitive nature are included (e.g., hiking, fitness). Team-based exercise reflects the volume of exercise activities spent in team-based sports, where multiple individuals work together to achieve a shared objective (e.g., field hockey, aerobic dance classes). Solitary exercise includes only the exercise or sports activities performed individually (e.g., jogging, swimming). Externally paced exercise represents the exercise behavior in activities in which the pace is determined by external factors, such as other participants or the weather conditions (e.g., judo, basketball, wind surfing). Internally paced exercise represents the volume of exercise behavior in activities in which the pace is determined by the exercisers themselves (e.g., tour cycling, inline skating, Nordic walking).

Here, we analyzed the total weekly volume of exercise (to be able to compare our results to previous findings), but also the volumes of weekly exercise behaviors specific to the six classes of exercise activity (competitive, non-competitive, team-based, solitary, externally paced, and internally paced). Based on previous twin and family studies, we expected to find a substantial broad sense heritability for the total volume of voluntary exercise behaviors, which should be mainly additive in nature. We further expected heritability of team-based, competitive and externally paced activities (field hockey, soccer) to be higher than that of solitary, non-competitive and internally paced activities (running, swimming). Enjoyment of team-based, competitive and externally paced activities (field hockey, soccer) is associated with physical fitness characteristics and cognitive and motor skills. Voluntary engagement in these activities therefore should reflect the heritability of these traits, which is known

to be high (Schutte, Nederend, Hudziak, Bartels, et al., 2016; Schutte, Nederend, Hudziak, de Geus, et al., 2016). Finally, based on previous findings using a dichotomous trait for exercise participation in a subsample of the current data (de Moor et al., 2011), we expected phenotypic assortment to play a main role in spousal resemblance.

Methods

Participants

The NTR is an ongoing national initiative focused on data-collection in twins and their extended family members. Detailed information on the longitudinal data collection procedures in the NTR can be found elsewhere (Ligthart et al., 2019). Every two to three years, extensive surveys are sent to parents of young twins, adult and adolescent twins and their parents and spouses. Almost all of these surveys included comparable questions on the type and volume of their exercise behaviors. For the present analyses, we used surveys filled out by participants in the age range of 16 to 65 years. Participants may have filled out surveys in multiple waves. For our present purposes, we selected one, if several data at waves were available. Specifically, we selected the wave in which the largest number of family members filled out the survey. For family members who did not participate in the chosen wave, we used the survey closest in time. If the family members differed greatly with respect to survey, we selected the last survey each member filled out. Finally, we only included participants who had at least one other participating family member (in any wave). Spouses of twins here count as family members. Throughout, we use 'spouse' to indicate the "partner with whom you share a lasting and stable relationship (like a marriage)", which was the wording used in the surveys.

The number of nuclear families with exercise data and with at least two family members was 19,543, and the total number of

participants with valid exercise data was 50,690. Table 5.1 describes the composition of these NTR pedigrees and their age characteristics. A more complex pedigree structure in some families was available than these nuclear family relationships. For instance, there was information on 18 grandparents of twins. There were 7 large (≥ 20 family members) pedigrees. The Mendel software fully exploits these complex family structures, providing extra sources of information, for the estimation of A, D, and C beyond that provided by parents, twins and siblings.

	N	M_{age}	SD_{age}
Families[b]	19,543	-	-
Individuals	50,690	38.63	13.65
MZ males	2,560	27.26	12.41
MZ females	5,201	29.88	13.63
DZ males from DZ male pairs	1,680	24.86	11.14
DZ females from DZ female pairs	2,976	27.20	12.00
DZ individuals from opposite-sex pairs	4,293	25.03	10.40
Fathers	12,674	48.21	7.41
Mothers	15,653	45.89	7.70
Brothers	5,268	26.72	11.75
Sisters	7,982	28.34	12.51
Spouse-pairs with offspring[c]	13,881	46.80	7.54
Spouse-pairs without offspring[c]	1,976	36.04	14.54

Table 5.1. Composition of families N: generally denotes number of individuals unless indicated by superscript. b: N=number of pedigrees with at least 2 individuals, c: N=number of spouse-pairs

Voluntary Exercise Behavior

Voluntary exercise behavior was assessed using survey questions in nearly all NTR questionnaires discussed above. Participants were first asked whether they regularly participate in

exercise in their leisure time ("Yes" or "No"). If the response was affirmative they were asked (1) which sport they participate in, (2) how many years they have been doing so, (3) how many months a year they do so, (4) how many times a week on average they do so, (5) how many minutes on each occasion they exercise on average. Metabolic equivalent of task (MET) values were assigned to each exercise activity following Ainsworth et al. (2011). This MET value reflects the energy expenditure of the type of exercise as a multiple of basal energy expenditure (approximately 1kcal/kg/hour). Using these data, all exercise activities were assigned a weekly MET-minutes value by calculating the product of (1) the intensity as the MET value, (2) the weekly frequency, and (3) the average duration in minutes. The total volume of exercise behavior of the participant (METmin) was the sum of the weekly METminutes across all exercise activities. Obligatory exercise such as physical education (PE) classes, or other non-voluntary activity such as biking as a form of transportation, as well as seasonal exercise, such as skiing during winter only, were excluded, as they are not part of regular voluntary exercise behavior. Weekly METmin scores for the exercise domains were calculated in a similar fashion, but with selective inclusion of the types of exercise activities that fit within each of the domains, i.e., competitive or non-competitive team-based or solitary activities, and internally or externally paced activities. The complete list of classification for each type of exercise is detailed in Supplementary Table 5.1. Test-retest reliability of total METmin was 0.82 as computed among 200 individuals who completed the questions about exercise in December 2004 and again 6 months later. Furthermore, using analysis of tracking coefficients of total METmin as well as the scores across the six domains showed moderate to high temporal stability for total volume of exercise across periods varying from 2 to 22 years, ranging from .38 to .77 with a median of .57 (van der Zee, van der Mee, et al., 2019).

Statistical Analyses

To obtain a first impression of familial resemblance for exercise behaviors, correlations were estimated for MZ and DZ twin pairs, for biological sibling pairs, for father–offspring and for mother–offspring pairs and calculating the Pearson's correlation of the METmin variables of the two. This was repeated for every exercise domain. Since we have previously found significant sex, age and sex*age effects on the total volume of exercise behavior (van der Zee et al., 2019), we first used linear regression to correct for age, sex and the age*sex effects on exercise behavior. The Pearson's correlations were based on the regression residuals. Two different spouse correlations were computed: one for parents of twins and one for twins and their spouses, who on average were 9 years younger than the parents of twins.

Variance components attributable to additive genetic factors (A), household influences (C), non-additive genetic factors (D), and unique environmental factors (E) were estimated using the Mendel software package (Lange et al., 2013). In all analyses, age, sex and age*sex were included as fixed effects rather than regressing them out beforehand as we did for the kinship correlations. As detailed in Boomsma et al. (2018), the algorithm, which takes the NTR administrative database as input, provides as output the pedigree structures in standard pedigree format, as used in software, such as Mendel or Merlin (Abecasis et al., 2002). Mendel has an obligatory field to indicate the presence of MZ twins in the pedigree. Optionally, extra fields can be included, to indicate whether persons share the same household within a pedigree. Whereas genetic relations do not change, household sharing can depend on the time of assessment of the exercise behaviors. Dutch offspring typically share a household with their parents and siblings (if any) until around age 18. After about 18 they establish their own households, which can be shared with their own spouse. Sharing a household was defined in four different ways:

Full Household: These household factors contribute to the resemblance of spouses, all parent-offspring pairs, and all twin and non-twin siblings pairs, with twin and non-twin offspring restricted to those under 18. Note that these effects are treated as age and birth-cohort invariant, i.e., they are the same for twins/siblings and their spouses in the pedigree as for the parents of twins/siblings.

Spouse Household: These household factors only contribute to spouse resemblance.

Sib Household: These household factors only contribute to the resemblance of twins and non-twin siblings that have shared a household up to age 18.

Twin Household: These household factors are specific to twins who have shared a household up to age 18.

To test for the sources of spouse resemblance, we used the approach suggested by van Grootheest et al. (2008), as shown in Figure 5.1, and briefly outlined in the introduction. Full maximum likelihood estimation was used in OpenMx (Neale et al., 2016) to obtain the variance-covariance matrix for the six family members expressed in Figure 5.1, separately for MZ, same-sex DZ twin families. Opposite-sex DZ twins were excluded from this analysis as the opposing sex of twins and their spouses would bias the results of r_2 and r_3, whereas in MZ and same-sex DZ twins the comparison is always either between same-sex or different-sex pairs. Age, sex, and the age*sex interaction were included as covariates in this model, and all correlations were constrained to be equal across MZ and DZ twins when no direct comparison was made (i.e., r_1, r_4, the 4 parent offspring correlation, and a residual correlation of parents to offspring-spouses). The correlations of interest obtained from the estimated variance-covariance matrix convey the resemblance between twins and their spouses (r_1), twins and the spouses of their co-twins (r_2), the spouse of twin 1 and the spouse of twin 2 (r_3), and the parents of the twins (r_4).

First, to confirm the presence of spousal resemblance, both the hypotheses $r_1 = 0$ and $r_4 = 0$ were tested by means of a 2-df likelihood ratio test. Second, to test for phenotypic assortment and social homogamy, four hypotheses were tested by means of 1 df likelihood ratio tests: (1) r_1 is equal to r_2, vs. $r_1 > r_2$; (2) r_2 is equal to r_3, vs. $r_2 > r_3$; (3) r_2 in MZ twins equals r_2 in DZ twins vs. r_2 in MZs > r_2 in DZs; and (4) r_3 in MZ twins equals r_3 in DZ twins vs. r_3 in MZs > r_3 in DZs. Significant effects ($p < 0.05$) favoring these four hypotheses were considered evidence for phenotypic assortment. Rejection of these hypotheses is considered to be evidence for social homogamy. The presence of marital interaction was tested by a 1 df likelihood ratio test of $r_4 = r_1$, vs. $r_4 > r_1$.

As a post-hoc test of the influence of phenotypic assortment on the estimates for the narrow-sense heritability (i.e., the additive genetic variance component) we performed mid-parent to offspring regression. We computed the mean parental values for exercise behavior (correct for age, sex and age*sex), and regressed these on the same exercise phenotype in one randomly selected child per family. In contrast to other estimations of narrow-sense heritability, this test is not sensitive to phenotypic assortment, and will not lead to biased estimation of additive genetic variance (Falconer & Mackay, 1996). Because the estimates for A for exercise behavior have been shown to decrease across the adult life span (Vink et al., 2011), we computed mid-parent to offspring regression in four age bins (20-45, 46-50, 51-55, 55-70 years).

Results

Means and standard deviations of the various exercise domains, as well as age for each family relation (parent, child, or twin) split by sex are included in Supplementary Table 5.2. All kinship correlations and the number of pairs in each combination are presented in Table 5.2. For total METhr, monozygotic (MZ) twin

correlations were 0.58, and same-sex dizygotic (DZ) twin and sibling correlations varied between 0.31 and 0.42, with little evidence of sex differences. Opposite sex DZ twin and brother-sister sibling correlations were slightly lower (~0.22). Parent-offspring correlations were between 0.12 and 0.23, which might reflect that parents and offspring share 50% of the additive genetic variance, but in contrast to DZ and sibling pairs no genetic dominance variance. The same pattern of results characterizes the other exercise domains with generally stronger familial resemblance estimates for team-based, competitive, and externally paced activities than for solitary, non-competitive, and internally paced activities.

Correlation	*Npairs*	*Total*	*Team*	*Comp*	*Sol*	*Ncomp*	*PaceI*	*PaceE*
All Spouses	13,002	0.31	0.50	0.42	0.31	0.30	0.32	0.44
Parents of twins	11,879	0.32	0.52	0.45	0.32	0.30	0.32	0.46
Twins - spousesa	1,767	0.28	0.37	0.31	0.29	0.26	0.29	0.33
MZM	1,280	0.58	0.70	0.67	0.50	0.52	0.42	0.67
MZF	2,589	0.58	0.77	0.72	0.56	0.55	0.43	0.72
DZM	824	0.37	0.55	0.45	0.44	0.52	0.32	0.45
DZF	1,468	0.42	0.58	0.53	0.42	0.42	0.28	0.54
DOS	1,857	0.23	0.35	0.24	0.23	0.23	0.20	0.27
Brother-brother	862	0.31	0.51	0.40	0.38	0.46	0.28	0.40
Brother-sister	3,281	0.22	0.30	0.23	0.19	0.19	0.16	0.24
Sister-sister	1,885	0.39	0.58	0.51	0.35	0.35	0.25	0.52
Mother-daughter	11,451	0.23	0.38	0.32	0.22	0.22	0.20	0.33
Mother-son	11,025	0.12	0.16	0.12	0.18	0.20	0.14	0.13
Father-daughter	9,465	0.21	0.34	0.27	0.20	0.20	0.17	0.29
Father-son	9,137	0.20	0.20	0.19	0.22	0.23	0.20	0.19

Table 5.2. Kinship Correlations. MZM: monozygotic male pairs; MZF: monozygotic female pairs; DZM: dizygotic male pairs; DZF: dizygotic female pairs; DOS: dizygotic opposite sex pairs; Total: Total METmin; Team: Team METmin; Comp: Competitive METmin; Sol: Solitary METmin; Ncomp: Non-competitive METmin; PaceI: internally paced METmin; PaceE: Externally paced METmin. [a]: if twins had twin offspring they were used in this row only, and not reused as parents of twins.

Mendel's estimates of genetic and environmental effects

Full results for all exercise domains and models tested are included in Supplementary Table 5.3. In all exercise domains, the best fitting model was the model including additive genetic, non-additive genetic (dominance), household, and unique environmental variance (ACDE) components. This was true across all four household definitions, indicating that heritability estimates are not very sensitive to how a shared household component is modelled. A summary of the model using the full household definition is presented in Figure 5.2, and

a representation of the other models is presented in Supplementary Figure 5.1. Broad-sense heritability of total METmin was 40%. The broad-sense heritabilities of team-, competitive- and externally paced METmin (46%, 43%, and 44%, respectively) were consistently higher than the broad-sense heritabilities of solitary-, non-competitive-, and internally paced METmin (33%, 30%, and 29%, respectively). These increases in heritability are largely explained by an increase in non-additive genetic effects. The effects of shared household components largely depended on its definition, with estimates of C for, e.g., total volume being low in the sibling (4%) and twin (8%) household models, and moderate in the full (20%) and spousal (24%) household models. The C component in the latter two models was driven by the spousal correlations.

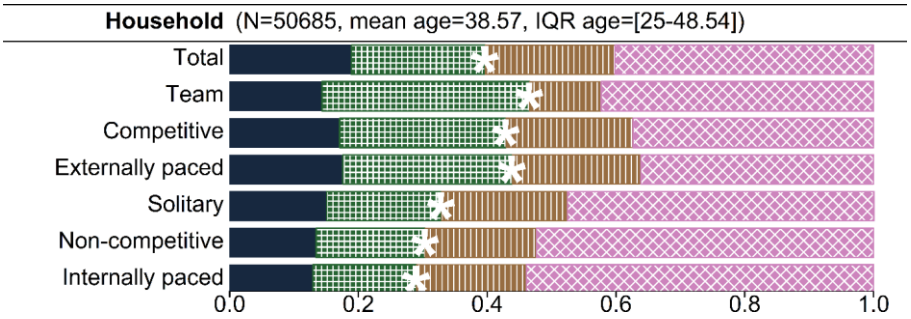


*Figure 5.2. Proportion of variance explained by **additive genetic (A)**, **non-additive genetic (D)**, **shared environmental (C)**, **unique environmental (E)** factors and broad sense heritability (*****) for different classes of exercise. The C component reflects the environment shared by spouses.*

Causes of spousal resemblance

The ML estimate of twin - spouse correlation for total METmin was 0.19 in the subsample used for the analysis of sources of spousal resemblance. The average age of the twins and spouses in this subsample was 37.3 years. Spouse correlations in parents of twins, who were on average 46.8 years old, were significantly higher (r = 0.35, $p = 5.77 \times 10^{-12}$), indicating marital interaction may contribute to spousal resemblance. Table 5.3 contains a summary of the hypotheses tested for the cause of spousal resemblance in the total volume of exercise behavior. Similar tables for the six different types of exercise activities can be found in Supplementary Tables 5.4 to 5.9. Spousal resemblance was significant across all exercise phenotypes ($p's < 0.05$) which is in line with the results from the different household models assessed in Mendel. The hypothesis of marital interaction, i.e., the correlation between spouses living together for a longer time (parents of twins) is larger than the correlation between those living together for a shorter time (twins and their spouses), was also supported across *all* exercise domains. The effect of marital interaction appeared to be particularly strong for team-, competitive- and externally paced sports activities.

Across all exercise domains, the support for social homogamy was greater than the support for phenotypic assortment. Evidence for phenotypic assortment was limited to the greater resemblance of twins and their own spouses (r_1) than the resemblance of twins and the spouses of their co-twin (r_2). We note that our data cannot discriminate between assortment at the time of mate selection and marital interaction as the source of this higher resemblance.

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
$r1 > 0$	1,608	0.20>0	82.46	1	1.08×10^{-19}	√			
$r4 > 0$	10,711	0.35>0	825.05	1	1.94×10^{-181}	√			
$r1 > r2$	1,420	0.20>0.11	17.02	2	2.02×10^{-4}		√	-	
$r2 > r3$	419	0.11>0.08	0.69	2	0.71		-	√	
$r2_{MZ} > r2_{DZ}$	465	0.11>0.11	0.01	1	0.92		-	√	
$r3_{MZ} > r3_{DZ}$	157	0.06>0.12	0.39	1	0.53		-	√	
$r4 > r1$	1,608	0.35>0.20	47.41	1	5.77×10^{-12}				√

Table 5.3. Spousal resemblance and its sources in total volume of exercise. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is $r1 > r2 > r3$, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is ($r1 = r2 \geq r3$), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$. Under marital interaction we expect $r4 > r1$; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with $\alpha = 0.01$.

Influence of phenotypic assortment on additive genetic effects

Regression coefficients obtained in the mid-parent offspring regression on all exercise phenotypes are displayed in Table 5.4. These provide a direct estimate of the narrow-sense heritability and are presented along with the narrow-sense heritability estimates obtained in the full household model analysis in Mendel. The regression coefficients from mid-parent regression in each phenotype were very close to the narrow-sense h^2 estimates from the extended family model, at each age bin tested. This shows that if phenotypic assortment was present, it had little to no influence on the Mendel estimate of the A variance component.

Parental age cutoffs	20-45yrs	45-50yrs	50-55yrs	55-70yrs	
M_{age} parents	42.67	47.85	52.59	59.68	
M_{age} offspring	19.00	20.09	22.08	27.48	
N_{pairs}	725	1369	1463	1238	Mendel narrow-sense h2
Total exercise	0.22	0.18	0.22	0.18	0.19
Team-based	0.21	0.18	0.13	0.14	0.14
Competitive	0.27	0.21	0.19	0.12	0.17
Externally paced	0.23	0.19	0.18	0.13	0.18
Solitary	0.23	0.15	0.19	0.19	0.15
Non-competitive	0.22	0.13	0.18	0.13	0.13
Internally paced	0.21	0.13	0.18	0.10	0.13

Table 5.4 Narrow sense heritability, estimated based on mid-parent offspring regression and from modelling full pedigree data in Mendel. N_{pairs}: Total number of complete pairs (at least one parent and at least one child) used in the regression. Age bins are based on the mean age of the parents. Estimates represent the coefficient from linear regression. Estimated A represents the estimate of A from the extended-family model tested in Mendel.

Discussion

We examined the sources of individual differences in voluntary exercise behaviors in the Netherlands using an extended twin pedigree design. Broad sense heritability estimates for total volume of exercise ranged from 34% to 41% in models using four different definitions of household sharing. Across the various classes of exercise activities, genetic contribution was largest to team-based, competitive and externally paced activities (39% < h^2 < 47%) with lower broad sense heritability for solitary, non-competitive and internally paced activities (26% < h^2 < 34%). These differences in heritability are largely explained by a larger component of non-additive genetic effects for team-based over solitary activities, a 7% increase for competitive over non-competitive activities, and an 8% increase for externally over internally paced activities. Results from the extended twin pedigree design further suggested that a shared sibling household (typically up till age ~18) explains 4-8% of the variance in adult exercise behavior, whereas sharing a household by spouses yielded higher estimates of the C variance component (20-24%) as it had to incorporate the spousal resemblance. The total spouse correlation was 0.32 for total exercise behavior and varied between 0.30 and 0.50 across the specific classes of exercise activities, which was best explained by marital interaction and social homogamy, with weaker evidence for phenotypic assortment.

Our broad sense heritability estimates for voluntary exercise behavior (40%) is 10% lower than the average of 50% obtained in twin registries from multiple countries (van der Zee & de Geus, 2019), which we believe reflects the larger numbers of adolescents and young adults in these studies. Heritability of voluntary behavior has been shown to peak in late adolescence (~65%) (Huppertz et al., 2016) and then gradually decrease across the life span (Vink et al., 2011).

A new finding in the current study is the clear evidence for the contribution of non-additive effects to the broad sense heritability, most prominent for team-based, competitive and externally paced activities. Previous use of extended twin designs for a variety of traits has similarly found higher non-additive and lower additive effects for behavioral traits than the classical twin design (Coventry & Keller, 2005). These non-additive genetic effects provide an explanation for the large discrepancy that has been found in the heritability estimates deriving from parent-offspring/sibling correlations in family studies compared to MZ/DZ twin based estimates in twin studies. Family studies (de Moor et al., 2011; Horimoto et al., 2011; Koopmans et al., 1994; Maia et al., 2014; Pérusse et al., 1988; Pérusse et al., 1989; Seabra et al., 2008, 2014) have systematically reported 11% lower heritability estimates than twin-only studies. Such differences may be partly attributed to the difference in these estimates to non-additive genetic effects, which contribute to resemblances of monozygotic twins and full siblings but not to resemblance of parents and offspring or other relatives such as cousins. The contribution of non-additive genetic effects may also have implications for genome-wide association studies (GWAS), as the traditional GWAS assumes additive genetic effects rather than take non-additivity into account.

Previous studies in adults twins have reported only small shared environmental influences on voluntary exercise behaviors with a meta-analysis estimating them to be about 10% (van der Zee & de Geus, 2019). As most of the studies in this meta-analysis used the classical twin design, the C estimates will largely reflect the sharing of environmental factors by the twins when they were still living at home. In our analyses, shared environmental effects were modeled in four different ways depending on the timing of survey completion: sharing a household as spouses at the time of survey completion, ever sharing a household either as siblings or as parents and offspring, sharing a

household as siblings, or sharing a household as twins. The latter two definitions of shared environmental effects assume that such effects linger after twins and siblings are separated from each other and form their own households (unshared by relatives). Our results using the SibHousehold and TwinHousehold models support the existence of a small impact of sharing a household until late adolescence on the total volume of exercise behavior, with estimates (4-8%) corresponding quite well with the meta-analysis.

In contrast, much larger shared household effects were found (20-24%) when using the SpouseHousehold or FullHousehold models. We surmise that these shared household effects derive largely from the spouse correlations in the twin/sibling and parental generations, which mostly reflected marital interaction across all types of exercise behaviors, including total exercise behavior. The latter contrasts with previous studies that did not find marital interaction for yes/no exercise participation (de Moor et al. 2011). The exercise-activity specific marital interaction suggests that the exercise behavior of a spouse constitutes an important environmental factor for the exercise behavior of a person. We note that marital interaction was weaker for solitary, non-competitive, internally paced activities than for team-based, competitive and externally paced activities. We further note that people tend to engage much less in team-based and competitive exercise activities over time and that solitary, non-competitive, internally paced activities are the largest source of adult exercise participation (van der Zee, van der Mee, et al., 2019). It may simply be difficult to maintain the commitments required for team-based and competitive activities without symmetrical interest and engagement of one's partner.

Marital interaction processes give rise to genotype-environment covariance, which can produce apparent shared environmental influences (Dolan et al., 2014) at the expense of additive

genetic influences. Likewise, if not explicitly modeled, phenotypic assortment may also affect A estimates, although spousal resemblance has to be fairly large to achieve noticeable bias. Phenotypic assortment can be discriminated from social homogamy by comparing the resemblances in twin and co-twin spouses (r_2) and in twin spouses and co-twin spouses (r_3) as a function of zygosity. Phenotypic assortment will make the spouses of MZ twins more alike to their co-twin and to each other than is found in the spouses of DZ twins. If non-random assortment is driven by social homogamy, r_2 and r_3 will *not* differ between MZ and DZ twins. Across all exercise phenotypes, social homogamy was favored over assortment at mate selection as the main source of spousal resemblance, but we note that social homogamy is tested as the null-hypothesis and power for detecting small-sized phenotypic assortment was low. Converging evidence that phenotypic assortment plays a minor role came from the mid-parent offspring regression analyses. The mid-parent offspring regression yields an estimate of narrow-sense heritability that is insensitive to assortment (Falconer & Mackay, 1996). Comparison of these estimates to the heritability estimates from Mendel showed that the latter were not overly biased by phenotypic assortment.

We acknowledge various limitations in our approaches. First, our results are based on twin families, and we do not know if and to what extent determination of exercise behavior in these families differs from that in non-twin families. However, we have noted previously (van der Zee, Schutte, et al., 2019) that there are no differences between the twins and the singleton siblings, and that the prevalence of the MET scores for regular voluntary exercise behavior in the Dutch twin families is very consistent with the prevalence of voluntary leisure time physical activity reported by 5 large-scale studies in Europe. Thus the results of the current study may be generalizable to the Dutch population at large.

Second, the way we carved up the total weekly volume into the six classes of exercise activity (competitive, non-competitive, team-based, solitary, internally paced, and externally paced) is far from ideal. Although not completely overlapping, these classes were not independent, in that the same activity, e.g., soccer would count in the team-based, competitive, and externally paced (paced) weekly exercise volumes.

A third limitation is the assumption that the same genes are expressed across sex and age. Parents and offspring's phenotypes were assessed at different ages. If the genetic correlation across the lifespan would be less than one for phenotypes that are analyzed this would result in a 'too low' parent-offspring correlation, compared to the sibling and DZ pair correlations, potentially causing a wrong conclusion about the presence of genetic dominance. There was some suggestion for a sex limitation model, but less so in the parent-offspring than in the sib data.

Fourth, we used the Mendel software as an efficient way to model the multigenerational familial resemblances, accepting that some effects could not be specified in the genetic model. A different approach would have been to use the 'Cascade' model (Keller et al., 2009). This model allows testing of all variance components as well as specific tests of social homogamy vs phenotypic assortment, and can even be made to incorporate vertical transmission (although we have encountered no actual examples of this). Using the simpler 'Stealth' model on the total volume of exercise behavior we previously found no evidence for vertical (or cultural) transmission suggesting that there is no direct 'copying' of parental exercise habits by the offspring (de Moor et al., 2011). This appears to be true for a large number of behavioral traits (Eaves, 2009) and we assumed here that it also applies to the exercise activities in the specific classes. It could be reasonably argued that for team sports this assumption may not hold.

In summary, regular voluntary exercise behavior, regardless of what type of exercise, is a moderate to highly heritable trait with substantial contribution by non-additive genetic factors. Competitive-team-based and externally paced exercise activities appear to be more heritable compared to non-competitive, solitary and internally paced activities, largely due to an increased effect of these non-additive genetic factors. In childhood and (early) adolescence the environment shared by siblings is a major contributor to exercise behavior (Huppertz et al., 2016) but having shared that environment plays only a minor role in voluntary exercise behaviors in adulthood. In adulthood, the exercise behavior of one's life partner becomes the more relevant environmental factor.

Genome-Wide Association Study of Voluntary Exercise Behaviors in the Netherlands

Under review as van der Zee, MD, Ehli, EA, Pool, R, Hottenga, JJ, van der Mee, DJ, Ligthart, L, Bartels, M, van Dongen, J & de Geus, EJC (2021) Genome-wide association study of voluntary exercise behaviors in the Netherlands.

Abstract

A growing body of literature shows a substantial genetic contribution to the variation in voluntary exercise behavior in humans, but the actual genetic variants for this behavior largely remain to be identified. Here, we used survey data to assess exercise activities in leisure time in 14,626 participants of the Netherlands Twin Register genotyped for 6,367,725 valid genome-wide SNPs to identify novel genetic variants influencing various forms of voluntary exercise behavior. We further used our genome-wide association (GWAS) results to see if these genetic variants replicated the candidate SNPs found in previous genetic association studies using a broader range of physical activity behaviors.

No genome-wide significant SNPs were found for total volume of exercise behavior, team-based, solitary, competitive, and internally or externally paced exercise behavior. For the volume of non-competitive exercise five statistically significant SNPs were detected, one of which, *rs147244851* located in the *Glypican-5* (*GPC5*) gene, was directionally consistent with a previous association to sedentary behavior. The other significant genetic variants (*rs11817437, rs1276262, rs11812727, rs34897771)* were all in an intergenic region on Chromosome 10. Replication of the associations reported in 27 previous studies did not exceed the threshold for multiple testing, but some evidence was found for *rs1858242*, located in an intron of the *TAFA1* gene (*p*-values for internally paced, non-competitive and solitary exercise ≈ 0.002), and *rs7020422*, located in an intergenic region between *RP11-112N13.1* and *RN7SKP191* (*p*-values for externally paced and team-based exercise 0.008 and 0.006 respectively).

Annotation of the significant and suggestive ($p < 5 \times 10^{-6}$) GWAS SNPs suggested red blood cell regulation as an important biological factor in exercise behavior, possibly related to exercise ability. Additionally, results support the inclusion of the volumes of team-

based versus solitary, apart from the total volume of exercise in future GWASs, as these capture two relatively orthogonal latent exercise phenotypes.

Introduction

Technological advancements made since the industrial era have continually decreased the necessity of physical activity (PA), both in work- and home-environments. This has drastically reduced physical hardships of workers as well as the amount of disabilities caused by jobs that demand heavy labor. However, there is also a detrimental effect of the reduced need for PA. Historically, the physically demanding environment caused genes in humans to be specifically selected to cope with high levels of PA. This resulted in a human physiology in which most systems (e.g. metabolic and cardio-vascular) do not function optimally unless regularly activated by sufficient PA (Booth et al., 2008). As a consequence, robust links are found between lower levels of physical activity and many mental and physical health outcomes (Weggemans et al., 2018).

Overall PA has proven difficult to measure reliably through self-report (Prince et al., 2008), but the more salient moderate to vigorous exercise activities voluntarily done in leisure time are more reliably recalled (de Moor et al., 2008; Haase et al., 2004; Stubbe et al., 2006). Regular exercise behavior in leisure time is also a component of PA that is readily modifiable, and a feasible target for behavioral intervention. To increase the success of such intervention, solid understanding of the psychological and biological factors influencing voluntary exercise behavior is needed (Lightfoot et al., 2018). One route to such understanding is employing genetics (Cesarini & Visscher, 2017).

A growing number of scientific research articles has shown a substantial genetic contribution to the variation in voluntary exercise behavior in humans (Bartels et al., 2012; de Geus et al., 2003; de Moor,

Stubbe, et al., 2007; den Hoed et al., 2013; Huppertz et al., 2012; Mustelin et al., 2009; Stubbe et al., 2006; van der Zee et al., 2020). Despite this large heritability, only a small number of genetic variants for exercise behaviors have been identified with studies using a candidate gene approach (Berentzen et al., 2008; Cole et al., 2010; Fuentes et al., 2002; Gielen et al., 2014; Hubáček et al., 2011; Huppertz, Bartels, Groen-Blokhuis, et al., 2014; Jozkow et al., 2013; Kostrzewa et al., 2015; Loos et al., 2005; Lorentzon et al., 2001; Mäestu et al., 2013; Murakami et al., 2011; Richert et al., 2007; Simonen, Rankinen, Pérusse, et al., 2003; Stefan et al., 2002; van der Mee et al., 2017; Wilkinson et al., 2013; Winnicki et al., 2004) of which even fewer (*LEPR, CASR, ACE, MC4R*) have been independently replicated at least once.

A main reason why candidate gene studies on exercise behavior have not been more successful is their potential lack of power. Genome-wide association studies (GWAS) have shown that the heritability of many complex behavioral traits derives from hundreds or thousands of common variants with very small effect sizes (Buniello et al., 2019). Most candidate gene studies may have been underpowered to detect them. Furthermore, meta-analyses of GWAS have shown that the genetic variants that survive replication across multiple cohorts often would not have been selected as likely targets upfront, questioning the selection of candidate genes based on current knowledge of biology. The agnostic approach of a GWAS is therefore the preferred strategy to identify the common genetic variants influencing population variance in complex behavioral traits.

Only a few genome-wide association studies (GWASs) in humans targeted voluntary exercise behavior thus far. de Moor et al. (2009) performed a GWAS in 2,750 unrelated individuals and some suggestive genetic variants related to regular exercise behavior were reported. Though there was partial replication of previous candidate gene-based association to variants in the *CYP19A1* and *LEPR* genes,

none of the genetic variants found in the study withstood the scrutiny of the current genome-wide significance cut-off ($p < 5\times10^{-8}$). Kim et al. (2014) performed a GWAS in a population-based Korean cohort (N = 8,842) and reported no genome-wide significant genetic variants. Lin et al. (2018) performed GWAS and meta-analysis using three different cohorts (total N = 22,219) assessing leisure time physical activity and likewise reported only suggestive genetic variants that would not withstand scrutiny of genome-wide significance. Klimentidis et al. (2018) performed a GWAS in the United Kingdom Biobank (UKB, N_{max} = 377,234) assessing various physical activity phenotypes, among which 'strenuous sports and other exercise', producing the first genome-wide significant genetic variants for voluntary exercise behavior. Shortly after this was published, Hara et al. (2018) published a GWAS on exercise behavior in Japanese adults with different significant genetic variants. Both Klimentidis et al. (2018) and Doherty et al. (2018) performed a GWAS, also in the UKB cohort, on various objectively assessed overall physical activity phenotypes (N_{max} = 91,084), which would partly include voluntary exercise behaviors. The total yield of genetic variants implicated in physical activity phenotypes from candidate gene and GWAS studies is listed in Table 6.1.

As indicated earlier, restriction to the more salient regular exercise and sports-related activities done voluntary and in leisure time may yield a more homogenous phenotype that is the most reliable measured, at least when self-report methods are used. The use of such a more narrowly defined trait may increase the power of genetic association studies (N. Cai et al., 2020). Ideally, each different type of exercise or sports activity would be considered a separate outcome. However, this would rapidly dilute sample sizes as even running, an often-endorsed activity, occurs only in about a quarter of the regular exercisers. Even so, voluntary exercise behavior can be meaningfully decomposed along three psychological dimensions: level of

competitiveness of the exercise activity, whether the activity is performed in a solitary or team setting, and whether the activity is internally paced like jogging or externally paced as occurs in sports activities involving opponents.

In this paper, we used standardized assessment of behavior in leisure time with genotyped participants of the Netherlands Twin Register (NTR, (Ligthart et al., 2019)) to identify novel genetic variants influencing the total volume of voluntary exercise behavior, the volume of exercise behavior spent in either competitive or non-competitive activities, the volume of solitary or team-related exercise behavior, and the volume of internally and externally paced exercise behavior. We further use genome-wide association results to see if they replicate the SNPs identified in the previous studies. Annotation of the significant and suggestive novel GWAS SNPs and the replicated SNPs is then used to attain a better understanding of the biological pathways that lead to individual differences in regular exercise behavior. This knowledge can provide leads for more personalized intervention on this crucial health behavior.

Variant	Previously reported b/OR	Previously reported p	References	Gene	Current GWAS b	Current GWAS se	Current GWAS p
rs116550874	1.72	1.63×10^{-7}	Lin et al. (2018)		No proxy		
rs11121691		> 0.10	Gielen et al. (2014)	MTOR	3.25	15.99	0.84
rs1801133		0.001	Murakami et al. (2011)	MTHFR	-11.76	14.64	0.42
rs1137101		0.016	Richert et al. (2007)	LEPR	12.00	13.78	0.38
		0.007	Stefan et al. (2002)				
rs12405556	OR: 1.24	0.00097	de Moor et al. (2009)	LEPR	4.93	15.48	0.75
	-0.005#	0.33	Hara et al. (2018)				
rs11586310	104.3#	4.38×10^{-5}	Kim et al. (2014)		37.65	33.09	0.26
rs187522732	0.0614	8.20×10^{-6}	Hara et al. (2018)		No proxy		
rs12612420	OR: 1.43	7.61×10^{-6}	de Moor et al. (2009)	DNAPTP6	15.07	18.45	0.41
	0.012#	0.0092	Hara et al. (2018)				
rs1161501	1.73	0.021	Wilkinson et al. (2013)	TPH2	No proxy		
rs1858242	0.031##	3.10×10^{-9}	Doherty et al. (2018)	TAFA1	-15.80	15.52	0.31
rs1248860	0.96	1.10×10^{-13}	Klimentidis et al. (2018)	CADM2	-3.64	13.61	0.79
rs2035562	-0.014	3.90×10^{-9}	Klimentidis et al. (2018)	CADM2	-9.42	14.80	0.52
rs62253088	1.05	1.00×10^{-19}	Klimentidis et al. (2018)	CADM2	-5.92	14.60	0.69
rs6280		0.88	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	DRD3	-11.45	14.89	0.44
	-0.02	0.26	van der Mee et al. (2017)				
rs7650960	-0.10	0.002	Lin et al. (2018)	CASR	-9.47	16.08	0.56
rs112909877[a]	0.11	0.0047	Lin et al. (2018)	CASR	-16.91	16.53	0.31
rs9811123	OR: 0.86#	0.03	de Moor et al. (2009)	CASR	-34.51	15.64	0.03
rs146555373	0.17	0.002	Lin et al. (2018)	CASR	12.51	22.01	0.57
rs1801725		0.01	Lorentzon et al. (2001)	CASR	-28.23	20.99	0.18
rs55716378	0.26	0.0025	Lin et al. (2018)	CASR	-30.61	20.07	0.13

Variant	Previously reported b/OR	Previously reported p	References	Gene	Current GWAS b	Current GWAS se	Current GWAS p
rs8192678		0.001***	Gielen et al. (2014)	PPARGC1A	6.94	14.52	0.63
rs6867384	67.28	3.18×10^{-5}	Kim et al. (2014)		-27.42	22.03	0.21
rs6891956	-66.67#	3.66×10^{-5}	Kim et al. (2014)		-27.42	22.03	0.21
rs11952141	-67.73#	2.92×10^{-5}	Kim et al. (2014)		-28.54	22.04	0.20
rs6880596	-67.66#	3.78×10^{-5}	Kim et al. (2014)		-28.34	22.01	0.20
rs10057067	53.42#	1.67×10^{-5}	Kim et al. (2014)	ITGA1	-0.43	14.77	0.98
rs159544	-0.97#	1.30×10^{-9}	Klimentidis et al. (2018)	CTC-436P18.1	24.00	13.99	0.09
rs26579	0.028##	2.60×10^{-9}	Doherty et al. (2018)	LINC00461	-4.37	14.08	0.76
rs25981	-0.028#,##	3.00×10^{-9}	Doherty et al. (2018)	EFNA5	-3.51	13.84	0.80
rs79173796		8.49×10^{-7}	Lin et al. (2018)	SLC22A4	No proxy		
rs3792874	0.31	8.33×10^{-7}	Lin et al. (2018)	SLC22A4	No proxy		
rs3792877		8.48×10^{-7}	Lin et al. (2018)	SLC22A4	No proxy		
rs3792878		8.61×10^{-7}	Lin et al. (2018)	SLC22A4	No proxy		
rs265981		0.94	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	DRD1	-17.44	14.30	0.22
	-0.011	0.51	van der Mee et al. (2017)				
rs2493869	51.44#	9.39×10^{-5}	Kim et al. (2014)	CDKAL1	26.74	14.13	0.06
rs10946808	-0.97#	9.90×10^{-10}	Klimentidis et al. (2018)	HIST1H2APS3	-22.99	15.45	0.14
rs149943	-0.019#	2.20×10^{-9}	Klimentidis et al. (2018)	OR2W2P	-5.34	19.85	0.79
rs3094622	0.02	1.40×10^{-9}	Klimentidis et al. (2018)	RPP21	No proxy		
rs1265074	-51.96	9.23×10^{-5}	Kim et al. (2014)	CCHCR1	7.62	17.24	0.66
rs2854277	0.032	2.60×10^{-10}	Klimentidis et al. (2018)	HLA-DQB1	No proxy		
rs2267668		0.005	Gielen et al. (2014)	PPARD	30.29	18.27	0.10
rs2076168		0.006	(Gielen et al., 2014)	PPARD	No proxy		
rs6454672	0.62###,#	0.22	Wilkinson et al. (2013)	CNR1	33.14	19.61	0.09

Variant	Previously reported		References	Gene	Current GWAS		
	b/OR	p			b	se	p
rs2764261	1.04	2.00×10^{-11}	Klimentidis et al. (2018)	FOXO3	16.23	14.43	0.26
rs17069951	246.8	3.91×10^{-5}	Kim et al. (2014)	CITED2	8.84	33.41	0.79
rs10252228	0.027	2.20×10^{-9}	Hara et al. (2018)		No proxy		
rs328902	-0.96#	5.50×10^{-10}	Klimentidis et al. (2018)	DPY19L1	-7.61	14.65	0.60
rs7791992	-0.014	5.70×10^{-10}	Klimentidis et al. (2018)		-28.22	14.10	0.05
rs34858520[b]	0.028##	4.20×10^{-9}	(Doherty et al., 2018)	CALN1	5.82	13.83	0.67
rs2519580	71.32#	4.62×10^{-5}	Kim et al. (2014)	GNGT1	26.21	16.87	0.12
rs2519573	-71.19	4.69×10^{-5}	Kim et al. (2014)	GNGT1	26.46	16.88	0.12
rs1043595	-0.014#	4.30×10^{-9}	Klimentidis et al. (2018)	CALU	9.89	15.16	0.51
rs1882094		> 0.10	Gielen et al. (2014)	NRF1	-13.53	15.09	0.37
rs7804463	-0.015#	1.20×10^{-11}	Klimentidis et al. (2018)	EXOC4	8.46	13.75	0.54
rs13243553	-1.04#	9.00×10^{-11}	Klimentidis et al. (2018)	EXOC4	11.35	14.11	0.42
rs940031	-92.64#	4.31×10^{-5}	Kim et al. (2014)		0.98	14.18	0.94
rs11781985	105.8	4.23×10^{-5}	Kim et al. (2014)	RP11-211C9.1	7.93	16.44	0.63
rs2988004	0.013#	4.10×10^{-9}	Klimentidis et al. (2018)	PAX5	-2.21	13.84	0.87
rs11791649	-107.6#	1.30×10^{-5}	Kim et al. (2014)		27.56	14.01	0.05
rs17228531	107.1	1.49×10^{-5}	Kim et al. (2014)		27.56	14.01	0.05
rs7020422	61.84	2.59×10^{-5}	Kim et al. (2014)		34.05	17.47	0.05
rs7023003	-0.0012#	0.75	Hara et al. (2018)		34.43	16.73	0.04
	65.58	4.67×10^{-5}	Kim et al. (2014)				
rs1611115		0.74	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	DBH	No proxy		
rs2519152		0.03	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	DBH	-7.50	13.86	0.59

Variant	Previously reported		References	Gene	Current GWAS		
	b/OR	p			b	se	p
	-0.014[#]	0.39	van der Mee et al. (2017)				
rs564819152[c]	0.028	4.20×10^{-9}	Doherty et al. (2018)	SKIDA1	8.86	14.59	0.54
rs10887741	OR: 0.76[#]	3.81×10^{-6}	de Moor et al. (2009)	PAPSS2	-12.76	14.41	0.38
	0.0044[#]	0.24	Hara et al. (2018)				
rs1819162	-0.17	0.0015	Lin et al. (2018)	ATAD1	17.25	14.87	0.25
rs3781411	-1.06[#]	3.00×10^{-10}	Klimentidis et al. (2018)	CTBP2	-41.97	20.35	0.04
rs1800955		0.36	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	DRD4	No proxy		
rs1800497	OR: 1.16	0.04	de Moor et al. (2009)	ANKK1	-5.33	17.18	0.76
		0.36	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)				
	0.05	0.51	van der Mee et al. (2017)				
rs6275		0.67	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	DRD2	11.12	15.03	0.46
		0.49	Jozkow et al. (2013)				
		0.023	Simonen, Rankinen, Pérusse, et al. (2003)				
rs10507652	113.1[#]	3.95×10^{-5}	Kim et al. (2014)	TDRD3	-13.55	15.20	0.37
rs28524846	-1.09[#]	1.30×10^{-6}	Lin et al. (2018)	MPP5	104.64	49.05	0.03
rs75930676	0.93	2.00×10^{-9}	Klimentidis et al. (2018)	SIPA1L1	No proxy		
rs12438610	-0.22	0.00084	Lin et al. (2018)	GABRA5	No proxy		
rs12902711	-0.17	0.0012	Lin et al. (2018)	GABRG3	-30.61	14.14	0.03
rs12595253	0.44	0.0065	Lin et al. (2018)	GABRG3	-3.87	25.91	0.88
rs8036270		4.61×10^{-5}	de Moor et al. (2009)	GABRG3	No proxy		

Variant	Previously reported		References	Gene	Current GWAS		
	b/OR	p			b	se	p
	0.0003	0.9331	Hara et al. (2018)				
rs72707657	-0.23	0.00052	Lin et al. (2018)	GABRG3	No proxy		
rs62020072	-0.29	0.0016	Lin et al. (2018)	RP11-108K3.1	-15.23	17.56	0.39
rs6493487	OR: 0.86[#]	0.02	de Moor et al. (2009)	RP11-108K3.1	2.85	16.10	0.86
rs743580	0.025	1.30×10^{-9}	Klimentidis et al. (2018)	PML	-6.14	13.78	0.66
rs17817449		0.94	Hubáček et al. (2011)	FTO	0.61	14.05	0.97
rs9939609	OR: 0.45	0.244	Berentzen et al. (2008)	FTO	0.81	14.05	0.95
		0.99	Hakanen et al. (2009)				
		0.26	G. Liu et al. (2010)				
rs166840	-1.03[#]	3.10×10^{-11}	Klimentidis et al. (2018)		9.98	13.87	0.47
rs55657917	0.30[#]	5.00×10^{-12}	Klimentidis et al. (2018)	RP11-105N13.4	-0.58	16.43	0.97
rs2696625[d]	-0.037	3.20×10^{-12}	Doherty et al. (2018)	KANSL1-AS1	-1.30	16.58	0.94
rs1799752		0.28	Fuentes et al. (2002)	ACE	No proxy		
		0.031	Mäestu et al. (2013)				
		0.0001	Winnicki et al. (2004)				
rs8066276	1.44[####]	0.012	Wilkinson et al. (2013)	ACE	20.68	13.92	0.14
rs12451328	OR: 0.88[#]	0.042	de Moor et al. (2009)	ACE	-14.87	14.08	0.29
rs8097348	OR: 1.36	6.68×10^{-6}	de Moor et al. (2009)	RP11-476K15.1	-23.83	16.51	0.15
	0.016[#]	0.854	Hara et al. (2018)				
rs59499656	0.028[#]	1.90×10^{-9}	Doherty et al. (2018)	SYT4	0.82	14.30	0.95
	0.23[#]	2.40×10^{-9}	Klimentidis et al. (2018)				
rs17066829		0.004	Cole et al. (2010)	MC4R	No proxy		
	OR: 1.07	0.2879	de Moor et al. (2009)				
rs138281308		0.0021	Cole et al. (2010)	MC4R	No proxy		
rs7242169		0.005	Loos et al. (2005)	MC4R	9.74	16.47	0.55
rs12462609	83.29[#]	2.04×10^{-5}	Kim et al. (2014)	CACNA1A	6.35	20.93	0.76

Variant	Previously reported b/OR	Previously reported p	References	Gene	Current GWAS b	Current GWAS se	Current GWAS p
rs111901094	-1.04	3.00×10^{-9}	Klimentidis et al. (2018)	GATAD2A	14.12	18.18	0.44
rs429358	0.022[#]	6.10×10^{-13}	Klimentidis et al. (2018)	APOE	34.12	18.43	0.06
rs363035	0.53[###][#]	0.005	Wilkinson et al. (2013)	SNAP25-AS1	39.17	28.48	0.17
rs6074898	-113.9[#]	1.42×10^{-5}	Kim et al. (2014)	MACROD2	1.40	17.88	0.94
rs6022999		0.0036	Kostrzewa et al. (2015)	CYP24A1	15.64	16.21	0.33
rs6092090		0.0038	Kostrzewa et al. (2015)		No proxy		
rs4680		0.09	Huppertz, Bartels, Groen-Blokhuis, et al. (2014)	COMT	11.51	13.90	0.41
	0.012	0.47	van der Mee et al. (2017)				

Table 6.1. Previously published genetic variants for exercise behaviors that were tested for association with total weekly MET-minutes in the current study. A more comprehensive version of this table is included in Supplementary Table 6.2. Previously reported b/OR: Previously reported regression coefficient or odds-ratio (indicated by OR) if explicitly specified in the original publication. If multiple values are reported (e.g., from different samples) the strongest estimate is included; Previously reported p: Previously reported p-value, if multiple values are reported the most significant p-value is included. Gene: Gene associated with the genetic variant, either by genomic location or as suggested by previous publication. Current GWAS: regression coefficient (b), standard error (se) and p-value from the GWAS on total volume of exercise behavior from the current study. [a]: Tagged by rs6795054; [b]: Tagged by rs34547894, [c]: Tagged by rs7084454, [d]: Tagged by rs7210219; [#]: Beta or odds ratio inverted from the original publication to match current studies effect allele if no tagging SNP was used was used; [##]: Beta from the original publication is for sedentary behavior, [###]: The alleles of rs8192678 vary depending on the reference sequence (C>T in GRCh38, G>A in PPARGC1A RefSeqGene), [####]: Tested allele unclear.

Methods

Participants

The NTR is an ongoing research initiative that includes twins and their relatives. Participants (both adults and children) are assessed every two to three years using extensive surveys. Methods and procedures used for survey data collection in the NTR are described in detail elsewhere (Ligthart et al., 2019). Participants were included when they had self-reported on exercise behavior in at least one survey (N = 78,525) and had genome-wide genotype data (N = 21,001). Participants were removed if they were found to be ancestral outliers based on genetic principal component analysis. The resulting sample size was 14,626. Both young (14 < age < 18 years) and adult (age ≥ 18 years) participants from the NTR were included. Two participants that underwent a gender-transformation were excluded from the analysis. A maximal cross-sectional dataset was construed, favoring data obtained in the age range from mid-adolescence to early adulthood, where heritability was shown to peak (de Geus et al., 2014). If repeated waves of survey data were available for a participant, the survey completed when the subject's age was closest to 18 was selected. The mean age of the final sample was 31.26 (SD = 16.82), and 44% was male.

Voluntary Exercise Behavior

The current study focuses on voluntary exercise behavior in leisure time. All NTR participants were asked what types of exercise activities they regularly engaged in, for how many years they had been doing so, how many months a year, as well as weekly frequency, and the average time spent on the activity. The weekly frequency was multiplied by the number of minutes exercised per occasion, and by the age-appropriate Metabolic Equivalent of Task (MET) score for that type of exercise (Ainsworth et al., 2011; Ridley et al., 2008) to derive a

MET-minutes/week score (METmin). The total volume of exercise behavior was computed as the sum of the METmin scores across all exercise activities endorsed. Six additional variables were computed based on the specific type of exercise and sports activities. The first was the total volume spent on competitive exercise activities, i.e. activities where an individual, or team, must compete against another individual, or team for victory. If a subject's only voluntary exercise behavior was spent in non-competitive sports (e.g. cardio, muscle training, or sports education) this was counted as zero METmin (as was done for all non-exercisers). Conversely, all METmin spent on non-competitive activities were also summed into a second variable for total volume spent on non-competitive exercise activities. The third variable was the sum of all METmin spent on team-related activities, in which exercisers work together to achieve a common goal. Conversely, all of the METmin spent on solitary exercise (i.e. horse sports, aiming based sports or fighting sports) was summed to a fourth variable of solitary METmin. The fifth additional variable was externally paced METmin where only exercise activities were included where the tempo or pace of the exercise was determined by external factors, not directly under the exercisers control such as opposing players, or ball movement. Internally paced METmin on the other hand includes only exercise activities where the exerciser is in full control over the execution of the exercise. Mean values of the seven different METmin scores can be found in Table 6.2, separately for males and females and across different age-bins.

	Ages		Total	Team	Comp	E-Pace	Solitary	N-Comp	I-Pace
Total	All	M	887.59	361.47	480.32	470.44	526.12	407.27	344.3
		SD	1126	768	866	855	879	781	726
	<18	M	1138.57	684.54	819.52	812.14	454.03	319.04	222.22
		SD	1158	969	1031	1031	754	642	524
	18-31	M	1116.14	514.38	640.23	619.22	601.76	475.91	395.3
		SD	1283	908	1022	992	997	882	837
	31-50	M	628.09	103.2	207.96	204.26	524.88	420.12	384.59
		SD	900	348	500	497	831	754	711
	>50	M	579.51	51.43	155.72	147.81	528.09	423.8	396.03
		SD	1003	243	453	441	960	866	836
Males	All	M	1120.16	562.59	715.96	698.12	557.57	404.2	370.85
		SD	1329	931	1032	1015	1015	902	857
	<18	M	1426.52	974.05	1150.22	1142.78	452.47	276.3	227.39
		SD	1312	1082	1138	1142	842	701	557
	18-31	M	1493.34	781.92	975.16	930.93	711.42	518.18	483.84
		SD	1508	1062	1199	1147	1198	1060	1054
	31-50	M	713.55	179.07	289.27	278.21	534.48	424.28	403.54
		SD	1073	478	589	573	962	887	862
	>50	M	683.38	89.01	213.95	200.16	594.37	469.43	448.07
		SD	1170	322	528	510	1115	1012	990
Females	All	M	725.82	221.58	316.41	312.07	504.24	409.41	325.83
		SD	926	591	683	680	770	685	619
	<18	M	856.78	401.24	495.92	488.58	455.55	360.87	217.17
		SD	900	743	790	785	656	575	489
	18-31	M	907.2	366.18	454.7	446.56	541.02	452.5	346.26
		SD	1086	772	854	847	859	766	684
	31-50	M	578.26	58.97	160.56	161.14	519.29	417.7	373.54
		SD	777	231	433	442	744	665	606
	>50	M	499.27	22.39	110.73	107.36	476.88	388.54	355.83
		SD	844	151	380	375	817	731	692

Table 6.2. Means (M) and standard deviations (SD) of the various exercise domains: Total volume of exercise (Total), volume of exercise in team-based activities (Team), competitive activities (Comp), externally paced activities (E-Pace), solitary activities (Solitary), non-competitive activities (N-Comp) and internally paced activities (I-Pace).

Genotypes

Prior to the analysis, samples with a call rate below 90% were removed. Samples with abnormal heterozygosity as determined by the inbreeding F value (<-0.10 and >0.10) in PLINK (Purcell et al., 2007) were also removed, as were samples with a mismatch between reported sex and biological sex from DNA. SNPs were filtered if the minor allele frequency (MAF) was smaller than 0.01, or if they differed from Hardy-Weinberg-Equilibrium significantly ($p < 1\times10^{-5}$). Additionally, the dataset was filtered based on missing values per SNP, with any SNP with >5% missing values removed, and on Mendelian errors, with any SNP with > 20 errors removed. All SNP quality control was performed on each of the genotyping platforms (Affymetrix 6.0, Affymetrix Perlegen, Illumina Genomic Screening Array, Illumina Human660, Illumina Omni 1M) separately. Datasets from all platforms were merged, phased, and imputed to the Genome of the Netherlands (GONL, (Genome of the Netherlands Consortium, 2014)) reference using Mach-admix (Fedko et al., 2015; E. Y. Liu et al., 2013).

After this first round of imputation SNPs were converted to best-guess genotypes using Plink 1.90 (Chang et al., 2015), and filtered a second time based on the following criteria: (1) a significant association with genotyping platform ($p < 1\times10^{-5}$), (2) a significant deviation from HWE ($p < 1\times10^{-5}$), (3) a Mendelian-errors-per-SNP rate over 3 SD from the mean, (4) an imputation quality (R^2) < 0.90. After this second phase of quality control the NTR participants that were in the GONL reference set were re-added (N = 286) to the dataset. This dataset was then imputed a second time using the Haplotype Reference Consortium (HRC) version 1.1 reference-set on the Michigan Imputation Server (McCarthy et al., 2016).

Correction for population stratification was first performed by principal component (PC) analysis. PC analyses were performed in the full set of SNPs after QC, and the first 10 genetic principal components

were saved and used as covariates in genetic analyses. Additionally, ethnic outliers were removed from the dataset, defined by principal component (PC) scores in the first 10 PCs laying outside the ranges of the European populations within the HRC reference.

After genotype imputation and PC analysis, participants with both genome-wide SNP data and voluntary exercise data were selected and SNPs were again filtered out if they had: (1) a minor allele frequency smaller then 0.01, (2) a significant deviation from HWE ($p < 1\times10^{-5}$), or (3) an imputation $R^2 < 0.90$.

Phenotypic analyses

For the phenotypic analyses all available exercise data of the maximal cross-sectional set was used, and one participant per family was randomly selected to prevent bias from larger families ($N = 23{,}681$). In order to chart the overlap between the seven different exercise phenotypes a correlation-matrix was computed. Due to the skewed distribution of the exercise phenotypes a Spearman rank-order correlation was used. To more formally assess the underlying factor structure and obtain the number of truly independent dimensions we performed exploratory factor analysis with varimax rotation.

Genetic analyses

A genome-wide complex trait analysis (GCTA; (Yang et al., 2011)) was performed to test the effect of each SNP on all seven exercise behavior dimensions separately. In this analysis the effect of SNPs is tested in the presence of the polygenic effect of all other autosomes, which is computed from the genetic relatedness matrix (GRM). Age, sex, genetic PC's, as well as chip-dummies were included as covariates. The current study used the Leave-One-Chromosome-Out (LOCO) approach. In this approach the GRM used in the GCTA analysis of a specific SNP is computed from all autosomes except the one

containing the SNP that is being analyzed. A genomic inflation factor lambda (λ_{GC}) was calculated using the genomic inflation method, and inflation was assessed through the intercept of LD-score regression (Bulik-Sullivan et al., 2015).

Next, in order to compare the correlation structure found in the phenotypic data to the genetic correlation structure, we attempted LD score regression using LDSC (Bulik-Sullivan et al., 2015). However, none of the SNP heritabilities were significantly larger than zero. As an alternative we therefore correlated the polygenic risk score (PRS) of each individual exercise phenotype with the PRS score of every other exercise phenotype. This was done using generalized estimating equation (GEE) to be able to correct for familial relatedness. By *z*-transforming the PRS standardized beta estimates are obtained from this regression analysis, which can be interpreted as a correlation coefficient.

GWAS annotation

For GWAS annotation we used the online Functional Mapping and Annotation of GWAS tool (Watanabe et al., 2017; Watanabe et al., 2019). For functional mapping, all suggestive SNPs ($p < 5\times10^{-6}$) were considered. SNPs with a high (>13) Combined Annotation Dependent Depletion (CADD, (Rentzsch et al., 2018)) score, indicative of a deleterious effect of the SNP are explored in more detail. Additionally, SNPs overlapping results of any previously reported GWAS for any trait are explored. For gene-based annotation, all suggestive SNPs, and SNPs in high (>0.6) LD with these SNPs were mapped to genes within 10 kilobasepair windows. Expression quantitative trait loci (eQTLs) were used to detect additional genes whose expression was affected. Next, the tissue-specific expression of mapped genes is discussed as well as significant (adjusted *p*-value < 0.05) overlap of mapped genes with existing gene sets and functional maps. Gene-sets based on

genome position (e.g. genes in chr4q31) and proximity (e.g. cancer gene neighborhoods) were excluded. Finally, to assess tissue- and cell-type specific regulatory components of the suggestive SNPs we used Functional element Overlap analysis of the Results of GWAS Experiments 2 (Breeze et al., 2020; Dunham et al., 2015), using the Consolidated Roadmap Epigenomics, 15-state markers reference data. For the latter annotation SNPs are filtered based on LD ($R^2 > 0.8$).

Replication of published genetic variants

A set of 103 genetic variants were identified in previously published genetic studies on exercise behavior and similar traits (see Table 6.1). These 103 SNPs were looked up in each of the seven GWAS results of the current study to assess reproducibility of these previously identified genetic associations. If the previously identified SNP was not present in the current GWAS results an attempt was made at finding a proxy variant by finding SNPs in high LD with the previously identified SNP ($R^2 \geq 0.9$) in the Central European population of 1000 Genomes Phase 3 using LDlink (Machiela & Chanock, 2015). No proxy was found for 21 out of the 103 SNPs. Of the 82 SNPs tested for replication, 6 were in high LD ($R^2 > 0.9$) with another included SNP, therefore using a Bonferroni correction, the *p*-value cutoff for statistically significant replication is $0.05/76 = 6.58 \times 10^{-4}$.

Software

Data manipulation, inspection, correlation and factor analyses were performed using Python 3 (Python Software Foundation, 2014). Genetic quality control and PC analyses were performed in PLINK (Purcell et al., 2007), polygenic risk scores were computed using LDpred (Vilhjálmsson et al., 2015), manipulation of vcf files was performed using vcftools (Danecek et al., 2011). Genetic association analyses were performed with the GCTA software (Yang et al., 2011).

Results

Phenotypic analyses

Individual Spearman rank-order correlations are presented in the lower triangle of Table 6.3. Total volume of exercise demonstrated a moderate to strong correlation to all six components (r_{min} = .47, r_{max} = .71). A high correlation was found between competitive, team-based and externally paced exercise (r_{min} = .85, r_{max} = .99), as well as between non-competitive, solitary and internally paced exercise (r_{min} = .76, r_{max} = .86). Correlations across these two sets of phenotypes were low or even negative (r_{min} = -.17, r_{max} = .07). In accordance with this correlation pattern, two factors were identified by confirmatory factor analysis with an eigenvalue greater than 1. Factor loadings on factor one were high for team, competitive, and externally paced exercise (0.95, 0.99, 0.98 respectively), and factor loadings on factor two were high for solitary, non-competitive and internally paced exercise (0.94, 0.99 and 0.91 respectively). Factor loadings of total volume of exercise were comparable for both factors (0.74 and 0.67). In short, apart from total exercise volume, two additional latent phenotypes could be discerned: volume of team, competitive and externally paced exercise, and volume of solitary, non-competitive and internally paced exercise.

	Total	Team	Comp	E-Pace	Solitary	N-Comp	I-Pace
Total		0.51	0.62	0.61	0.65	0.49	0.55
Team	0.63		0.75	0.76	-0.16	-0.15	-0.05
Comp	0.71	0.85		0.98	0.10	-0.18	-0.04
E-Pace	0.70	0.85	0.99		0.08	-0.18	-0.06
Solitary	0.60	-0.15	0.07	0.06		0.81	0.73
N-Comp	0.47	-0.15	-0.17	-0.17	0.83		0.80
I-Pace	0.47	-0.11	-0.10	-0.11	0.76	0.86	

Table 6.3. Correlations between exercise phenotypes and standardized regression coefficients of regressions between exercise PRSs. The lower triangle contains Spearman correlation coefficients between exercise phenotypes. The upper triangle contains standardized regression coefficients from GEE regressing the PRS for one exercise phenotype on another. All correlations are statistically significant ($p_{max} = 1.23\times10^{-19}$). All PRS coefficients are statistically significant ($p_{max} = 1.15\times10^{-3}$).

Genetic analyses

A total of 1,074 participants from a variety of non-European ancestries were identified to be ethnic outliers. To avoid stratification artefacts, GWAS was restricted to the remaining 14,626 individuals of European ancestry. Of the 41,145,760 SNPs in the full imputed dataset, 33,248,922 SNPs had a minor allele frequency below 0.01, 46,680 SNPs deviated significantly from Hardy-Weinberg equilibrium, and 33,495,123 SNP had a poor imputation quality. After quality control 6,367,725 SNPs remained for subsequent analyses.

Volume of total voluntary exercise behavior

No genome-wide significant SNPs were found for total volume of exercise behavior. A set of 15 SNPs was suggestive at $p < 5\times10^{-6}$ (see Supplementary Table 6.1), the majority of which (10 SNPs) were located on chromosome 9. Both the LD intercept (1.009, SE = 0.007) and λ_{GC} (1.005) suggest no statistical inflation of the results. The Manhattan plot is shown in Figure 6.1.

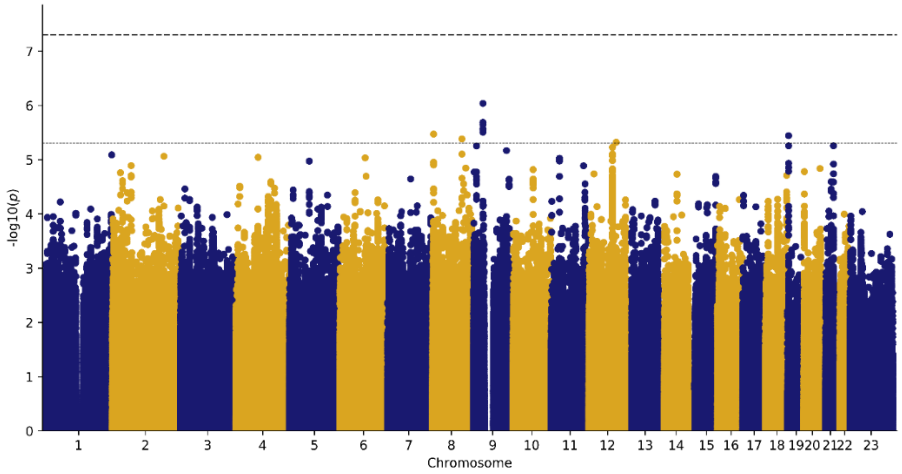


Figure 6.1. Manhattan plot of genome wide analysis results for total volume of exercise. Y-axis represents -log10 transformed p-values. The upper dotted line represents genome-wide significance ($p = 5\times10^{-8}$), the lower dotted line represents the cutoff for suggestive SNPs ($p = 5\times10^{-6}$).

Volume of team, competitive and externally paced exercise

No genome-wide significant SNPs were found for team-based, competitive and externally paced exercise behavior. Sixteen suggestive SNPs were found for team-based exercise, 55 suggestive SNPs were found for competitive exercise, and 28 suggestive SNPs were found for externally paced exercise. Two SNPs were identified in both team-based and competitive exercise, five SNPs were identified in both team-based and externally paced exercise and 15 SNPs were identified in both competitive and externally paced exercise. Three of these SNPs were identified in all three phenotypes. Both the LD intercept, λ_{GC} suggest no statistical inflation of the result in either team-based (LD intercept = 1.009, SE = 0.007, λ_{GC} = 1.002), competitive (LD intercept = 1.007, SE = 0.007, λ_{GC} = 0.999), or externally paced (LD intercept = 1.008, SE = 0.007, λ_{GC} = 0.999) exercise.

Volume of solitary, non-competitive and internally paced exercise

No significant SNPs were found for solitary and internally paced exercise behavior, but for volume of non-competitive exercise five statistically significant SNPs were detected, namely *rs11817437*, *rs1276262*, *rs11812727*, *rs34897771*, and *rs147244851*. Additionally, 16 suggestive SNPs were found for solitary exercise, 721 suggestive SNPs were found for non-competitive exercise, and 449 suggestive SNPs were found for internally paced exercise. Five SNPs were identified in both solitary and non-competitive exercise, six SNPs were identified in both solitary and internally paced exercise, and 408 SNPs were identified in both non-competitive and internally paced exercise. Five of these SNPs were identified in all three phenotypes. Both the LD intercept, and λ_{GC} suggest no statistical inflation of the result in either solitary (LD intercept = 1.008, SE = 0.007, λ_{GC} = 0.993), non-competitive (LD intercept = 1.004, SE = 0.007, λ_{GC} = 1.002), or internally paced (LD intercept = 1.003, SE = 0.007, λ_{GC} = 0.990) exercise. The relatively large number of suggestive SNPs for solitary and non-competitive exercise is explained by a cluster of SNPs on chromosome 4 in high linkage disequilibrium (LD), 628 of the suggestive SNPs for non-competitive exercise are within this cluster and 366 of the suggestive SNPs for internally paced exercise are within this cluster.

Manhattan plots for all exercise phenotypes are provided in Supplementary Figure 6.1.

Cross-phenotype PRS correlations

Standardized β-coefficients resulting from the GEE models of the combination of exercise phenotypes can be found in the upper triangle of Table 6.3. Results of PRS generated with a prior of 0.5 are included as results did not notably differ between priors. The highest coefficients, and therefore greatest correlation was found between the

PRSs for team-based, competitive and externally paced exercise (β_{min} = 0.75, β_{max} = 0.98), as well as between solitary, non-competitive and internally paced exercise (β_{min} = 0.73, β_{max} = 0.81). These results are in accordance with the phenotypic correlation structure and the overlap found in the genetic variants influencing the various exercise phenotypes, and further confirm the two factor structure of team, competitive and externally paced exercise activities versus solitary, non-competitive and internally paced exercise activities.

GWAS Annotation

For each exercise phenotype annotated SNPs, eQTLs, mapped genes, tissue specificity and significant gene set matches are included in Supplementary Tables 6.3-6.7. FORGE2 analysis of tissue- and cell-type regulatory components are included in Supplementary Table 6.8. We highlight the main findings below.

Volume of total voluntary exercise behavior

From the result of total volume of exercise behavior one suggestive SNP with a high CADD score was identified (*rs62620995, p* = 4.12×10^{-6}) which is identified as a missense variant in the *DCSTAMP* gene. Another SNP (*rs10503256*) has previously been implicated in Schizophrenia (Bergen et al., 2012). Based on suggestive SNPs, SNPs in high LD, and eQTL effects of these SNPs, ten genes were included in gene annotation (see Supplementary Table 6.5). None of the ten included genes demonstrated significant tissue-specific expression. The mapped genes were significantly enriched for transmembrane transport of water urea, glycerol and carbohydrates (MSigDB C5; (Liberzon et al., 2011; Subramanian et al., 2005)) due to the inclusion of the *AQP7* and *AQP3* gene. No significant tissue- and cell-type regulatory components were identified at a false discovery rate (FDR) below 0.05.

Volume of team-based, competitive and externally paced exercise

From the results of team-based exercise one suggestive SNP (*rs11569190*, $p = 2.91 \times 10^{-6}$) was in high LD with SNPs previously implicated in hair color, *rs2288634* (Kichaev et al., 2019) and *rs3804772* (Morgan et al., 2018). Six genes were included in gene annotation (see Supplementary Table 6.5). None of the six included genes demonstrated significant tissue-specific expression. The mapped gene set was significantly enriched for genes previously implicated in total hippocampal volume due to the inclusion of *ATP1B3* and *TFDP2* (Buniello et al., 2019).

Annotation of the results of competitive and externally paced exercise yielded highly comparable results as found for team-based exercise. No significant tissue- and cell-type regulatory components were identified at an FDR below 0.05 for team-based, competitive and externally paced exercise.

Volume of solitary, non-competitive and internally paced exercise

From the result of solitary exercise one suggestive SNP (*rs10939529*, $p = 1.66 \times 10^{-6}$) is in high LD with two SNPs with a high CADD score, *rs74654257* and *rs76453796* in an enhancer region on chromosome 4. One suggestive SNP (*rs12679098*, $p = 3.00 \times 10^{-6}$) was found that is in high LD with a SNP (*rs978152*) previously implicated in coronary artery calcification (Wojczynski et al., 2013), and one SNP (*rs2673604*) previously implicated in interleukin-8 levels (Ahola-Olli et al., 2017). Three genes were included in gene annotation (see Supplementary Table 6.5), none of which demonstrated significant tissue-specific expression or significant overlap with previously defined gene sets.

From the results of non-competitive exercise one suggestive SNP (*rs11724752*, $p = 1.02 \times 10^{-6}$) is in high LD with five SNPs with a high CADD score, two of which have been identified as a synonymous variant in the *FREM3* gene (*rs11100802, rs1545437*), one of which located in an enhancer region on chromosome 4 (*rs6858195*), and two of which an intronic variant of the gene *AC107223.1* (*rs35305926, rs11100803*). Additionally this same SNP (*rs11724752*) is in high LD with six other SNPs previously implicated in various red blood cell related phenotypes, specifically: *rs13134327* previously implicated in glycated hemoglobin levels (Wheeler et al., 2017), *rs4835097* previously implicated in blood protein levels (Sun et al., 2018), and *rs4835473, rs13101482, rs4323050* previously implicated in fraction of reticulocytes (immature red blood cells) and reticulocyte counts (Astle et al., 2016). Twelve genes were included in gene annotation (see Supplementary Table 6.5), none of which demonstrated significant tissue-specific expression. The mapped gene set was significantly enriched for genes related to heme (a precursor for hemoglobin) metabolism (Liberzon et al., 2011), as well as for severe malaria and response to cognitive-behavioral therapy in anxiety disorder (Buniello et al., 2019) due to the inclusion of *GYPE, GYPB, GYPA* and *FREM3* genes, with *FREM3* only contributing to severe malaria. Several significant (FDR $q < 0.01$) regulatory components were identified for non-competitive exercise. The vast majority of these were found in cells in a quiescent/low chromatin state, epigenetically the final state of cell differentiation, where replication and differentiation is restricted (Srivastava et al., 2010). These regulatory elements were identified in various tissues such as T-memory cells, mononuclear cells from peripheral blood, aorta, fetal heart and fetal lung fibroblasts. These results can largely be attributed to a chunk of SNPs on chromosome 4, previously mentioned to cause the relatively large number of suggestive SNPs for non-competitive exercise, compared to solitary and internally paced exercise. Post-hoc

analysis of the significant SNPs on chromosome 10, using the Biobank-Based Integrative Omics Studies (BIOS) atlas (Zhernakova et al., 2017) revealed these SNPs to be a significantly associated with the methylation site *cg08891559* in blood, but not with RNA expression levels of any transcript.

From the results of internally paced exercise one suggestive SNP (*rs6844400*, $p = 1.67 \times 10^{-6}$) is in the same LD-block as *rs11100802, rs1545437, rs6858195, rs35305926* and *rs11100803* discussed in the results of non-competitive exercise. Likewise, it is in high LD with *rs13134327, rs4835097, rs4835473, rs13101482* and *rs4323050* previously implicated in various red blood cell related phenotypes. Twenty genes were included in gene annotation (see Supplementary Table 6.5). This set was significantly enriched for genes affecting blood pressure due to the inclusion of *GYPA*, *GPC5, LRP1B,* and *MECOM* (Buniello et al., 2019). No significant tissue- and cell-type regulatory components were identified for internally paced exercise.

Replication of published genetic variants

A summary of replication results in total volume of exercise behavior is included in Table 6.1. Comprehensive results for each of the seven exercise phenotypes are included in Supplementary Table 6.2.

The strongest evidence for replication was found for two genetic variants, depending on the phenotype studied in the current GWAS. First, *rs1858242*, which is located in an intron of the *TAFA1* gene, previously associated with sedentary behavior (Doherty et al., 2018) and in our GWAS negatively associated with internally paced, non-competitive and solitary exercise with *p* values around 0.002 in these phenotypes. Second, *rs7020422*, which is located in an intergenic region between *RP11-112N13.1* and *RN7SKP191*, previously associated with exercise behavior (Kim et al., 2014) and in our GWAS associated

with externally paced and team-based exercise with *p* values of 0.008 and 0.006 respectively.

Publicly available summary statistics

Summary statistics from all seven separate GWAS analyses performed in the current paper are made publicly available at https://doi.org/10.34894/ZLSPFG (van der Zee, 2021).

Discussion

We set out to identify genetic variants influencing voluntary exercise behavior in a representative Dutch sample (*N* = 14,626), using more refined phenotyping of exercise and a GWAS to both detect novel variants and replicate reported variants from previous studies. Novel genetic variants were identified for non-competitive exercise only. The genetic variant *rs147244851* was found to be significantly associated with non-competitive exercise, and was also the top-performing SNP in internally paced and solitary exercise. It is of interest, because it is located in the *Glypacin-5* (*GPC5*) gene which was previously suggested to be a genetic marker for sedentary behavior (Comuzzie et al., 2012), though no evidence was found for association with gene expression. The other significant genetic variants for non-competitive exercise are all in an intergenic region on Chromosome 10 with little evidence for regulatory functions of these SNPs, apart from an association with methylation site *cg08891559*.

Annotation of SNPs that reached a suggestive threshold ($p < 10^{-6}$) for the association with the total weekly volume of exercise behavior identified one of these results (*rs62620995*) to be a missense variant in the *DCSTAMP* gene. The protein product of this gene is considered a master regulator of osteoclastogenesis, the generation of bone-resorbing cells (Kukita et al., 2004). It plays an important role in Paget's bone disease (Mullin et al., 2019) and *DCSTAMP*-knockout mice have

increased bone mass (Yagi et al., 2005). The specific genetic variant *rs6262099* has previously been suggested to have an impact on *DCSTAMP* expression and osteoclast morphology (Laurier et al., 2017). Larger bone mass may support a larger exercise ability or resilience to injury, both of which could contribute to higher levels of exercise behavior.

Based on strong phenotypic and genetic overlap we detected two distinct phenotypes that can meaningfully add to using just the total weekly volume of exercise behavior: team-based, competitive and externally paced exercise, and solitary non-competitive and internally paced exercise. Annotation of suggestive results of team-based, competitive and externally paced exercise suggested genetic overlap with genes that have previously been implicated in hair color. It is important to note, however, that these results originate from a large LD block of SNPS on chromosome 4, spanning over 350,000 base pairs. This LD block spans three genes, namely *ATP1B3*, *TFDP2*, and *GK5,* the first of which was not only associated with hair color (Kichaev et al., 2019; Morgan et al., 2018), but also with red blood cell distribution (Astle et al., 2016). *TFDP2* and *GK5* have been associated with kidney function (Köttgen et al., 2010; Tin et al., 2013) and mean corpuscular volume, a measure of volume of erythrocytes in a blood sample (Kullo et al., 2010), and *GK5* was additionally associated with height (Tachmazidou et al., 2017) and eosinophil count (M. H. Chen et al., 2020). Interestingly, annotation of suggestive results of solitary, non-competitive and internally paced exercise also identified multiple genes associated with a number of red blood cell phenotypes such as glycated hemoglobin levels (Wheeler et al., 2017), blood protein levels (Sun et al., 2018), fraction of reticulocytes and reticulocyte counts (Astle et al., 2016) and heme metabolism (Liberzon et al., 2011)

Because these results derive from suggestive, rather than significant results, they should be interpreted carefully until replicated

in a larger sample. Even so, the results hint at an important role of red blood cells in voluntary exercise behavior. Red blood cells, carrying hemoglobin, are an important factor in endurance capacity (Berglund & Hemmingson, 1987; Ledingham, 1977) through their role in oxygen transport from the lungs to the rest of the body, and through a number of different pathways such as decreasing vascular resistance through the release of nitrous oxide (Stamler et al., 1997). For a comprehensive review of mechanisms through which red blood cells may affect exercise we refer the reader to Mairbäurl (2013). Previous work has indicated that higher endurance capacity can be an important motivational factor in the adoption and maintenance of regular exercise (de Geus & de Moor, 2008; Schutte et al., 2019).

We used our GWAS results to test for replication of previously reported associations with a broader range of physical activity phenotypes. Strongest evidence for replication was found for *rs1858242* in internally paced, non-competitive and solitary exercise ($p \approx 0.002$), as well as for *rs7020422* in externally paced and competitive exercise ($p = 0.006$ and $p = 0.008$ respectively).

To conclude, we identified novel significant genetic variants for non-competitive exercise behavior, and GWAS annotation of different exercise phenotypes suggest that red blood cells are a biological determinant of engagement in regular exercise behavior. Additionally, the current paper strongly suggests that the use of only total volume of voluntary exercise behavior may not suffice in measuring this complex behavioral trait. Separate analyses for each individual sports/exercise activity would be optimal but considering the large sample sizes needed for genetic studies, this is not feasible. The inclusion of the volumes of team-based versus solitary, apart from the total volume of exercise, is proposed as a good compromise, as these capture two relatively orthogonal latent factors, for which large sample sizes can still be easily obtained.

The role of regular exercise behavior in many health-related outcomes is well established (Weggemans et al., 2018). As this behavior is readily modified at an individual level it is considered a prime target for personalized intervention (Buford et al., 2013). However, it is important to keep in mind that the barriers to, and benefits of, exercise differ from individual to individual in part based on differences in biology (Schutte et al., 2019). Meta-analysis of our GWAS results, which are made publicly available, with those of others using these voluntary exercise phenotypes could generate polygenic risk scores for exercise behavior, and ideally for the different types of exercise behavior, that could help tailor interventions to match the individual's preference and biology, improving their effectiveness.

The Role of Personality in Voluntary Exercise Behaviors: a Mendelian Randomization Study

Manuscript in preparation as van der Zee, MD, de Vries, LP, Pool, R, Hottenga, JJ, Ehli, EA, Bartels, M, Dolan, CV & de Geus EJC (2021). The role of personality in voluntary exercise behaviors: a Mendelian randomization study.

Abstract

Personality traits, in particular extraversion, neuroticism and conscientiousness, have been accredited with a causal effect on a key target for health promotion: regular exercise behavior. The causal hypothesis derives entirely from the robust prospective association between these personality traits and exercise found in population studies because the use of traditional intervention methods to test effects of deliberate changes in personality is challenging. Genetic and environmental confounding are difficult to rule out from such observational data. Here, we use triangulation across different genetically informative methods to falsify the causal hypothesis, where failure to falsify across these methods strengthens our confidence in the causal hypothesis.

Using a sample from the Netherlands Twin Register (N = 36,636) we first establish the cross-sectional and longitudinal associations between five personality traits (extraversion, neuroticism, agreeableness, openness, conscientiousness) and the total volume of weekly exercise, as well as the volume of team-based/competitive versus solitary/non-competitive exercise activities. To incrementally control for (genetic) confounding, we assess the regression of monozygotic intrapair differences in both types of traits, use twin-sibling structural equation models to confirm that all latent factors influencing personality also influence exercise, and test simultaneously for causality and genetic pleiotropy using Mendelian Randomization within the context of a Direction-of-Causation twin model.

Strongest support for the causal hypothesis was found for the effect of extraversion on total volume of exercise (one SD extraversion leads to 0.12 SD more exercise) and solitary/noncompetitive activities (effect size 0.08), followed closely by the effects of conscientiousness on total exercise (effect size 0.11) and solitary/noncompetitive

activities (effect size 0.12). Reverse causality testing suggested that these effects were bidirectional. No support was found for (reverse) causal effects of neuroticism, agreeableness or openness on any of the exercise behaviors. These results shed new light on the causality underlying the well-established relationship between personality traits and exercise behavior.

Introduction

Despite the well-established relationship between regular physical activity and cardiovascular disease (Aune et al., 2015), cancer (Kyu et al., 2016), type 2 diabetes (Boulé et al., 2001), as well as mental health (Chekroud et al., 2018) much of the world's population does not meet the guidelines for physical activity (Guthold et al., 2018). This does not reflect a lack of appreciation of the benefits of physical activity or a lack of the intention to be more active. Instead, strong forces appear to intercede between the intention to be more physically active and the actual enactment, a discordance alluded to as the intention-behavior gap (Rhodes & de Bruijn, 2013). These forces can be external to a person, including unfavorable built environments (Bauman & Bull, 2007), but the high heritability of leisure time physical activity (van der Zee, Schutte, et al., 2019) and in particular voluntary exercise behaviors (van der Zee et al., 2020) suggests that person-specific traits are also at play.

One of the (heritable) traits that has long been associated with exercise behavior is personality (see Rhodes and Smith (2006) and Wilson and Dishman (2015) for meta-analysis and review). Personality is mostly described as individual-level tendencies to show consistent patterns of thoughts, actions and feelings, which are observable across ages, genders and cultures (McCrae et al., 2000; Wilson & Dishman, 2015). The most widely used taxonomy for personality traits, and the one we will adhere to in this paper, is the five-factor model (Digman, 1990). In this model an individual's personality is described in terms of scores on five conceptually independent dimensions, namely: (1) neuroticism, (2) extraversion, (3) openness to experience, (4) agreeableness, and (5) conscientiousness. Neuroticism is the propensity to experience negative emotions, extraversion is the preference for social stimulation, openness to experience is a need for novelty, agreeableness is a readiness to agree with others during

interpersonal conflict, and conscientiousness is having good self-discipline and a strong drive to achieve (McCrae & Costa Jr, 2008). Of these personality traits neuroticism, extraversion and conscientiousness are consistently associated with exercise behavior, associations between openness and exercise behavior are found less consistently (Rhodes & Smith, 2006; Sutin et al., 2016; Wilson & Dishman, 2015).

The oldest and still dominant perspective on the personality–exercise association is that it reflects a causal effect of personality traits on the adoption and maintenance of voluntary exercise behaviors (Flemming, 1934). That particular personality traits, mainly low neuroticism, high extraversion and high conscientiousness, are enriched in successful athletic adolescents was taken to mean that a certain personality profile was a requirement to successfully pursue an athletic career (Allen & Laborde, 2014). While this may be true for high-level athletes, it is not necessarily true for the bulk of (recreational) exercisers with more mundane levels of performance. In the population at large, the association between personality and exercise behavior could well be causal but it might also derive from a "third" underlying factor such as socio-economic background and social networks, or pleiotropic genes independently affecting the (neuro)biology crucial to these traits. With regard to the latter, ample evidence has been published that both personality and voluntary exercise behavior are highly heritable traits, with between 40-60% of the variance explained by genetic factors (Bouchard Jr & McGue, 2003; Jang et al., 1996; van der Zee & de Geus, 2019; van der Zee et al., 2020; Vernon et al., 2008). Furthermore, the existence of reverse causality has been hypothesized to play a role too, such that regular exercise participation can modify personality (Stephan et al., 2014).

Establishing a causal role for personality traits in health behaviors is important. It could help shape interventions that better fit

a specific individual. This applies in full to regular exercise. The strong gap between the intention to engage in regular exercise and actual exercise behaviors is in part due to the appetitive and aversive responses to acute exercise, which may be moderated by personality traits, or derive from genetic factors shared between personality and these affective responses to exercise (de Geus, 2020). Surprisingly little empirical work has been done to establish the direction of causality in the personality – exercise association, and even less of this work took genetic confounding into account. A study in a large twin sample that accounted for genetic confounding showed that there may be a causal effect of extraversion on voluntary exercise behavior, whereas the association with neuroticism was mainly through genetic pleiotropy (de Moor & de Geus, 2018). However, like nearly all studies in this field, this study used a rather 'monolithic' approach to exercise behavior. That is, different types of exercise activities over the course of a week are summed into a single weighted exercise score, which is either a frequency (minutes per week), or a weighted average taking duration and the energy expenditure (MET) of the activity into account. Looking at the total METminutes spent per week does not discriminate between e.g. team-based, competitive sports like football or hockey versus solitary non-competitive activities like running, lap swimming or race biking.

As extensively argued by S. Eysenck and Eysenck (1968), personality may not just influence the volume of exercise behavior, but also the type of exercise activities chosen. Facets of personality, such as being outgoing and sociable (extraversion) or emotionally reactive and easily disturbed (neuroticism) may specifically influence the preference for team-based exercise but not impact on solitary activities. The common use of a weekly METminutes score may therefore act to dilute the personality-exercise association. Ideally, METminutes for each individual type of exercise activity would be used

(i.e. fitness, hockey, football, etc.) but doing so would significantly reduce the number of available observations, unless the sample used is very large. As a compromise we previously proposed a classification of all activities in six domains: (1) team-based, (2) solitary, (3) competitive, (4) non-competitive, (5) internally paced, and (6) externally paced exercise activities. Good temporal stability and substantial heritability for the volume of exercise in each of these six domains has been found (van der Zee et al., 2020; van der Zee, van der Mee, et al., 2019). Based on strong phenotypic and genetic overlap these six domains could be further reduced to two distinct phenotypes that capture almost all of the variance in total weekly volume of exercise behavior: (1) team-based exercise activities, which are often competitive and externally paced, and (2) solitary exercise activities which are often non-competitive and internally paced (van der Zee et al., 2021).

In the current study, we use a baseline and five year follow-up measurement of personality and exercise behavior in a large sample of twins and their family members from the Netherlands Twin Register (Ligthart et al., 2019). We use the five-factor model for personality and quantify exercise as total METminutes weekly but also METminutes weekly spent on team-based competitive and solitary non-competitive activities separately. Because it is difficult to 'prove' causality without an intervention, which is hard if not impossible to do on personality traits, we instead use a falsification approach. Across a set of causal inference methods for observational data that apply an increasingly more stringent control of genetic (and early shared environment) confounding, we test whether the predictions from the causal model hold. This triangulation across increasingly more robust tests of causality will demonstrate how the choice of design can affect the conclusion, but also can provide strong support for causality – if the different attempts to falsify the causal hypothesis fail.

We begin by examining the cross-sectional association which should be significant under a causal model, although that does not rule out reverse causality and genetic or environmental confounding. Next, prospective association should also be significant under the causal model, such that personality traits at baseline predict exercise behavior at follow-up. This rules out reverse causality as the (sole) source of the cross-sectional association, but is still compatible with confounding by genetic or environmental factors. To rule out genetic confounding, the intrapair twin difference in personality is regressed on the intrapair difference in exercise behavior in genetically identical (monozygotic) twin pairs. This regression is expected to be significant under the causal model. It could, however, be due to confounding by unique environmental factors. These can be addressed in another design that makes clear testable predictions under the causal model. A bivariate structural equation model of all latent genetic and environmental factors that influence personality and exercise behavior should reveal that the genetic and environmental correlations are both significantly different from zero (Bartels et al., 2012; de Moor et al., 2008).

A final test of causality can be done by using two-sample Mendelian Randomization (MR), as extensively detailed elsewhere (Davey Smith & Ebrahim, 2003). Briefly, if a subject is genetically liable to score high on a personality trait, and that personality trait causes exercise behavior, then the genetic variants associated with personality in large GWAS analyses on the five personality scales (de Moor et al., 2015; Lo et al., 2017) should also be associated with exercise behavior. Reverse causality can be ruled out by showing that the genetic variants that lead to increased exercise behavior are not associated with the personality traits. One important demand for MR is that the genetic instruments used must not suffer from weak instrument bias. A polygenetic risk score (PRS) based on tens of

thousands of genome wide SNPs weighted for their association with a personality trait would best qualify for this. But such a PRS would likely violate another important assumption of MR, namely that the genetic variants used as instrumental variables for personality do not have direct effects on exercise behavior that are independent of personality. However, this 'no pleiotropy' assumption can be relaxed by combining MR analysis with direction-of-causation modeling in a classical twin model (Minică et al., 2018). This method was used here to test causal effects of personality while simultaneously estimating and accounting for concurrent genetic pleiotropy.

Based on earlier findings we expect that, even when accounting for (genetic) confounding, we will not need to reject the hypothesis that personality causes exercise behavior. Specifically we expect high levels of extraversion and conscientiousness, and low levels of neuroticism to cause a higher overall volume of exercise (Allen & Laborde, 2014; Rhodes & Smith, 2006). We further expect that low levels of extraversion and high levels of conscientiousness cause higher overall volume of solitary, non-competitive exercise (H. J. Eysenck et al., 1982), and that high levels of extraversion predict team-based, competitive exercise (Courneya & Hellsten, 1998; Eagleton et al., 2007; Kirkcaldy & Furnham, 1991).

Methods

Sample

The Netherlands Twin Register (NTR) is an ongoing research initiative collecting data in twins and their extended family members. Detailed information on data collection in the NTR can be found elsewhere (Ligthart et al., 2019). Only surveys completed when participants were between 18 and 65 years (inclusive) old were included. If two surveys were available in any given participant, and the surveys were completed two years apart or more, the first survey was

included in the baseline (T1) dataset, and the second survey was included in the follow-up dataset (T2). For participants that had multiple sets of two surveys that were spaced apart by two years or more, first the sets with the lowest amount of missing values for the personality and exercise traits were selected, after which the sets were selected that had the baseline-to-follow-up difference closest to 5 years. If multiples sets meeting these criteria remained, the set where the age of the participant at baseline was closest to 18 was selected. For a graphical overview of this selection process, see Supplementary Figure 7.1. This selection process was used to generate a longitudinal dataset with a minimum difference of two years between baseline and follow-up and a median difference of 5 years (range 2-24).

If only a single survey was available for a participant, personality and exercise data were still included in the baseline dataset, to obtain a maximal sample size for this T1 dataset. This resulted in a T1 dataset which included 36,636 participants (60.1% female) with a mean age of 37.2 (SD = 14.8), and a T2 dataset which included 18,554 participants (62.1% female) with a mean age of 43.2 (SD = 14.0). A total of 17,434 participants had complete data at two time points with mean ages 38.9 (SD = 14.0) at T1 and 43.8 (SD = 14.0) at T2.

Measures

Voluntary exercise behavior

Participants in the NTR were asked detailed questions on what exercise activities they regularly engaged in, and, per activity, the weekly frequency and the average duration of a single exercise bout. Each exercise activity was recoded to a metabolic equivalent of task (MET) intensity level (Ainsworth et al., 2011). The sum over all exercise activities, multiplying weekly frequency by average duration (in minutes) and the corresponding MET value forms the total weekly MET-minutes value.

We further classified all activities as team-based or solitary exercise, competitive or non-competitive exercise, and externally or internally exercise, as we have done previously (van der Zee et al., 2020; van der Zee, van der Mee, et al., 2019). Team-based exercise activities are activities where a group of exercisers work together towards a shared objective, and solitary exercise is always performed by one individual. In competitive exercise an individual or team must compete against another individual or team for victory, in non-competitive exercise this competitive element is absent. In externally paced exercise the pace of the exercise is determined by external factors (factors such as opponents, teammates, a ball, etc.), in internally paced exercise the pace of the exercise is up to the performer (a full mapping of exercise activities to these dimensions is given in Supplementary table I in van der Zee et al. (2020)). Weekly MET-minutes scores for the team-based and solitary activities were calculated in a similar fashion as for the total weekly MET-minutes value, with the difference that only exercise activities conforming to the team and solitary classification respectively were summed.

Next, two factor scores were created from these exercise phenotypes, as we have previously demonstrated these exercise domains to be highly correlated (van der Zee et al., 2021). Factor loadings are included in Supplementary Table 7.1. The first factor represents competitive team-based exercise, and the second factor represents non-competitive solitary exercise. These two factors, as well as total volume of exercise were included in further analyses.

Exercise data were skewed with a part of the participants having zero METminutes, i.e. not engaging in any exercise. However, previous research has shown that transformation of skewed data does not improve the small known bias of the twin models (Derks et al., 2004) that may overestimate unique environmental effects on exercise at the expense of shared environmental effects.

Personality

In most NTR surveys neuroticism, extraversion, conscientiousness, agreeableness and openness were measured using the Dutch translation of the NEO Five-Factor Inventory (Costa & McCrea, 1992; Hoekstra et al., 1996), which contains 12 questions for each personality trait. Item scores were recoded such that a higher score reflected a higher trait-score. In the earlier surveys, only neuroticism and extraversion were measured using a Dutch version of the Eysenck Personality Questionnaire, the ABV (Wilde, 1970). Since the number and scale of items differs between the ABV and NEO questionnaires, the summed personality scores were *z*-transformed on a per-trait per-survey basis. By design, surveys where the NEO is included are preferred over surveys with the ABV. Conscientiousness, agreeableness and openness were treated as missing values in surveys with the ABV.

Genotypic data

Methods used for genotyping and quality control are included in the Supplementary Methods. To compute the PRS for personality in the NTR, to be used as genetic instruments in the MR-DoC model, genome-wide association study (GWAS) summary statistics of all five personality traits from the 23andMe cohort (Lo et al., 2017) were used. For neuroticism, additional summary statistics were used from the Genetics of Personality Consortium (GPC) meta-analysis (de Moor et al., 2015), leaving out the NTR cohort, and from item number 20127 in the UK Biobank (Sudlow et al., 2015). The GPC assessed neuroticism in 23 cohorts using various surveys and the items were harmonized using an item-response theory approach (van den Berg et al., 2014).

For reverse causality testing with MR-DoC, we also computed a PRS for voluntary exercise behavior in the NTR using the GWAS

summary statistics in the UKB for strenuous sports and other exercise (SSOE) as described in full by Klimentidis et al. (2018).

Statistical Analyses

To test our causal hypotheses we used a number of different statistical methods aimed to falsify the causal hypothesis. For each analysis, the exercise phenotypes were *z*-transformed on a per survey basis as we had done for the personality traits. By using z-transformed data for personality and exercise, the estimates for the causal effect sizes become comparable across the different types of analyses. All data manipulation, plotting and analyses were performed in Python 3.7 (van Rossum & Drake Jr, 2009) unless stated otherwise.

Cross-sectional and prospective associations

First, both cross-sectional (T1 and T2) as well as longitudinal (T1→T2) phenotypic associations are computed for the 15 pairings of personality-exercise traits as the standardized beta from the regression of exercise traits on the personality trait, using age and sex as covariates, and using generalized estimating equations (GEE) to correct for family relatedness.

MZ intrapair regression of differences in exercise behavior on differences in personality

The next model controlled for confounding of genetic and shared environmental factors, by focusing on the intrapair difference scores of MZ twins. The scores of the second-born twin were subtracted from the scores of the first-born twin to obtain a within-pair difference score for the 5 personality traits and 3 exercise phenotypes. The within-pair difference scores of the personality traits were then regressed on the within-pair difference scores in the exercise behaviors, both at T1 and T2 using ordinary least-square regression.

Twin-sibling models

Next, OpenMx version 2.18 (Neale et al., 2016) in R version 4.0 (R Core Team, 2013) was used to fit various twin-sibling structural equation models to assess genetic and environmental overlap underlying the phenotypic association between personality and exercise. Using the fact that monozygotic (MZ) twins share all of their genetic material, and dizygotic twins (DZ) twins as well as siblings share half of their genetic material on average, these models can decompose the variance and covariance of traits into (co)variance explained by genetic and environmental components. Four distinct components are used in the models for the current paper. First, the additive genetic (A) component, which reflects variance explained by effects of independent genetic variants. Second, the non-additive genetic (D) component, which reflects interaction effects between genetic alleles. Third, the shared environmental (C) component, which reflects the effect of the environment shared within a family, but can also be affected by various sources of spousal resemblance (see van der Zee et al. (2020)). Last, the unique environmental component (E) which reflects environmental effects unique to an individual, as well as measurement error.

For model identification reasons, the effects of C and D cannot be estimated simultaneously using only twins and siblings. Therefore, an ADE and an ACE model were fitted first based on the pattern of twin correlations. Quantitative sex-differences in the variance components were tested by constraining them to be equal across sexes in reduced models. Finally, the comparative fit of the even more reduced AE model was tested. The fit of nested models was compared using a log-likelihood ratio test (LRT), as the difference in minus two times the log-likelihood (-2LL) between nested models, which follows a χ^2 distribution where the difference in the number of estimated parameters equals

the degrees of freedom. If the p-value from a χ^2 test between nested models is greater than the alpha it is indicative of a non-significant drop in overall fit of the model, hence the more parsimonious model is selected for further analyses. Since an AE model was found to be the most parsimonious model for the majority of phenotypes we decided to use an AE model for the bivariate twin models below, in order to make the r_A and r_E estimates from all models more readily comparable.

Two bivariate twin-sibling models were fitted where genetic and environmental contributors to variance and covariance of two traits are assessed. We first fitted this model in a cross-sectional scenario assessing variance and covariance of personality at T1 and exercise at T1, and as a replication, of personality at T2 and exercise at T2. We then used the longitudinal structure of the data by selecting personality traits at T1 and the exercise phenotypes at T2 (de Moor et al., 2008), as displayed in Figure 7.1a. The procedure is comparable to the bivariate models mentioned previously, but now the overlap is assessed between the variance components influencing personality at baseline and exercise behavior at follow-up. Under the causal hypothesis, the genetic (r_A) and environmental (r_E) correlation between personality at baseline and exercise behavior at follow-up should be non-zero and a model fixing them to zero should provide a significant deterioration in fit.

The longitudinal twin-sibling models were repeated to also test for reverse causality. This is done by reversing the phenotypes at T1 and T2. At T1 we now used the exercise phenotypes and at T2 the personality traits.

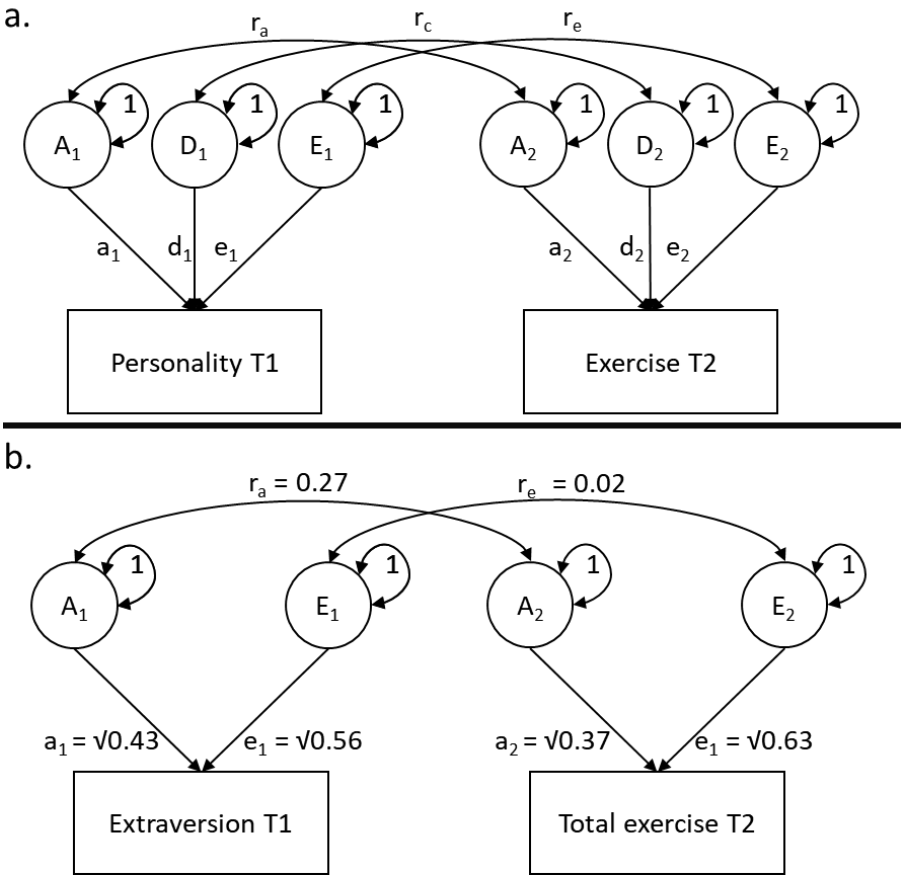


Figure 7.1. Longitudinal bivariate biometric model. Where additive genetic (A), non-additive genetic (D), and unique environmental (E) latent factors are fitted, as well as their correlations (r_a, r_d and r_e respectively). Variances of the latent factors are fixed at 1. The top panel (a) represents the full model, the bottom panel (b) includes the results of this model for extraversion and total volume of exercise.

Genetic correlation based on GWAS summary statistics

Apart from using twin models, the genetic correlation (r_g) between personality and exercise behavior can also be computed based on GWAS summary statistics using LD score regression (Bulik-Sullivan et al., 2015). We therefore re-assessed the genetic correlations using the available GWAS summary statistics for the five NEO personality traits and SSOE. Genetic correlations and SNP heritabilities were calculated using LDScore regression 1.0 (Bulik-Sullivan et al., 2015) in Python 2.7 (van Rossum & Drake Jr, 1995).

Polygenic risk scores (PRS)

Since multiple summary statistics are available for neuroticism, the PRS was computed once on only the 23andMe sample (labeled PRS Neuroticism), and once on a meta-analysis of 23andMe, UK-Biobank, and leave-NTR-out GPC (labeled PRS MetaNeuroticism). This meta-analysis was performed using sample size weighted GWAS meta-analysis as described by Baselmans et al. (2019). PRS were computed in the NTR using the available summary statistics data on self-reported exercise SSOE from UKB and the five NEO personality traits from 23andme, as well as on MetaNeuroticism (23andMe, UK-Biobank, and leave-NTR-out GPC). The PRS for each trait was calculated using LDpred 1.0 (Vilhjálmsson et al., 2015) in Python 3.7 using a range of priors (1.0, 0.5, 0.3, 0.2, 0.1, 0.5 and 0.01). Only SNPs were included that had a minor allele frequency greater than 0.01, an imputation quality above 0.8, a call rate above 95%, and did not significantly deviate from Hardy-Weinberg equilibrium ($p > 10^{-6}$).

Because our MR-DoC model is robust against pleiotropy, we used the personality PRS as our genetic instrumental variables. The personality traits in the NTR were first regressed on the PRS based on the consortium GWAS for the corresponding personality traits to verify these PRS were strong genetic instruments in the NTR. Similarly, the

total weekly METminutes and competitive/team-based and non-competitive/solitary activities were regressed on the PRS based on the UKB GWAS for SSOE. Regressions were fitted using GEE to correct for familial relations, and included age, sex, genotyping platform and the first ten genetic principal components as covariates.

Mendelian Randomization – Direction of Causation model (MR-DoC)

After verification of the PRS as strong genetic instruments, the PRS with the highest proportion of explained variance for each personality trait was used in MR-DoC models that leverage the twin-family structure of our data to minimize the effects of potential assumption violation (Minică et al., 2020; Minică et al., 2018). A visual representation of the MR-DoC model is included in Figure 7.2a. We used T1 data exclusively where we had the largest number of complete twin pairs with genotype data (needed for the PRS computation). To test for potential causal effects of personality on exercise behavior taking into account the presence of genetic pleiotropy, the fit of the MR-DoC model with a direct causal effect from a personality trait on an exercise phenotype in the twin participants is compared to the MR-DoC model where the causal path is fixed to 0. If the latter model produces a statistically significant deterioration of fit this is indicative of a causal relationship. The MR-DoC model also directly estimates and accounts for the pleiotropic effects of the PRS for personality traits on the exercise phenotypes.

The MR-DoC analyses were repeated to also test for reverse causality. This was done in a model that uses a direct causal path from an exercise phenotype to a personality trait in the twin participants, and the PRS for SSOE as the genetic instrument for the exercise phenotypes. MR-DoC models were fitted using OpenMx version 2.18 in R version 4.0.

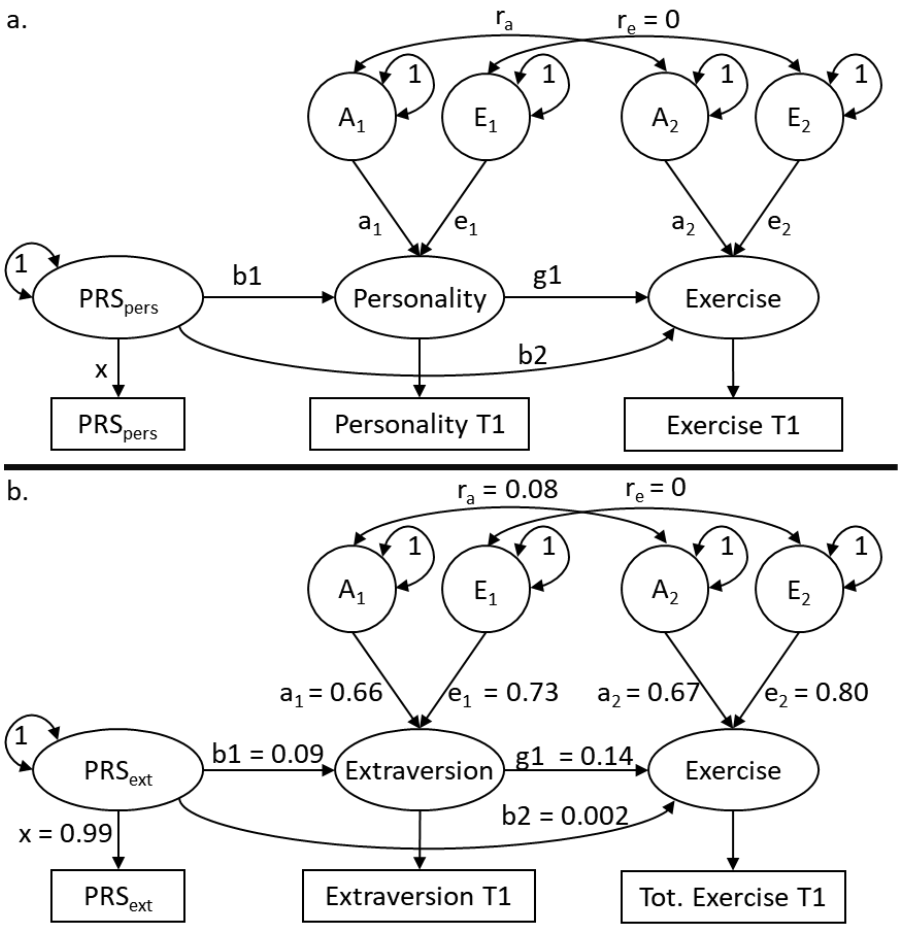


Figure 7.2. The MR-DoC model. This model includes additive genetic (A) and unique environmental (E) factors, genetic correlation (r_a), as well as a polygenic score for personality (PRS_{pers}). For identification purposes the unique environmental correlation is not estimated. b1: the direct effect of this PRS on personality, b2: the pleiotropic effect of this PRS on the outcome, g1: the direct estimate of the causal effect. x: Path-estimate of the latent factor for the PRS on the observed PRS (should be close to 1). The top panel (a) represents the base model, the bottom panel (b) includes the results of this model for extraversion and total volume of exercise.

Triangulation of the various tests of the causal hypothesis

Under the **causal hypothesis**, all the aforementioned methods should produce a specific pattern of results for the personality trait - exercise phenotype pairs. Failure to produce the expected pattern of results decreases the confidence that the causal hypothesis is true for that pairing. First, cross-sectional associations between personality and exercise behavior at each time point, as well as longitudinal associations from baseline to follow-up must be significant and in the expected direction. Second, the MZ intrapair differences in the exposure must be significantly associated to the MZ intrapair differences in the outcome at both time points in the expected direction. Third, any variance component influencing the personality trait must also influence the exercise phenotype, meaning that additive genetic correlations AND unique environmental correlations must all be significantly different from zero, and have the expected sign. Fourth, the direct path from personality to exercise phenotype in the MR-DoC analysis should be significant and in the expected direction.

Our methods also provide an indication for the plausibility of two alternative hypotheses: (1) the association is explained by **genetic pleiotropy**, which would be supported by a significant pleiotropic effect in the MR-DoC model, and significant genetic correlations in twin and GWAS data; (2) **reverse causality** which would be supported by longitudinal twin-sibling models showing non-zero genetic and environmental correlations between exercise behavior at baseline and personality at follow-up and a significant direct causal path from an exercise phenotype to a personality trait in the MR-DoC analysis.

Statistical significance and multiple testing

The number of different statistical tests performed varies somewhat per analysis, but globally we test 15 pairings of the three

exercise phenotypes and five personality phenotypes. Per analysis, alpha was set to 0.05/(3*5) = 0.0033. This may be considered a liberal alpha, as we used seven different methods to test causality. However, we summarize the support from the various methods by a yes/no dichotomy in a closing table, separately for all of the personality trait exercise phenotype pairings. It is important to note that to support causality in a single pairing, seven falsification tests had to be passed (cross-sectional, longitudinal and MZ intrapair associations, all (non)additive genetic and unique or shared environmental correlations > 0 in both the T1 and T2 cross-sectional and the T1→T2 longitudinal bivariate models, the MR-Doc causal path g1 > 0).

Results

Descriptive statistics from the T1 and T2 datasets, as well as complete longitudinal twin pairs of different zygosities are shown in Table 7.1.

	T1			T2			Complete longitudinal pairs				
	M	SD	N	M	SD	N	MZM	DZM	MZF	DZF	DOS
age	37.2	14.8	36636	43.2	14.0	18554	462	240	1288	564	300
Total	736.3	1117.1	36557	707.4	1040.5	18531	462	239	1288	564	300
Competitive	309.0	721.2	36557	222.6	575.8	18531	462	239	1288	564	300
Non-competitive	427.5	844.2	36557	485.0	856.3	18531	462	239	1288	564	300
Team-based	216.5	619.8	36557	135.9	458.4	18531	462	239	1288	564	300
Solitary	519.9	929.0	36557	571.7	929.4	18531	462	239	1288	564	300
Externally paced	301.6	715.5	36557	215.3	562.9	18531	462	239	1288	564	300
Internally paced	382.9	805.2	36557	447.9	829.6	18531	462	239	1288	564	300
Competitive/Team	0.0	1.1	36557	-0.1	1.0	18531	462	239	1288	564	300
Non-competitive /Solitary	0.0	1.0	36557	0.1	1.0	18531	462	239	1288	564	300
Neuroticism	29.3	9.0	26513	28.5	8.9	15445	370	169	1129	464	221
Extraversion	41.8	9.1	26513	41.4	9.0	15445	370	169	1129	464	221
Openness	37.8	5.7	26513	37.8	5.7	15445	370	169	1129	464	221
Conscientiousness	44.6	5.3	26513	45.0	5.1	15445	370	169	1129	464	221
Agreeableness	43.7	4.8	26513	43.9	4.7	15445	370	169	1129	464	221

Table 7.1. Means (M),standard deviations (SD) and number of samples (N) in the baseline dataset (T1), and the follow-up dataset (T2). Exercise variables (Total, Competitive, Non-competitive, Team-based and Solitary) are given in METminutes/week. Competitive/Team and Non-competitive/Solitary are the principal components. For the purposes of this table neuroticism and extraversion scores from the ABV were transformed to match the range of the NEO-FFI. MZM: Monozygotic male pairs, DZM: Dizygotic male pairs, MZF: Monozygotic female pairs, DZF: Dizygotic female pairs, DOS: Dizygotic other-sex pairs.

Cross-sectional and prospective associations

All cross-sectional, and longitudinal associations between personality and exercise were assessed using regression models, results of which are included in Table 7.2. Strongest associations between personality and exercise were observed between extraversion and all three exercise phenotypes, both cross-sectionally (regression coefficients between 0.11 and 0.14) and longitudinally (regression coefficients between 0.10 and 0.13). Neuroticism showed the expected negative associations and conscientiousness showed the expected positive associations. No associations were found between agreeableness and any of the exercise behaviors. Longitudinal and cross-sectional associations between openness and total exercise were not significant, possibly due to opposite relationships between openness and competitive/team exercise (coefficients between -0.03 and -0.06) and openness and non-competitive/solitary exercise (coefficients between 0.05 and 0.07).

		Baseline T1	Follow-up T2	Personality T1 Exercise T2			Reverse Longitudinal Exercise T1 Personality T2
	Exercise	**b (SE)**	**b (SE)**	**b (SE)**	Exercise		**b (SE)**
Neu	Total	-0.10 (0.01)	-0.08 (0.01)	-0.08 (0.01)	Total	Neu	-0.07 (0.01)
	Competitive/Team	-0.10 (0.01)	-0.08 (0.01)	-0.07 (0.01)		Ext	0.13 (0.01)
	Non-competitive/Solitary	-0.07 (0.01)	-0.06 (0.01)	-0.06 (0.01)		Ope	0.00 (0.01)
Ext	Total	0.14 (0.01)	0.15 (0.01)	0.13 (0.01)		Agr	-0.02 (0.01)
	Competitive/Team	0.11 (0.01)	0.11 (0.01)	0.10 (0.01)		Con	0.07 (0.01)
	Non-competitive/Solitary	0.12 (0.01)	0.14 (0.01)	0.11 (0.01)	Competitive /Team	Neu	-0.07 (0.01)
Ope	Total	-0.00 (0.01)	0.04 (0.01)	0.01 (0.01)		Ext	0.09 (0.01)
	Competitive/Team	-0.06 (0.01)	-0.03 (0.01)	-0.05 (0.01)		Ope	-0.05 (0.01)
	Non-competitive/Solitary	0.06 (0.01)	0.07 (0.01)	0.05 (0.01)		Agr	-0.00 (0.01)
Agr	Total	-0.00 (0.01)	-0.02 (0.01)	-0.01 (0.01)		Con	0.02 (0.01)
	Competitive/Team	0.01 (0.01)	-0.01 (0.01)	0.01 (0.01)	Non-competitive /Solitary	Neu	-0.05 (0.01)
	Non-competitive/Solitary	0.01 (0.01)	-0.01 (0.01)	0.00 (0.01)		Ext	0.11 (0.01)
Con	Total	0.08 (0.01)	0.07 (0.01)	0.07 (0.01)		Ope	0.04 (0.01)
	Competitive/Team	0.02 (0.01)	0.03 (0.01)	0.01 (0.01)		Agr	-0.01 (0.01)
	Non-competitive/Solitary	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)		Con	0.08 (0.01)

*Table7.2. Left hand panel: Cross-sectional (T1, T2) regression coefficients (**b**) and standard errors (**SE**) from generalized equation estimation (gee) regression between personality traits (**Neu**roticism, **Ext**raversion, **Ope**nness, **Agr**eeableness and **Con**scientiousness) and exercise behaviors (**Total** volume of exercise, **Competitive/team** factor score and **Non-competitive/solitary** factor score). Middle panel: Longitudinal (T1->T2) regression coefficients (**b**) and standard errors (**SE**) from GEE models between personality and exercise behaviors. Right hand panel: Reverse regression coefficients (**b**) and standard errors (**SE**) from GEE models between exercise behaviors and personality. Non-significant estimates are in italics*

MZ intrapair regression of differences in exercise behavior on differences in personality

Results of all MZ intrapair difference analyses are included in Table 7.3, and the MZ intrapair regression of extraversion on total exercise is also visually presented in Figure 7.3. Significant coefficients at T1 were identified between neuroticism and total exercise (β_{T1} = -0.09, SE = 0.02), extraversion and total exercise (β_{T1} = 0.09, SE = 0.02), extraversion and competitive/team-based exercise (β_{T1} = 0.06, SE = 0.02), conscientiousness and total exercise (β_{T1} = 0.10, SE = 0.02), and conscientiousness and non-competitive/solitary exercise (β_{T1} = 0.08, SE = 0.02). At T2, directionally consistent results were found, but coefficients did not as often reach significance because less twin pairs were available than at T1. Significant coefficients at T2 were identified between extraversion and total exercise (β_{T2} = 0.10, SE = 0.03) and extraversion and non-competitive/solitary exercise (β_{T2} = 0.09, SE = 0.03).

		Baseline T1			Follow-up T2	
		N_{pairs} = 2,359			N_{pairs} = 1,491	
		beta	**SE**		**beta**	**SE**
Neuroticism	Total	-0.094	0.021		-0.043	0.026
	Competitive/Team	-0.058	0.020		-0.012	0.029
	Non-competitive/Solitary	-0.032	0.022		-0.039	0.028
Extraversion	Total	0.093	0.022		0.098	0.026
	Competitive/Team	0.062	0.021		0.049	0.028
	Non-competitive/Solitary	0.055	0.022		0.091	0.028
Openness	Total	0.012	0.024		-0.006	0.028
	Competitive/Team	-0.035	0.022		-0.035	0.030
	Non-competitive/Solitary	0.053	0.024		0.032	0.030
Agreeableness	Total	0.012	0.021		-0.021	0.026
	Competitive/Team	0.005	0.020		0.006	0.028
	Non-competitive/Solitary	0.008	0.021		-0.038	0.027
Conscientiousness	Total	0.101	0.021		0.052	0.026
	Competitive/Team	0.041	0.020		0.019	0.029
	Non-competitive/Solitary	0.077	0.021		0.076	0.028

*Table 7.3. Regression of monozygotic intrapair differences in personality traits (Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness) on the intrapair differences in exercise behaviors (Total volume of exercise, Competitive/team factor score and Non-competitive/solitary factor score). Regression estimates (**beta**) and standard errors (**SE**) are given separately for baseline (T1) and follow-up (T2) data sets. Non-significant estimates are in italics.*

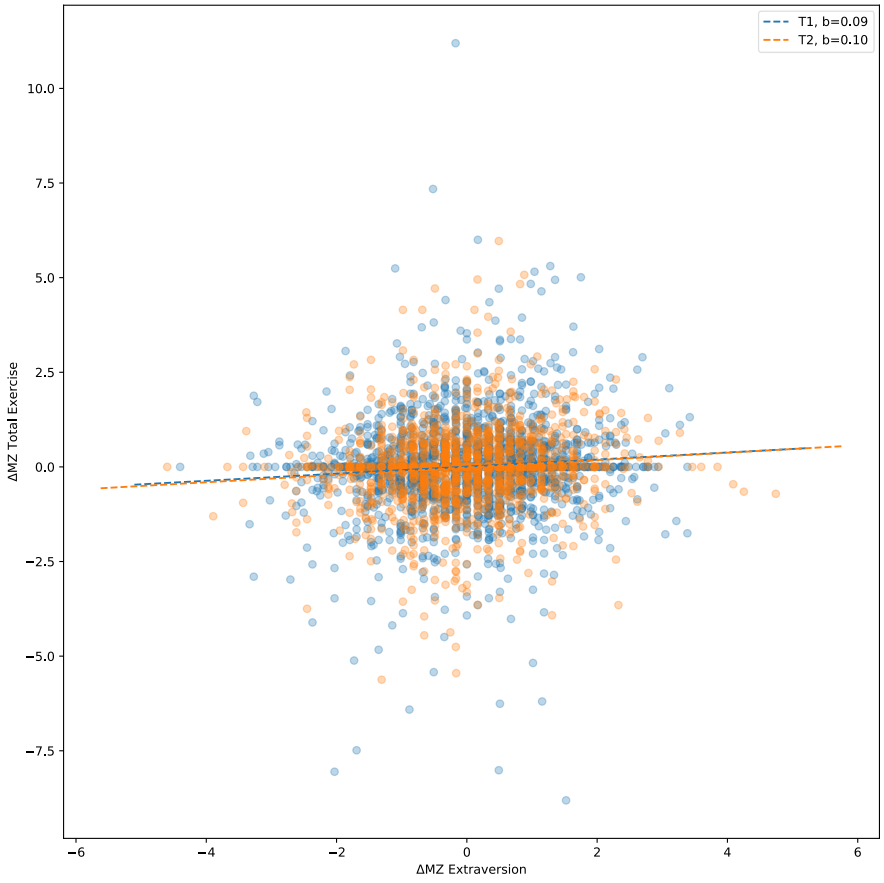


Figure 7.3. Within monozygotic twin-pair regression. Blue dots represent the association between the differences in total exercise between MZ twins (ΔMZ Total Exercise) and the differences in extraversion (ΔMZ Extraversion) at baseline (T1). Orange dots represent follow-up (T2). The lines displayed in this figure are the fitted regression lines, and are colored similarly, blue for T1, orange for T2.

Twin-sibling models

A summary of the results of the cross-sectional and longitudinal bivariate twin-sibling analyses, all using AE models, are included in Supplementary Tables 7.2 and 7.3 respectively. For illustration, the path estimates for the longitudinal model of extraversion and total exercise are depicted in Figure 7.1b.

Results of the cross-sectional bivariate models indicate a significant genetic correlation between neuroticism and any exercise, extraversion and any exercise, openness and competitive/team-based exercise, and conscientiousness and non-competitive/solitary exercise. Furthermore, significant environmental correlations were observed between neuroticism and total exercise, extraversion and total exercise, extraversion and non-competitive/solitary exercise, conscientiousness and total exercise, and conscientiousness and non-competitive/solitary exercise, although the effect size was only about 25% of that of the genetic correlation.

The same pattern can be observed in the results from the cross-sectional bivariate model where data at timepoint 2 was used, though fewer parameters were significant, particularly for the environmental correlation due to the lower effect size in combination with the reduced sample size. Separately looking at males and females also yielded the same pattern of results, again more often significant in the females where sample size was larger than in males.

The results of the longitudinal bivariate twin-sibling models (Supplementary table 3) were highly consistent in direction and magnitude with those from the cross-sectional model, but fewer statistically significant estimates were obtained, likely due to the drop in power due to the decreased sample sizes as a result of using

longitudinal data. Genetic correlations between extraversion and any exercise, as well as the genetic correlation between conscientiousness and non-competitive/solitary exercise were still statistically significant. However, none of the longitudinal environmental correlations for these pairs reached significance.

Genetic correlation based on GWAS summary statistics

All genetic correlations and h^2_{snp} estimates can be found in Supplementary Table 7.4. For neuroticism, standard errors were smaller for the meta-analysis than for the stand-alone 23andme sample, suggesting these meta-analysis results are more accurate. Strongest within-domain genetic correlations were found between MetaNeuroticism and agreeableness (r_g = -0.30, SE = 0.05), MetaNeuroticism and extraversion (r_g = -0.20, SE = 0.04), extraversion and openness (r_g = 0.34, SE = 0.05). Significant cross-domain genetic correlations were found between SSOE and extraversion (r_g = 0.14, SE = 0.04), openness (r_g = 0.15, SE = 0.04), and neuroticism (r_g = -0.22, SE = 0.03).

Polygenic risk scores (PRS)

Results (R^2 and p-values) for all PRS analyses are provided in Supplementary Table 7.5. Priors with the highest $R^2$for each respective phenotypes were included in the MR-DoC analysis. Prediction of personality traits using PRS are in line with expectations as the PRS constructed from the meta-analysis of neuroticism summary statistics provided the most predictive power for neuroticism in NTR (max(R^2) = 0.020, p = 2.9×10^{-40}). For openness (R^2 = 0.008, p = 1.1×10^{-16}), agreeableness (R^2 = 0.004, p = 2.3×10^{-9}), conscientiousness (R^2 = 0.005, p = 6.1×10^{-11}) and extraversion (R^2 = 0.010, p = 1.5×10^{-20}) the PRS provided most predictive power for their respective NTR phenotypes. The PRS for SSOE explained most variance in non-competitive/solitary exercise

in NTR (R^2 = 0.004, p = 8.9×10^{-10}), closely followed by the total volume of exercise (R^2 = 0.003, p = 1.3×10^{-7}). The PRS for SSOE was also a significant predictor of competitive/team-based exercise but the variance explained in this domain is notably lower (R^2 = 0.001, p = 0.002) and we consider this PRS not suitable as strong genetic instrument for Mendelian Randomization.

Mendelian Randomization – Direction of Causation model

A summary of the parameters of each best-fitting MR-DoC model (AE+PRS) is included in Supplementary Table 7.6. Path estimates of the model including extraversion as exposure and total exercise as outcome are depicted in Figure 7.2b.

Significant bidirectional effects were found between extraversion and total volume of exercise, with the path estimate for an effect of extraversion on exercise being slightly higher (g_1 = 0.12, *SE* = 0.03) compared to the estimate for the reverse direction (g_1 = 0.10, SE = 0.03). A similar pattern of significant bidirectional effects was observed between extraversion and non-competitive/solitary exercise (g_1 = 0.08, SE = 0.03 and g_1 = 0.07, SE = 0.02). Likewise, bidirectional effects were found between conscientiousness and total volume of exercise (g_1 = 0.11, *SE* = 0.03, and g_1 = 0.10, SE = 0.02), as well as between conscientiousness and non-competitive/solitary exercise (g_1 = 0.11, *SE* = 0.03, and g_1 = 0.12, *SE* = 0.02).

Testing the causal hypothesis

An overview of the results on tests for a causal effect of personality on total and non-competitive exercise is included in Table 7.4. The complete results, including competitive/team-based exercise is included in Supplementary Table 7.7.

			GEE Association		MZ Intrapair	Cross-sectional Bivariate Twin Model		Longitudinal Bivariate Twin Model		LDreg	MR-DoC	
	Exercise		beta (T1)	beta (T1->T2)	beta(MZ)	rA	rE	rA	rE	Rg	g1	b2
Neuro-ticism	Total	Results	-0.10*	-0.08*	-0.09*	-0.17*	-0.07*	-0.16*	-0.04	n.a.	0.00	0.02
		Causality	✓	✓	✓	✓		X			X	
	Ncomp/ Sol	Results	-0.07*	-0.06*	-0.03	-0.13*	-0.03	-0.14*	-0.03	-0.22*	0.00	0.00
		Causality	✓	✓	X	X		X		✓	X	
Extra-version	Total	Results	0.14*	0.13*	0.09*	0.23*	0.08*	0.27*	0.02	n.a.	0.12*	0.01
		Causality	✓	✓	✓	✓		X			✓	
	Ncomp/ Sol	Results	0.12*	0.11*	0.06	0.21*	0.06*	0.27*	0.01	0.14*	0.08*	0.01
		Causality	✓	✓	X	✓		X		✓	✓	
Openness	Total	Results	-0.00	0.01	0.01	-0.10*	0.01	-0.02	-0.01	n.a.	0.00	0.00
		Causality	X	X	X	X		X			X	
	Ncomp/ Sol	Results	0.06*	0.05*	0.05	0.05	0.05	0.05	0.04	0.15*	0.00	0.00
		Causality	✓	✓	X	X		X		✓	X	
Agree-ableness	Total	Results	-0.00	-0.01	0.01	-0.05	0.02	0.03	-0.02	n.a.	0.00	0.00
		Causality	X	X	X	X		X			X	
	Ncomp/ Sol	Results	0.00	-0.03	0.00	0.01	0.04	0.17	-0.07	0.10	0.00	0.00
		Causality	X	X	X	X		X		X	X	
Conscien-tiousness	Total	Results	0.08*	0.07*	0.10*	0.07	0.08*	0.11	0.05	n.a.	0.11*	0.01
		Causality	✓	✓	✓	X		X			✓	
	Ncomp/ Sol	Results	0.09*	0.09*	0.08*	0.13*	0.08*	0.21*	0.04	0.03	0.12*	0.00
		Causality	✓	✓	✓	✓		X		X	✓	

*Table 7.4. Result summary of personality traits and **Total** and non-competitive/solitary (**Ncomp/Sol**) exercise. *: p < 0.0033, **Causality**: A checkmark (✓) in this row indicates the above result supports the causal hypothesis; an X indicates that the above results falsify the causal hypothesis, **beta(T1)**: Cross-sectional phenotypic association, **beta(T1->T2)**: Longitudinal phenotypic association, **beta(MZ)**: Beta estimate from the within-MZ difference model, **rA**: Genetic correlation from the respective biometric model, **rE**: Environmental correlation from the respective biometric model, **Rg**: Genetic correlation from LDscore regression, **g1**: Causal-path estimate from the MR-DoC model, **b2**: Genetic pleiotropy estimate from the MR-DoC model.*

For neuroticism, negative associations and some support was found for the causality-based hypothesis, regardless of which exercise domain was assessed. Since the MR-DoC model, being our strongest test for causal effects, did not find significant estimates for neuroticism and any exercise, and given the consistent negative genetic correlations, genetic pleiotropy seems the most likely cause for the association between neuroticism and exercise behaviors.

For extraversion, positive association and support for both genetic pleiotropy and causality-based hypotheses were found in different exercise domains. Where results of extraversion and competitive/team exercise only support genetic pleiotropy, results of extraversion and total or non-competitive/solitary exercise support causality (and genetic pleiotropy to a lesser extent).

For openness, associations are negative for competitive/team exercise, and positive for non-competitive/solitary exercise. Furthermore, only results of openness and non-competitive/team exercise seem to suggest an association. The near-zero associations between openness and total exercise are likely the result of the opposing estimates for competitive/team exercise on one hand, and non-competitive/solitary exercise on the other.

For agreeableness, though some slightly negative associations were found, no support for either pleiotropy or causality was found.

For conscientiousness associations with total and non-competitive/solitary exercise were significantly negative, whilst associations with competitive/team exercise were slightly negative or not significant. Furthermore, most support was found for a causal relationship between conscientiousness and total or non-competitive/solitary exercise.

Testing reverse causality

An overview of the results regarding the causal effect of total and non-competitive exercise on personality is included in Table 7.5. The complete results, including competitive/team-based exercise is included in Supplementary Table 7.8.

			GEE Association		MZ Intrapair	Cross-sectional Bivariate Twin Model		Longitudinal Bivariate Twin Model		LDreg	MR-DoC	
x	Personality		beta (T1)	beta (T1->T2)	beta(MZ)	rA	rE	rA	rE	Rg	g1	b2
Total	Neuroticism	Results	-0.10*	-0.07*	-0.09*	-0.17*	-0.07*	-0.16*	-0.01	-0.22*	0.00	0.00
		Causality	✓	✓	✓	✓		X		✓	X	
	Extraversion	Results	0.14*	0.13*	0.09*	0.23*	0.08*	0.31*	0.01	0.14*	0.10*	0.02
		Causality	✓	✓	✓	✓		X		✓	✓	
	Openness	Results	-0.00	0.00	0.01	-0.10*	0.01	-0.01	-0.02	0.15*	0.00	0.03
		Causality	X	X	X	X		X		✓	X	
	Conscientiousness	Results	0.08*	0.07*	0.10*	0.07	0.08*	0.11*	0.04	0.03	0.10*	0.00
		Causality	✓	✓	✓	X		X		X	✓	
	Agreeableness	Results	-0.00	-0.02	0.01	-0.05	0.02	-0.08	0.02	0.10	0.00	0.04
		Causality	X	X	X	X		X		X	X	
Non-competitive/Solitary	Neuroticism	Results	-0.07*	-0.05*	-0.03	-0.13*	-0.03	-0.13*	-0.01	-0.22*	0.00	0.00
		Causality	✓	✓	X	X		X		✓	X	
	Extraversion	Results	0.12*	0.11*	0.06	0.21*	0.06*	0.24*	0.04	0.14*	0.07*	0.02
		Causality	✓	✓	X	✓		X		✓	✓	
	Openness	Results	0.06*	0.04*	0.05	0.05	0.05	0.12*	-0.03	0.15*	0.00	0.04
		Causality	✓	✓	X	X		X		✓	X	
	Conscientiousness	Results	0.09*	0.08*	0.08*	0.13*	0.08*	0.14*	0.06	0.03	0.10*	0.00
		Causality	✓	✓	✓	✓		X		X	✓	
	Agreeableness	Results	0.01	-0.01	0.01	0.02	0.01	-0.02	0.02	0.10	0.00	0.04
		Causality	X	X	X	X		X		X	X	

*Table 7.5. Result summary of Total and non-competitive/solitary exercise and Neuroticism, Extraversion, Openness, Agreeableness, Conscientiousness. *: p < 0.0033,* ***Causality****: A checkmark (✓) in this row indicates the above result supports the causal hypothesis; an X indicates that the above results falsify the causal hypothesis,* ***beta(T1)****: Cross-sectional phenotypic association,* ***beta(T1->T2)****: Longitudinal phenotypic association,* ***beta(MZ)****: Beta estimate from the within-MZ difference model,* ***rA****: Genetic correlation from the respective biometric model,* ***rE****: Environmental correlation from the respective biometric model,* ***Rg****: Genetic correlation from LDscore regression,* ***g1****: Causal-path estimate from the MR-DoC model,* ***b2****: Genetic pleiotropy estimate from the MR-DoC model.*

For total volume of exercise behavior a negative association was observed with neuroticism, with most results in favoring genetic pleiotropy over causality. Positive associations and support for the causality-based hypothesis was found with extraversion, and to a lesser extent conscientiousness. Regular exercise appears to increase the scores on these personality traits. No associations were found with openness and agreeableness.

Since the PRS for SSOE was not found to be a strong genetic instrument for competitive/team-based exercise, we did not explicitly test for reverse causality for this type of exercise.

For non-competitive/solitary exercise, similar to total exercise, negative associations and results in line with genetic pleiotropy were found with neuroticism. Support for the causality-based hypothesis was found with conscientiousness and extraversion. Regular non-competitive/solitary exercise appears to increase the scores on these personality traits.

Discussion

We employed triangulation across a number of different methods, including 'classic' prospective epidemiology, intrapair regression in identical twins, multivariate twin-sibling estimation of genetic and environmental correlation, and 2-Sample Mendelian Randomization to address causality in the personality – exercise behavior relationship while taking into account genetic confounding. Results varied across personality traits and types of exercise behavior. Our results are congruent with a bidirectional causal effect of extraversion and conscientiousness on the total volume of weekly exercise behavior, and specifically on activities that are solitary and non-competitive. Neuroticism showed the well-known negative association with exercise behaviors, but this association seems to derive from genetic pleiotropy rather than causal effects. We find no

role for agreeableness and openness to experience as determinants or even correlates of exercise behavior.

The consistently significant negative association between neuroticism and total weekly exercise and the positive associations between extraversion/conscientiousness and total weekly exercise are all in line with the three previous meta-analyses (Rhodes & Smith, 2006; Sutin et al., 2016; Wilson & Dishman, 2015), which found significant effects of these personality traits in the same range as our cross-sectional association estimates ($r \approx 0.10$). Furthermore, the support for extraversion and conscientiousness being factors that increase voluntary exercise behavior was in line with our hypotheses. These hypotheses were based on the extant literature, that had however, not adequately accounted for (genetic) confounding. Whereas genetic confounding is likely to take place, we now find support for an additional causal effect of extraversion and conscientiousness on the total volume of exercise. Contrary to our hypothesis, the effects for extraversion were driven by solitary/non-competitive activities and not by team-based/competitive activities. Our expectation was that the latter type of activities would be more appealing to persons high on facets of extraversion like being sociable, assertive, and outgoing. This expectation was not borne out by the data. However, we note that our test may not have been optimal as a much lower number of participants engaged in team-based activities than solitary activities. We have previously shown that team-based competitive exercise peaks in late adolescence and is not a big component of total exercise behavior in the fourth decade of life where most of our participants resided (van der Zee, van der Mee, et al., 2019).

As observed previously, individuals high in conscientiousness tend to be organized and disciplined, have higher self-efficacy (Rhodes et al., 2003), and their motivation to exercise tends to come from internal, rather than external, sources (Ingledew & Markland, 2008).

Such factors are known to be conductive to increase adoption of a health behavior and reduce drop-out (Ryan & Deci, 2007). A causal effect of conscientiousness on total weekly exercise is therefore not surprising. Furthermore, that these effects were mostly driven by solitary/non-competitive activities and not team-based/competitive activities was also in line with our hypothesis. We expected that self-discipline would be particularly relevant when external cues and the social pressures of team-mates or a waiting opponent are absent. Moreover, the health aspect of exercise may be what draws the high conscientious to exercise, rather than its social esthetic, or enjoyment aspects (Courneya & Hellsten, 1998).

In the extant literature, the association between high neuroticism and low exercise behavior has received most of the attention because it is congruent with the association between regular exercise and anxious-depressive symptoms, for which neuroticism is a risk factor. Here, our primary hypothesis was that high neuroticism is a cause of lower exercise behavior, but we also tested the reverse idea that neuroticism could be reduced by regular exercise, which could explain part of the antidepressant effect of exercise (Stubbs et al., 2018). We find no support for either of these hypotheses. In keeping with a previous study on a subset of these data (de Moor & de Geus, 2018) we find genetic pleiotropy between neuroticism and total exercise to be the most likely explanation for this association. In particular, the non-significant MR-DoC estimate for the causal path argues against a causal impact of exercise behavior on neuroticism. We note that the absence of a causal effect of exercise on neuroticism clearly does not preclude a causal effect on depression, for which substantial evidence has accrued (de Geus, 2020). Instead, the effects on extraversion and conscientiousness may play a larger role in the beneficial psychological effects of exercise than they are currently given credit for.

Conflicting results have been reported for the link between openness to experience and exercise behaviors. Openness was a significant correlate of physical activity levels in the meta-analyses by Wilson and Dishman (2015) and Sutin et al. (2016), but not in the meta-analysis by Rhodes and Smith (2006). Our results provide a tentative suggestion for these discrepancies, because we see substantial differences in the associations between openness and exercise depending on which exercise domain is considered. Consistently no association has been found between agreeableness and exercise (Rhodes & Smith, 2006; Sutin et al., 2016; Wilson & Dishman, 2015) and we fully confirm this here. Regular exercisers are not more (or less) agreeable than non-exercisers.

Completely contrary to our hypotheses was the surprisingly consistent support for the existence of reverse causality, where exercise itself was a cause of changes in extraversion and conscientiousness. We did not strongly anticipate the possibility of an effect of regular exercise behavior on personality based on the idea that personality traits are relatively immutable factors, with tracking over time largely determined by genetic factors (Costa Jr et al., 2019). Absolute levels of personality scores are known to change over time, even in adulthood (Terracciano et al., 2010), but we expected this to be mainly an effect of maturation (Costa Jr et al., 2019) and life events (Specht et al., 2011) with little if any effect of changes in (health) behaviors. Our results suggest otherwise, and we conclude that the effects of extraversion and conscientiousness on exercise behavior are reciprocated by similarly sized effects of exercise behavior on extraversion and conscientiousness. Changes in personality and exercise behavior appear to reinforce one another and such bidirectional causality can be expected to increase the temporal stability of both traits across the life span. Notably, age-dependent increase in tracking coefficients for personality (Anusic & Schimmack,

2016; Hampson & Goldberg, 2006; Milojev & Sibley, 2014; Roberts & DelVecchio, 2000; Terracciano et al., 2010) and regular exercise (van der Zee, van der Mee, et al., 2019) have indeed been found.

This pattern further suggest that personality traits often associated with athletic success, high conscientiousness in particular (Steca et al., 2018), may not only provide a predispositional advantage to achieve this success, but that these personality traits may be further amplified by the prolonged exercise activity required to achieve this athletic success. In line with this hypothesis Lochbaum et al. (2010) find strong almost dose-response-like associations between how long participants have engaged in regular exercise (non-exercisers, less than six concurrent months, and more than six concurrent months) and extraversion and conscientiousness, in the expected direction based on our bidirectional findings.

We did not find support for any reverse causal effects of competitive/team-based exercise but attribute this to a limitation in the MR-DoC model for this test. The PRS for exercise behavior was based on the UKB summary statistics for SSOE, as this is the only phenotype that is somewhat comparable to the weekly METhours of voluntary exercise assessed in the NTR (Klimentidis et al., 2018). However, in the age range of the UKB population, the SSOE report will predominantly reflect solitary/non-competitive activities (van der Zee, van der Mee, et al., 2019). Indeed the PRS based on SSOE did not significantly predict this type of activity in the NTR, disqualifying it for the MR part of the MR-DoC test. This limitation extends to the PRS for total exercise and non-competitive/solitary exercise, although to a milder extent. The variance explained in non-competitive/solitary exercise by this PRS was almost an order of magnitude lower than the variance explained in some personality traits by their respective PRS (R^2 = 0.004 compared to R^2 between 0.004 and 0.02), which may have had an effect on the MR-DoC estimates where this PRS was used.

Unfortunately, at the time of writing, the only available GWAS summary statistics for the exact exercise behaviors used in NTR has been done in this same population (van der Zee et al., 2021), the use of which would have induced bias in the MR-DoC estimates in the current study.

Another limitation of the current study is a potential lack of power to detect a significant environmental correlation in the longitudinal twin-model, due to a much smaller sample size in the longitudinal dataset (N = 18,554) compared to the cross-sectional dataset (N = 36,636). As shown by de Moor et al. (2008), a bivariate genetic model on traits with cross-sectional correlation of 0.10, needs 3864 twin pairs to have a power of 80% to detect a significant environmental correlation > 0.10. This drops to 1704 when the cross-sectional correlation is 0.15. For personality, we had at most 2353 twin pairs contributing to the longitudinal analyses, which would not have been sufficient to declare the small environmental correlations (0.05 rather than 0.10) significant. This is likely the cause of the more robust findings in the cross-sectional bivariate model and the MR-DoC model when compared to the longitudinal bivariate models.

Despite these limitations, our combined use of multiple different tests in an attempt to falsify causal associations between personality and exercise provides a uniquely powerful assessment in phenotypes like personality where traditional case-control designs can't be used. The highly consistent findings in line with causal effects of extraversion on total and non-competitive/solitary exercise, as well as of conscientiousness on total and non-competitive exercise, strongly suggest there is a bidirectional causal association between these specific personality traits and exercise behavior.

The implications of the bidirectional causal associations between personality and exercise we found in this paper are broad. First, they may provide insight into the mechanisms behind the association between exercise and mental health (de Geus, 2020) and

the effectiveness of exercise therapy for various mental disorders such as depression (Haller et al., 2018; Knapen et al., 2015; Stanton & Reaburn, 2014), anxiety (Jayakody et al., 2014) and substance use disorder (Hallgren et al., 2017), and others (see Ashdown-Franks et al. (2020) for review). Whereas the focus has been on neuroticism as a mediator between exercise and these disorders, they are also strongly predicted by extraversion and conscientiousness (Klein et al., 2011; Kotov et al., 2010), the traits where we find most support for bidirectional causal relationships. This leads us to hypothesize that exercise may affect mental health outcomes through these personality traits. Second, our results support the idea that the effectiveness of exercise intervention can be increased by the use of personality-informed interventions (Chapman et al., 2014), as personalized interventions are likely more effective than broad interventions (Buford et al., 2013). The demonstrated benefits of increasing the adoption and maintenance of regular exercise behavior globally, which may include desirable personality changes, further increases the urgency of the development and implementation of better interventions.

Summary and Discussion

Summary

The past years have seen a remarkable surge of scientific discoveries in human genetics, and complex-trait genetics specifically. This has been largely facilitated by the development of affordable genome-wide arrays and aided by the development of new statistical methods to use the data from these arrays for different purposes. The genetic epidemiology of regular voluntary exercise behavior is no exception to this trend. As clear indicators of this trend, the largest reviews to date of twin-, family and gene-finding studies on exercise behavior (Aasdahl et al., 2021; van der Zee & de Geus, 2019; van der Zee, Schutte, et al., 2019) have been published during my PhD project, some of which are included in this thesis. In the same time frame, four of the six genome-wide association studies on physical activity traits (Doherty et al., 2018; Klimentidis et al., 2018; Lin et al., 2018) including sports and exercise behaviors (Doherty et al., 2018; Klimentidis et al., 2018; Lin et al., 2018; van der Zee et al., 2021) saw the light and these have produced the first genome-wide significant genetic variants.

Throughout this dissertation I have employed several methods to further our understanding of determinants of regular voluntary exercise behavior, with an emphasis on genetic factors. In the first three chapters of this thesis, I explore the prevalence and stability of exercise, and review currently available literature concerning genetic epidemiology of exercise. These chapters are followed by an extended-family study, a genome-wide association study (GWAS), and finally a study on the causal effects of personality on exercise, using genetically informed designs. A recurring theme throughout this dissertation was the separation of the total volume of exercise into activities in different domains. The main findings in this thesis are summarized below, followed by a general discussion on the genetic epidemiology of exercise, and my thoughts on future developments in this field.

Prevalence and stability of voluntary exercise behavior

As mentioned in the introduction of this thesis, voluntary exercise behavior is a prime target for behavioral intervention to increase physical activity. Since the health benefits only persist as long as the exercise activities persists, the ultimate goal of such interventions is to develop life-long habits of regular exercise. In other words, the ultimate goal is to not only increase the prevalence, but also increase the *tracking* (i.e. the stability over time) of exercise behavior. The age- and sex-specific prevalence are generally reported in all studies on exercise behavior, however an assessment of tracking is more rare as it requires a large longitudinal dataset. Creating such a dataset is a costly endeavor, especially when the aim is to assess tracking across the lifespan. It is, therefore, not surprising that many previous studies on tracking of exercise focused only on a specific range, such as from childhood to adolescence (Aarnio et al., 2002; Francis et al., 2013; Pahkala et al., 2013; R. Tammelin et al., 2014).

In **Chapter 2** of this thesis, we used a large dataset (N = 43,889) from the Netherlands Twin Register (NTR) to assess prevalence and tracking coefficients across the largest part of the lifespan (ages 8 to 80), with varying distances between the baseline exercise measurement and follow-up (2 to 22 years). We assessed the prevalence and tracking of total weekly volume of exercise as well as of the volume of activities in six specific exercise domains: team-based, solitary, competitive, non-competitive, externally paced, and internally paced. Generalized estimating equations (GEE) were used throughout to correct for family-relatedness in the NTR sample.

The prevalence of exercise increases during childhood and early adolescence, with the highest prevalence found in mid-adolescence, around age 16 for both men and women. The largest contributor to this peak is the prevalence of team-based and competitive exercise activities, which are the dominant exercise

activities in the Netherlands till around age 25. These activities show a strong linear decline after age 16 the slope of which is attenuated around age 40. After age 25, solitary activities, often non-competitive in nature, become the mainstay of the total weekly volume of exercise behavior. The prevalence of these solitary non-competitive exercise activities, remarkably, remains relatively stable from ages 18 to 64. After age 64 the prevalence of non-competitive solitary exercise also starts to decline, though at a slower rate than team-based competitive exercise did during adolescence. Throughout the life span, men have on average 25% higher levels of total exercise than women, but the sex difference is most pronounced in adolescence and early adulthood. This is because the discrepancy in total exercise between men and women during adolescence could be largely attributed to the greater prevalence of team-based and competitive exercise in men.

The tracking of exercise, regardless of type of exercise was moderate to high (2-year tracking coefficients between .38 and .77) with slightly higher tracking coefficients in men. In accordance with previously published tracking studies (Sallis et al., 1996; Telama et al., 2006) we found tracking strength to decrease as the distance between starting age and follow-up age increased. For example, tracking coefficients for total volume of exercise had a median of .57 across all 2-year intervals down to a median of .38 for 22-year intervals. Also in accordance with previous literature we found tracking coefficients generally increased with starting age resulting in lower tracking from childhood into adulthood compared to the higher coefficients within adulthood. For example, in females the 4-year tracking coefficient of the total volume of exercise from age 10 at baseline to age 14 at follow-up was 0.43, whereas the tracking coefficient from age 48 at baseline to age 62 at follow-up was 0.81. However, the increase in tracking strength with increasing age was entirely driven by a sharp increase in tracking for solitary and non-competitive activities. For team-based

and competitive activities, tracking was fairly stable or even gradually decreased across the life span, possibly because the prevalence of such activities decreases. These results suggest that adult exercisers engaged in solitary exercise become increasingly less likely to drop out, and vice versa, that non-exercising adults become increasingly less likely to ever take up exercise, even if it is solitary self-paced and non-competitive.

To conclude, regular voluntary exercise behavior declines with age, but this is largely driven by a decrease in team-based activities. Solitary exercise activities show a much more modest decline in prevalence. The significant female disadvantage in exercise prevalence is moderated by type of exercise and age. Throughout, the moderate to high tracking coefficients confirm that exercise behavior is a rather stable individual characteristic.

Heritability of regular exercise behavior

A large volume of genetic studies has accrued in adult populations that looked at regular voluntary exercise behavior, as well as three other physical activity phenotypes, namely total physical activity (TPA), moderate-to-vigorous physical activity (MVPA), and leisure-time physical activity (LTPA). While TPA, MVPA and LTPA all encompass voluntary exercise behavior, they are much broader in also including non-exercise activities like walking, bicycling or gardening, and activities that are not fully voluntarily controlled by an individual, such as physical activity for transport or manual labor. Total physical activity places no restriction on intensity of the activity, whereas MVPA uses only activities with a minimum intensity (generally activities with an intensity value of more than 3 or 4 times the resting metabolic rate). LTPA again does not select activities based on intensity, but on whether they are performed voluntarily in leisure time and most closely resembles voluntary exercise behavior. The key distinction

being that LTPA also includes all non-exercise or sports performed in leisure time, such as walking or dancing, on top of exercise and sports activities.

Whereas the past studies on exercise genetics have unanimously shown an important genetic contribution to physical activity behaviors, estimates of heritability have varied considerably, even for a similarly defined exercise phenotype (de Geus et al., 2014; Lightfoot et al., 2018). In **Chapter 3** of this thesis, we therefore performed a systematic search and a meta-analysis of adult twin studies on regular voluntary exercise behavior, as well as three other physical activity phenotypes. Twin-studies are particularly well-suited to estimate the heritability of a trait by comparing the difference in within-pair resemblance of monozygotic (MZ) or dizygotic (DZ) twins. A greater resemblance in MZ twins compared to DZ twins suggests additive genetic factors (A) affect the variance in a trait and additional non-additive genetic factors (D), if the difference is more than a factor two. This is because MZ twins share (nearly) 100% of their genetic variants, whilst DZ twins share, on average, about 50% of additive genetic and 25% of non-additive genetic variance. Conversely, if the resemblance in MZ and DZ twins does not differ significantly this suggests that common environmental factors (C) influence the variance of a trait. This is because both MZ twins and DZ twins share all of their common environmental factors, such as parenting styles or prenatal environment, by definition. Finally, the differences between MZ twins can only be ascribed to unique environmental factors (E). Of note, E also includes any phenotypic measurement error.

A set of 27 independent adult twin studies were included, some of which used objective measures (such as accelerometers), others used self-reported surveys. We meta-analyzed the A and C estimates of each study (almost no D was reported) for males and females separately and corrected for the original study sample size.

Contribution of A was significant and substantial in all traits, with the lowest meta-analytic estimates found for MVPA (A_{males} = 0.41, $A_{females}$ = 0.44), and the highest estimates found for TPA (A_{males} = 0.51, $A_{females}$ = 0.48) and exercise behavior (A_{males} = 0.48, $A_{females}$ = 0.51). The contribution of C was only found to be consistently significant in voluntary exercise behavior (meta-analytic estimate 0.10 for both sexes), but not for the broader physical activity traits. No significant sex-differences were found for any of the A and C estimates resulting from the meta-analysis.

In **Chapter 4** of this thesis, we additionally included parent-offspring and sibling correlations from family studies in our review of studies assessing the genetic and environmental contributions to regular voluntary exercise behavior. Twin studies rely on a number of assumptions, one of which is the equal-environment assumption, which states the environmental effects are equal for MZ and DZ twins. Sibling correlations and parent offspring correlations are not relying on this assumption. However, an assumption in estimating A from parent-offspring in the standard family-based design correlations is that the transmission of trait resemblance is entirely due to genetic and not to cultural transmission. In the case of cultural transmission, the behavior of parents directly influences the behavior of the offspring (i.e. parents act as a role-model), thus increasing the influence of common environmental factors. In previous work in the NTR, de Moor et al. (2011) found a minimal (3%) effect of cultural transmission in boys only, despite their participants being adolescents, for whom shared environmental factors are still significant. In keeping, the heritability estimates resulting from the family-studies on exercise that we reviewed in Chapter 4 suggest that those found in twin-only studies are indeed not overly biased.

More alarmingly, however, **Chapter 4** did find spouse correlations to be consistently significant and substantial, ranging

from 0.16 to 0.48. This indicates noticeable assortment for exercise behavior, which violates another assumption of the classical twin design, namely that the trait should not influence an individual's choice of a partner. If such assortative mating is present, and not corrected for, DZ twins in a family may share more than 50% of genetic factors influencing a trait, which in turn will cause an overestimation of the influence of common environmental factors, and an underestimation of additive genetic factors.

In **Chapter 5** we addressed assortment in combination with a final major shortcoming of the twin- and family-studies included in chapters 3 and 4. These previous studies only estimated the effects of A, C and E, but could not simultaneously assess the effects of non-additive genetic factors (D). To achieve this, we used data from 50,690 adolescent and adult participants from 19,543 nuclear pedigrees in the NTR. All family-relations amongst these individuals and within these nuclear pedigrees were known, and this knowledge was leveraged by using the Mendel software package (Lange et al., 2013). Different definitions of 'shared environment' or household effects were tested, namely full household (where C contributes to spousal, parent-offspring, twin- and sibling resemblance), spouse household (where C only contributes to spousal resemblance), sibling household (where C only contributes to resemblance of twins and siblings), and twin household (where C only contributes to resemblance of twins). While still focused on voluntary exercise behavior, we expanded on previous twin-family studies by not only estimating the contribution of the variance components to the total volume of exercise but also to subsets of activities in different domains like team-based or solitary or competitive and non-competitive exercise.

Depending on which definition of household effects was used the broad-sense heritability (h^2, the percentage of variance explained by A and D) ranged from 34% to 41%. Higher heritability estimates were

found for team-based, competitive and externally paced exercise behavior (39% < h^2 < 47%), compared to solitary, non-competitive and internally paced exercise (26% < h^2 < 34%), and were largely attributable to an increased contribution of non-additive genetic factors. Across all exercise domains the contribution of C strongly depended on which definition was tested, with low estimates found under the twin- or sibling household definitions, and higher estimates for the spouse- and full household definitions. For total volume of exercise, for example, the percentage of variance explained by sibling household effects was 4%, by twin household effects 8%, by full household effects 20%, and by spouse household 24%. These large differences mainly reflected the high spousal correlations found across all exercise domains (0.30 < r < 0.50).

Chapter 5 also examined the causes of spousal resemblance for regular exercise behavior, with a particular interest on how they could bias twin models. The only study that had examined spousal resemblance of exercise behavior was published by de Moor et al. (2011), who found that this resemblance was best explained by phenotypic assortment. To re-assess the sources of this spousal resemblance we adopted the approach outlined by van Grootheest et al. (2008) in Chapter 5. Using this model allowed us to test three distinct sources of spousal resemblance in exercise behavior. First phenotypic assortment, which describes preferential mating between partners that more resemble each other's exercise behavior. Second social homogamy, which refers to similar partner selection based on sharing a milieu in which exercise is common. Finally marital interaction, which describes partners growing more similar over time. This could mean that the non-exerciser starts exercising, influenced by the other, or vice versa. Our results suggested marital interaction was a significant contributor in all exercise domain, and there was more evidence in favor of social homogamy over phenotypic assortment. This increases

confidence in twin studies of voluntary exercise behavior that will be most strongly biased by phenotypic assortment, to a lesser extent by social homogamy, and even less by marital interaction.

To conclude, results of chapters 3-5 are in line with those of previous twin and family studies regarding the heritable nature of regular voluntary exercise behavior as well as broader physical activity traits in adulthood, and the near absence of lingering effects of being raised in a shared family environment. Like so many other behavioral traits, "about half of the variance" in adult exercise behavior can be explained by heredity. This heritability, however, is a result of both additive, and non-additive effects, particularly but not exclusively for team-based, competitive and externally paced exercise. Marital interaction seems to be consistent source of influence on spousal resemblance in regular voluntary exercise behavior and there is likely an additional role for social homogamy. These effects are of greater influence than phenotypic assortment.

Gene-finding in regular voluntary exercise behavior

As mentioned in the introduction of this thesis discovering the genetic variants underlying the heritability of regular voluntary exercise behavior has proven to be difficult, despite the well-established moderate-to-high heritability. A large number of candidate-gene studies have been performed in which a single or a few genetic variants were assessed for an association with regular exercise or broader physical activity traits (see review in Chapter 6 but also Aasdahl et al. (2021)). Only very few candidate genes, however, have been independently replicated and then not even always at the level of a specific variant but rather at the gene level. A genome-wide association study (GWAS), where millions of genetic variants are tested simultaneously, is a more robust and hypothesis-free method for identifying genetic variants. The first three GWASs on regular

voluntary exercise behavior, published by de Moor et al. (2009), Kim et al. (2014) and Lin et al. (2018), did not yield any significant results. The two GWASs that followed, published by Klimentidis et al. (2018) and Hara et al. (2018), did yield the first genome-wide significant hits for sports and exercise behaviors. Genetic variants for specific types of exercise activities, however, were not tested.

In **Chapter 6** of this thesis, we performed a GWAS of the total weekly volume of exercise, as well as a GWAS for activities in the team-based, solitary, competitive, non-competitive, and externally/internally paced domains as defined previously. We used a sample of the NTR (N = 14,626) for which genome-wide data in the form of millions of single nucleotide polymorphisms (SNPs) was available as well as survey data on exercise activities.

No genome-wide significant SNPs were found for total, team-based, solitary, competitive, externally paced and internally paced exercise. For non-competitive exercise five novel genome-wide significant SNPs were found. Four of these variants (*rs11817437*, *rs1276262*, *rs11812727*, *rs34897771*) were located in an intergenic region on chromosome 10 and have not been found in any previous genetic study of exercise behavior. One of these significant variants (*rs147244851*) was located in the *Glypican-5* (*GPC5*) gene. Though this gene has previously been found to be associated with sedentary behavior (Comuzzie et al., 2012), the variant found in our GWAS on exercise behavior was different from the variant associated with sedentary behavior. Since there is no evidence for either our exercise variant, or the previous sedentary behavior variant affecting gene expression, we cannot say with certainty that the direction of the effect found in these two GWASs is congruent.

We used our suggestive ($p < 5\times10^{-6}$) GWAS results to perform several functional annotation analyses. Through these analyses one of the variants (*rs62620995*) found for total volume of exercise was

identified as a missense (gene-product altering) variant in the *DCSTAMP* gene. This gene plays a crucial role in bone-development (Kukita et al., 2004), furthermore the same variant found in our GWAS has been shown to affect *DCSTAMP* gene expression and osteoclast (bone-resorbing cells) morphology (Laurier et al., 2017). Across the different types of exercise activities, the suggestive GWAS results were consistently enriched for genes that were related to a number of red blood cell phenotypes, such as glycated hemoglobin levels, mean corpuscular volume, and heme metabolism.

In chapter 6 we also aimed to replicate the 103 genetic variants identified in previous candidate-gene studies and test all these variants (or a genetic proxy of these variants) for association. This replication used a much more relaxed *p*-value than the GWAS, namely 0.05/103, i.e. the 'standard' α corrected for multiple testing. Strikingly, no significant evidence was found for replication of any of the previously reported genetic variants.

To conclude, in chapter 6 we identified novel genetic variants for regular voluntary non-competitive exercise activities specifically. Results of this chapter tentatively suggest bone development and red blood cell physiology to be biological determinants of voluntary exercise behaviors.

The pathways from genetic variation to regular voluntary exercise behavior

The recent discovery of significant genetic variants for physical activity traits, including regular voluntary exercise behavior, allows us to use new methods that leverage these variants to test causal hypotheses on the role of potential 'intermediate phenotypes' on the path from genes to exercise behavior. Personality, a trait known to be heritable, is one of these potential determinants of regular voluntary exercise. There is robust evidence for an association between these

two behavioral traits (Rhodes & Smith, 2006; Wilson & Dishman, 2015). The oldest, and still predominant perspective on this association is that it reflects a causal effect of personality traits on exercise behavior (Flemming, 1934), though an inverse causal effect has also been hypothesized (Stephan et al., 2014), and there is much room for genetic and environmental confounding. Testing the causal hypothesis, however, is not possible with traditional methods such as a controlled trial, as we cannot directly experimentally modify a subject's personality. This is where new genetically informed designs come in, including those based on twin-families and/or GWAS derived variants.

In **Chapter 7** we used a number of these designs to test the causal effects of the 'big five' personality traits (neuroticism, extraversion, agreeableness, openness and conscientiousness) on the total volume of exercise behavior, as well as on the volume of competitive/team and non-competitive/solitary exercise activities. First, we assessed the cross-sectional and longitudinal associations between the five personality traits and the various exercise behaviors. Second, as a direct test of causality we regressed the intrapair differences in MZ twins in exercise behaviors on the intrapair difference in the personality traits, thereby controlling for possible genetic and shared environmental confounding. Third, we tested a crucial requirement to uphold the causal hypothesis, namely that all genetic and environmental correlations between personality and exercise behavior would be significant. This was done in two different twin-sibling models, each with its own strengths. In the first model, to reduce impact of measurement error, we used the longitudinal data (two observations per participants) in a bivariate model as multiple indicators of the 'true phenotype'. In the second model, to reduce impact of reverse causality, we used longitudinal data in a prospective design, such that personality at baseline was a predictor for exercise behavior at follow-up. We separately re-assessed the genetic

correlation in independent data, using the available leave-NTR-out GWAS summary statistics for personality (de Moor et al., 2015; Lo et al., 2017) and for sports and exercise behavior (Klimentidis et al., 2018). Finally, the summary statistics from these GWASs were used to generate polygenetic scores (PRS) for personality traits and exercise behavior to be used as genetic instruments in a direct test for causality using the Mendelian randomization direction of causation (MR-DoC) model (Minică et al., 2018).

Results provided the strongest support for the causal hypothesis, i.e. by resisting falsification, for the association between extraversion and total volume of exercise, closely followed by the association between conscientiousness and non-competitive/solitary exercise and conscientiousness and total volume of exercise. Of note, we failed to show a causal effect of neuroticism on exercise behavior, which is in line with a previous cross-lagged panel analysis by de Moor and de Geus (de Moor & de Geus, 2018). Interestingly there was some evidence for a positive association between openness and non-competitive/solitary exercise, but no association between openness and total volume of exercise, likely due to the (non-significant) negative associations between openness and team-based exercise. Reverse causal effects of exercise behavior on extraversion, conscientiousness and openness were found consistently, although the effect sizes for such reverse causality were consistently smaller in our strongest estimator of the causal effect, the MR-DoC model.

From the results of chapter 7 we conclude that there is likely a bidirectional causal relationship between the personality traits extraversion, conscientiousness and openness to experience and regular voluntary exercise activities.

Discussion

The main conclusions and contributions of this thesis can be divided into three main topics, discussed separately below. First, the prevalence and stability of exercise, which mainly concerns chapter two, and the other chapters to a lesser extent. Second, the genetic epidemiology of voluntary exercise, where chapters 3, 4 and 5 mainly address the heritability, chapter 6 addresses gene-finding, and chapter 7 serves as an example of how this knowledge can be used to further study the pathways from genes to voluntary exercise behavior. Finally, the added value of a subdivision of total weekly exercise into activities in specific exercise domains, which has been a common theme throughout this thesis.

Exercise prevalence and stability

The results of chapter 2, as well as exercise descriptive statistics in other chapters provide yet more evidence that the weekly volume of regular exercise decreases with age and depends on sex. The peak of exercise behavior is in mid-adolescence, after which it decreases sharply until at about age 40 where it shows a more gradually slowing downward trend that re-accelerates around age 65. Males tend to exercise more than females at all ages although the rate of decrease after mid-adolescence is greater in males. These findings are by no means novel, and this pattern is a well-known characteristic of exercise behavior (Azevedo et al., 2007; Eime et al., 2016; Hirvensalo & Lintunen, 2011; R. Tammelin et al., 2014). The results of this thesis do, however, shed some new light on this pattern by splitting exercise activities into the various domains introduced previously. Our results indicate that the strong decrease in exercise after mid-adolescence is almost entirely due to a decrease in team-based competitive exercise in both males and females, whereas solitary non-competitive exercise remains relatively stable. This is in line with what was stated earlier by

Telama et al. (2005) based on work by Burton (1988): "there are large numbers of young athletes who would prefer to continue competitive sport but do not do so because they are screened out because of low achievement or because they perceive a sport as too demanding in the relation to their perceived ability". The strong emphasis on the *perceived* demand and *perceived* ability is further in line with a pathway model by de Geus (2020), which suggests a feedback loop between exercise behavior and perceived ability, where high perceived ability acts as positive reinforcement for exercise behavior, but low ability as punishment.

The above patterns in exercise prevalence suggest that the transition from competitive team-based sports to solitary exercise is currently a 'weak link' from a public health perspective. This is where a major drop-out occurs. Social roles may start to interfere with team-obligations and/or the level of organized team sports for adults is not as advanced as that for children and adolescents. It is not clear why many exercisers do successfully switch to other exercise activities, whereas others don't. Interestingly, the perceived value of competitive elements does not seem to change. We found a decreasing correlation between team-based and competitive exercise in parallel to an increasing correlation between solitary and competitive exercise throughout adulthood. Individuals favoring a competitive element tend to switch from a team-based (e.g. field hockey or soccer) to a competitive solitary exercise (e.g. tennis or squash) where they still obtain the perceived benefits of competition.

Exercise behavior that survives the transition into adulthood shows an increasingly higher temporal stability. Adult regular exercisers seem to have found a set of exercise activities that is sufficiently reinforcing to be maintained, possibly because it optimally matches their genotypes.

Genetic epidemiology of exercise

Previous findings suggest that the heritability of exercise behavior increases from childhood to adolescence, with a peak at around 18 years (Huppertz et al., 2016; Vink et al., 2011). As reviewed in Schutte et al. (2018), this occurs in parallel to a large change in the contribution of shared environmental influences. During childhood, sibling resemblance in exercise behaviors is strongly determined by the shared environment in which they grow up, but this shared environmental influence strongly wanes during adolescence. The strong decline of contribution of shared environmental factors may be due to the parental motivation increasing the volume of exercise behavior in their children (Beets et al., 2010). The effect of this parental motivation does decrease over time, giving way to increased influence from peers and coaches (Chan et al., 2012).

When the adolescents become young adults, heritability appears to decrease again to reach the levels reported in chapters 3 and 4, where we conclude that across the full adult life span about half of the variance in exercise behavior is contributable to genetic factors. Fine-grained analysis of the adult trajectory (as is available for children and adolescents) is currently lacking. However, various indicators suggest that this reflects a decrease in heritability from the peak in adolescence to lower levels in later stages of adulthood where the decrease becomes asymptotic. For example, the estimate of 48% from the meta-analyses in chapter 3 is higher than our finding in chapter 5, where we find a slightly lower heritability of roughly 40%, depending on the type of exercise. This is likely due to the large proportion of studies in the meta-analysis, with a relatively low mean age (~28) compared to the extended twin-family study in chapter 5 that had a mean age of 40 years.

This decreased influence of genetics is paired to a gradual increased influence of person-specific environmental influences.

There is a huge number of factors that can come into play in exercise behavior in adulthood such as entering a stable relationship, having children, job demands, caretaking obligations, or health problems (Dishman et al., 1985; Seefeldt et al., 2002; Simonen, Videman, et al., 2003). Furthermore, as identified in our extended twin-family study, we find a significant contribution of spouses on each other's exercise behaviors. Assuming individuals do not share a household with their spouse before adulthood and given the significance of marital interaction we find in chapter 5, the spousal influence on exercise behavior after partnering up may slowly increase up to 24% of explained variance.

Combining all results obtained in the NTR cohort I would like to propose an extension to the figure by Huppertz et al. (2012) showing the change in relative contributions of genetic and environmental factors to the total volume of regular voluntary exercise behavior as a function of age.

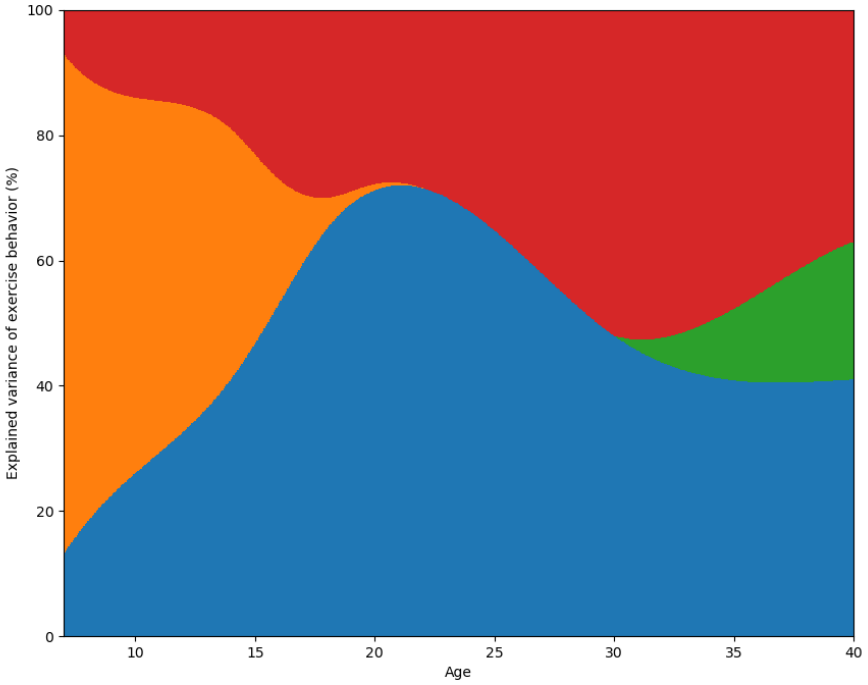


Figure 8.1. Percentage of variance in regular exercise behavior explained by unique environmental factors, environmental factors shared by siblings, genetic factors, and environmental factors shared by spouses over time.

It is important to note that the genetic factors displayed in my Figure 8.1 includes both additive, and non-additive genetic factors. Due to the limitations of the classical twin model, only one of common environmental factors, or non-additive genetic factors can be included in the model. If non-additive genetic factors are not specifically defined in the model, their contribution to the variance will end up in the additive genetic component. Given our finding that about half of the heritability (and even more so in team-based and competitive exercise) is attributable to non-additive genetic factors it seems likely that these factors also play a role earlier in life. More research is needed to accurately separate non-additive from additive genetic factors

throughout the lifespan, so for now the blue-colored area reflects broad sense heritability.

Four main tenets are visualized in this figure. First, the earlier mentioned strong decrease in the contribution of environmental factors shared by siblings from childhood (age 7) to late adolescence, to the point where these factors are no longer a significant contributor (around 16 years). Second, a strong increase in contribution of genetic factors up to a peak heritability of 70% in late adolescence and early adulthood (around 20 years) that levels off to 40% at age 40. Third, the near linear increase in the influence of unique environmental factors across the life span. Fourth, spousal factors become an important component of these environmental factors from age 30 onwards and form a substantial (de)motivator to engage in exercise after age 35.

There are different possible explanations for the sharp increase in the relative importance of genetic factors during adolescence followed by a gradual decrease to middle adulthood. First, the changes in the percentage of variance explained by genetics may simply be a result of changes in the importance of common and unique environmental factors. In other words, the *true* genetic effects may not change, only their relative contribution. However, a true increase in absolute genetic variance was demonstrated from childhood to adolescence (Huppertz et al., 2016) making this explanation unlikely. That unique environmental influence start to have a cumulative effect from adolescence towards middle adulthood makes more sense and a change in the relative contribution of E and A could explain part of the gradual decrease in heritability in this period.

However, the peak in adolescence and the decline in adulthood could also be related to the change in the prevalence of team-based and competitive exercise activities. It is very high in adolescence but wanes in adulthood. Because we have found team-based and competitive exercise to have an overall higher heritability in chapter 5,

decreased participation in primarily these types of exercise activities would lead to a proportional decrease in heritability of total volume of exercise. Second, the decreasing heritability may be a direct result of different genetic effects coming into play at different ages. In other words, a *true* change in the genetic effects may occur. Such a change may be related to a behavioristic model for the genetic determinants of exercise behavior presented in de Geus (2021). Core determinants of exercise behavior are grouped in two instrumental conditioning loops continuously influencing the adoption and maintenance of regular exercise activities. One loop is related to the affective responses during and after exercise, which is linked to genetic effects on the neurobiological impact of exercise on reward/punishment circuits; the other loop is related to the rewarding effects of being able to perform well on a valued activity, which is linked to genetic effects on core elements of exercise ability including fitness, strength, low injury sensitivity and body composition.

The difference in heritability estimates found for different types of exercise activities may be the result of different genetic effects on these types of exercise. Team-based competitive activities would be favored by genetic variants influencing exercise ability whereas solitary and non-competitive activities would be favored by genetic variants influencing the affective response to exercise. In other words difference in heritability estimates can reflect that these activities are, at least partly, influenced by different genetic variants. If this is the case, then we would expect (1) to find different GWAS results for team-based and competitive exercise compared to solitary and non-competitive exercise, and given the changing proportions of different types of exercise behavior across various ages (2) to find different GWAS results for total exercise behavior depending on the age of the sample.

In line with these expectations, we find different results for the different types of exercise activities in our GWAS in chapter 6, especially when comparing team-based competitive exercise activities to solitary non-competitive activities. Furthermore, we find no direct overlap between our GWAS results in a sample with a mean age of 31 and the self-reported total volume of exercise behavior in the largest GWAS available for self-reported sports and exercise behavior in the UKB, which is a sample aged 40-69 years. Nor did we find overlap with the variants detected in the second largest GWAS by Hara et al. (2018) in participants with a mean age of 55. Furthermore, the genome-wide significant variants found in the UKB by Klimentidis et al. (2018) appeared to be largely brain-related. In contrast our GWAs in younger population, found variants related to red blood cell phenotypes.

It is attractive to think that our GWAS tapped more into the genetic variants influencing ability whereas the UKB GWAS tapped more into genetic variants influencing the affective response to exercise. There is a clear association between red blood cell physiology and exercise ability (Mairbäurl, 2013) mainly due to their role in oxygen-delivery to peripheral muscle. A case in point is the story of Eero Mantyranta, who had stellar biathlon capability due to a rare mutation in the EPO receptor gene (Enriquez & Gullans, 2012). It is therefore not surprising that red blood cell increasing substances such as erythropoietin (EPO) are banned from professional sports as performance-enhancing drugs (World Anti-Doping Agency, 2021). Thus it seems very plausible that individuals with a genetic makeup for an increased number (or improved physiology) of red blood cells perform better compared to their peers. The perception of higher ability would in turn make these individuals more likely to continue exercising as part of (competitive) team sports activities, in line with de Geus (2020) as discussed earlier.

Clearly, this is utter speculation. There are multiple other possible sources of the differences between the handful of exercise GWASs, most notably the modest sample size in our NTR GWAS. Apart from insufficient power, there is potential bias due to selective participation (both in the UKB and NTR cohorts) resulting in population stratification and poor harmonization of instruments used to establish voluntary exercise participation. The different findings might also be the result of different moderation of genetic variants across different countries (i.e., gene-environment interaction). Larger GWAS on the various types of exercise in different populations and at different ages are very much needed to progress our understanding of the molecular genetic basis of voluntary engagement in regular exercise behaviors and its change across the lifespan.

Exercise domains

The main exercise phenotype used throughout this thesis was the total amount of sports and exercise activities weekly, recast as a measure of energy expenditure on these activities. This involves the use of fixed intensity codes for specific activities based on the average values obtained in past experimental studies (Ainsworth et al., 2011; Ridley et al., 2008). This had the advantage of making our exercise phenotype comparable to those used in a large number of studies in the extant literature. A disadvantage of this measure, compared to e.g. number of minutes weekly spent on exercise, is that it introduces measurement error because individuals may perform these activities at different actual intensity levels based on differences in motivation or efficiency related to body composition and/or motor competence. However, acknowledging this disadvantage, we could now at least compare our results to a large body of past work that also used a METminutes-based score.

The additional use of energy expended on team-based, solitary, competitive, non-competitive, externally paced and internally paced exercise activities was more challenging in this regard. The inclusion of these separate exercise domains was inspired by van der Mee et al (2017) and has become a common theme throughout this thesis. As a consequence, many results presented in this thesis, i.e. those related to these domains specifically, cannot be directly compared to existing literature, as to my knowledge almost no previous genetic studies have used these domains. However, I believe the results presented in this thesis strongly justify the use of these separate domains, as they yield differential and novel findings, compared to total volume of exercise alone. Most strikingly was the differential course across the life span of team-based, competitive and solitary, non-competitive activities in chapter 2 as well the larger broad sense heritability of the former in chapter 5 due to a more substantial non-additive component.

Do all these domains need to be included in the analyses? It may depend on the application. Because of the clear meaning the different domains have for professionals and the clear role these may play in possible interventions it is my recommendation to adhere to these domains when the aim of future research is directly aimed at developing or improving exercise interventions. For genetic research, we did note substantial redundancy between team-based, competitive and externally paced activities on the one hand, and solitary, non-competitive, and internally paced activities on the other hand. Internally (self) paced versus externally paced, in particular, did not add a lot to the domain solitary vs team-based. However, when examining the role of the genetics of executive functioning on exercise preference this distinction did prove to be meaningful (van der Mee et al., 2017). Somewhat more distinction exists between the team-based vs solitary domain and the competitive and non-competitive domain, although substantial overlap was found for these dimensions too. Even

so, the inclusion of the full set of team-based versus solitary and competitive versus non-competitive may be maintained when research concerns, for instance, the role of personality in exercise (H. J. Eysenck et al., 1982).

Whereas some research questions may require all three exercise domains to be used in full, evidence from both the phenotypic and genotypic analyses performed throughout this thesis suggest that two underlying factors capture the six METminutes values across the three exercise domains rather well. These latent factors can be approximated by using the team-based versus solitary activities but an actual factor score computed across the six different exercise phenotypes, as used in chapter 7 may be optimal. On top of the inclusion of the separate domains or the two factor scores, the total volume of exercise behavior should also be included since (1) the correlations between total volume and the two factors or the full set of the six METminutes scores are only moderate and (2) the vast majority of previously published literature only includes total volume of exercise behavior, as discussed earlier, thus allowing for a more direct comparison to past findings.

Future directions for exercise phenotyping

Throughout this thesis self-report, or parent-report in children included in chapter 2, was used to measure the time spent on all exercise activities. As discussed in multiple chapters of this thesis, voluntary exercise activities are salient and thus reliably recalled by self-report in surveys. Total physical activity on the other hand is much harder to accurately recall retroactively, it is too easy to significantly over- or underestimate the time spent being physically active in a day, let alone a week. It is therefore not surprising that the accelerometer (and pedometer to a lesser extent) has all but replaced the self-report survey in studies on physical activity (Silfee et al., 2018).

Nowadays the vast majority of the population carries highly valid accelerometers with them everywhere all the time, at least one in (nearly) every smartphone, and sometimes additional ones in devices such as fitness bands or smartwatches. The large amount of data generated by these devices provides a lot of opportunity for future research. Major challenges, however, are *to obtain* the data from the often commercial suppliers like Fitbit, Google, Apple or Polar, and *to standardize* across the many proprietary versions of the hard- and software that regularly spring into life. For this reasons, many researchers have resorted to dedicated research devices for accelerometer signals, of which the Actigraph seems most popular, paired to scoring standards and open software solutions in R to do so (Migueles et al., 2021). Unfortunately, where total physical activity was harder to measure through self-report, voluntary exercise behavior cannot be assessed by relying solely on an accelerometer. Since the accelerometer only records raw movement on three axes, it is perfectly suited to measure total physical activity, however without a record or log about what type of activity was performed (and why), there is no easy way to tell which bout of activity is attributable to voluntary regular exercise behavior.

The future is bright here though, as innovative machine learning methods become increasingly capable at recognizing patterns in vast quantities of unstructured data, making them very promising methods to accurately detect exercise bouts from raw accelerometer data (Hoogendoorn & Funk, 2018). Ongoing research to make these classifiers more and more accurate are making it feasible to use accelerometers in large-scale exercise studies in the future. The benefits of this are plenty, for example we can more accurately estimate the exact amount of time an individual is engaged in exercise, and how often. Additionally, as these models develop, they will inevitably produce a better estimate of the true energy expenditure

during exercise compared to survey-based data in conjunction with MET-constants (Freedson et al., 2011; Lawal & Bano, 2020; Staudenmayer et al., 2009; Yen & Lin, 2020).

Another exciting avenue for research is brought about when accelerometers are combined with short self-report queries that are sent out on an hourly basis as part of people's daily lives, also known as ecological momentary assessment (EMA). Smartphone-based EMA can measure an individual's intent to engage in, attitudes towards *voluntary* exercise, and experienced enjoyment. Combining EMA with accelerometers can accurately assess the differences in people's reported or intended exercise activities, acute psychological effects of exercise, and momentary determinants of exercise, together with the objective volume of exercise they engage in (Dunton et al., 2015; Liao et al., 2017; Liao et al., 2015; Maher et al., 2017).

In spite of these exiting developments, survey-based research into exercise behavior will retain its value in the foreseeable future. No matter how good the accelerometers and the accompanying (machine learning) models become at estimating time and energy spent on exercise, the scale of the samples required for genetic analyses may not be easy to obtain with these methods due to the challenges mentioned previously. Also, various trait-like characteristics are still well captured by surveys like personality or attitudes. One other relevant trait that we recently tackled is an individual's exercise *liking*. In over 157,000 individuals from the UK Biobank, we sought to complement and extend previous findings on the genetics of physical activity behaviors by performing genome-wide association studies of liking of several exercise-related behaviors plus an additional derived overall activity-liking trait. We identified a total of 17 loci, along with an additional eight for the overall trait, only some of which overlap with loci previously identified for physical activity behavior. Replication in over 7,000 adults from the NTR showed directional consistency in 13

out of 17 loci (Klimentidis et al., 2021). If the aim of a future study is to aid the development of more personalized exercise interventions, an individual's exercise liking and attitudes towards the behavior are, in my opinion, at least as important (if not more so) as the objective energy expenditure.

Future directions for finding molecular determinants of exercise behavior

As sample sizes increase, future GWASs will be able to tell us more about the complex genetic underpinnings of exercise behavior, particularly when cohorts combine forces in meta-analyses. Throughout this thesis I have shown that these meta-analyses would do well to take the changing composition of total exercise behavior into account. 'Total exercise' at age 20 is not strictly the same behavior as 'total exercise' at age 60, which may make meta-analysis of studies across different ages more challenging. Additionally, as discussed previously, a GWAS of exercise activities is likely sensitive to selection bias compounded by population stratification. Luckily the development of genetic methods does not seem to be slowing down and new methods to tackle issues like this are being developed or have already been developed. For example to tackle the issue of different compositions of total volume of exercise across the various dimensions, future research could employ the multivariate GWAS meta-analysis approach as described by Baselmans et al. (2019). Additionally, to tackle the issue of stratification or indirect genetic effects future research could use a within-sibling GWAS (Howe et al., 2021).

Variations in the DNA code, the focus in this thesis, are not the only possible molecular underpinnings of voluntary exercise. There is likely a strong relationship between epigenetic mechanisms and voluntary exercise behavior as well (van Roekel et al., 2019; Voisin et al.,

2015). Epigenetics does bring up the thorny issue of causality. In genetics, the direction of causality is simple, as genetic variants can (indirectly) affect behavior in many ways but the reverse is near impossible (barring extremely rare genetic mutations due to things like radiation). In epigenetics, this is not as simple because epigenetic changes such as DNA methylation or histone modification may well have a causal effect on exercise behavior, but the reverse may just as well hold, or the effect may be bidirectional creating an epigenetic feedback loop (Lightfoot, 2020). Further complicating the epigenetics of exercise is the fact that epigenetic profiles are different in different cell types (Maurano et al., 2012). Nevertheless, I believe epigenetics has great potential in exercise behavior as it can lead to more insights into the molecular underpinnings of the health benefits of exercise. Longitudinal epigenetic assessments combined with longitudinal measures of (objective) exercise behaviors can likely aid in distinguishing these two directions of causality. Using twin samples this can be further aided by an MR-DoC design.

Another avenue through which we may well learn more about the biological underpinnings of the health benefits of exercise behavior is the gut-microbiome. The interest in the complex ecosystem of bacteria and other micro-organisms in the gut is relatively young and the immense diversity of the gut-microbiome with over 3 million unique genes from over 1000 unique bacterial species (Qin et al., 2010) poses significant challenges for future research. This diversity is in turn affected by a combination of host genetics (Kurilshikov et al., 2021), and numerous environmental factors (Rothschild et al., 2018), and these effects can vary between species, posing further challenges. Despite these challenges, however, numerous studies in both humans and animals have already found evidence for associations between exercise and the gut-microbiome (see Mailing et al. (2019) and Dalton et al. (2019) for review).

To summarize, I think there is a lot to look forward to in the field of exercise genetics. With the combined growth in popularity of daily-wear accelerometers, advancements in methods to accurately label this data and advancements in EMA methods, there is a wealth of new data just on the horizon. By combining this data with novel and increasingly bias-resistant genetic methods in ever increasing sample sizes, and expanding from the genomic to other multi-omics levels, we stand to learn a lot more about the molecular determinants of exercise behaviors. Combining these results across cohorts in large-scale meta-analyses will yield the required statistical power to detect the small effects that are inherent to any highly polygenetic behavior. If, on top of that, future research also analyzes the different types of exercise activities separately, this will likely not only yield more scientific findings, but also provide avenues for personalized interventions that have higher success rates than those currently on offer.

Appendix

References

Aaltonen, S., Ortega-Alonso, A., Kujala, U. M., & Kaprio, J. (2010). A longitudinal study on genetic and environmental influences on leisure time physical activity in the Finnish Twin Cohort. *Twin Research and Human Genetics, 13*(5), 475-481.

Aaltonen, S., Ortega-Alonso, A., Kujala, U. M., & Kaprio, J. (2013). Genetic and environmental influences on longitudinal changes in leisure-time physical activity from adolescence to young adulthood. *Twin Research and Human Genetics, 16*(2), 535-543.

Aarnio, M., Winter, T., Peitonen, J., Kujala, E. M., & Kaprio, J. (2002). Stability of leisure-time physical activity during adolescence – a longitudinal study among 16-. 17-, and 18-year-old Finnish youth. *Scand J Med Sci Sports, 12*.

Aasdahl, L., Nilsen, T. I. L., Meisingset, I., Nordstoga, A. L., Evensen, K. A. I., Paulsen, J., . . . Skarpsno, E. S. (2021). Genetic variants related to physical activity or sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 18*(1), 15.

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. *Nat Genet, 30*(1), 97-101.

Ahola-Olli, A. V., Würtz, P., Havulinna, A. S., Aalto, K., Pitkänen, N., Lehtimäki, T., . . . Raitakari, O. T. (2017). Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet, 100*(1), 40-50.

Ainsworth, B. E., Haskell, W. L., Herrmann, S. D., Meckes, N., Bassett, D. R., Jr., Tudor-Locke, C., . . . Leon, A. S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc, 43*(8), 1575-1581.

Allen, M. S., & Laborde, S. (2014). The role of personality in sport and physical activity. *Current directions in psychological science, 23*(6), 460-465.

Anderssen, N., Jacobs Jr, D. R., Sidney, S., Bild, D. E., Stemfeld, B., Slattery, M. L., & Hannan, P. (1996). Change and secular trends in physical activity patterns in young adults: a seven-year longitudinal follow-up in the Coronary Artery Risk Development in Young Adults Study (CARDIA). *Am J Epidemiol, 143*(4), 351-362.

Anusic, I., & Schimmack, U. (2016). Stability and change of personality traits, self-esteem, and well-being: Introducing the meta-analytic stability and change model of retest correlations. *Journal of personality and social psychology, 110*(5), 766.

Ashdown-Franks, G., Firth, J., Carney, R., Carvalho, A. F., Hallgren, M., Koyanagi, A., . . . Stubbs, B. (2020). Exercise as Medicine for Mental and Substance Use Disorders: A Meta-review of the Benefits for Neuropsychiatric and Cognitive Outcomes. *Sports medicine, 50*(1), 151-170.

Astle, W. J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A. L., . . . Soranzo, N. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell, 167*(5), 1415-1429.e1419.

Aune, D., Norat, T., Leitzmann, M., Tonstad, S., & Vatten, L. J. (2015). Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. *Eur J Epidemiol, 30*(7), 529-542.

Azevedo, M. R., Araújo, C. L., Reichert, F. F., Siqueira, F. V., da Silva, M. C., & Hallal, P. C. (2007). Gender differences in leisure-time physical activity. *Int J Public Health, 52*(1), 8-15.

Barnett, T. A., Gauvin, L., Craig, C. L., & Katzmarzyk, P. T. (2008). Distinct trajectories of leisure time physical activity and predictors of trajectory class membership: a 22 year cohort study. *International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 57.

Bartels, M., de Moor, M. H. M., Van der Aa, N., Boomsma, D. I., & de Geus, E. J. C. (2012). Regular exercise, subjective wellbeing, and internalizing problems in adolescence: causality or genetic pleiotropy? *Frontiers in genetics, 3*, 4.

Baselmans, B. M. L., Jansen, R., Ip, H. F., van Dongen, J., Abdellaoui, A., van de Weijer, M. P., . . . Social Science Genetic Association, C. (2019). Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics, 51*(3), 445-451.

Bassett Jr, D. R., Cureton, A. L., & Ainsworth, B. E. (2000). Measurement of daily walking distance-questionnaire versus pedometer. *Med Sci Sports Exerc, 32*(5), 1018-1023.

Bateson, W., & Mendel, G. (2013). *Mendel's principles of heredity*: Courier Corporation.

Bauman, A. E., Brown, W., & Plotnifkoff, R. (2016). Re: Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response metaanalysis for the Global Burden of Disease Study 2013. . *BMJ, 354*, i3857.

Bauman, A. E., & Bull, F. (2007). Environmental correlates of physical activity and walking in adults and children: a review of reviews. *London: National Institute of Health and Clinical Excellence*.

Beets, M. W., Cardinal, B. J., & Alderman, B. L. (2010). Parental social support and the physical activity-related behaviors of youth: a review. *Health Educ Behav, 37*(5), 621-644.

Berentzen, T., Kring, S. I. I., Holst, C., Zimmermann, E., Jess, T., Hansen, T., . . . Sørensen, T. I. A. (2008). Lack of Association of Fatness-Related FTO Gene Variants with Energy Expenditure or Physical Activity. *The Journal of Clinical Endocrinology & Metabolism, 93*(7), 2904-2908.

Bergen, S. E., O'Dushlaine, C. T., Ripke, S., Lee, P. H., Ruderfer, D. M., Akterin, S., . . . Sullivan, P. F. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol Psychiatry, 17*(9), 880-886.

Berglund, B., & Hemmingson, P. (1987). Effect of reinfusion of autologous blood on exercise performance in cross-country skiers. *Int J Sports Med, 8*(3), 231-233.

Berlin, J. A., & Colditz, G. A. (1990). A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol, 132*(4), 612-628.

Beunen, G., & Thomis, M. (1999). Genetic determinants of sports participation and daily physical activity. *International journal of obesity, 23*(3), S55-S63.

Boomsma, D. I., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3*(11), 872-882.

Boomsma, D. I., Helmer, Q., Nieuwboer, H. A., Hottenga, J. J., de Moor, M. H. M., van den Berg, S. M., . . . Dolan, C. V. (2018). An extended twin-pedigree study of neuroticism in the Netherlands Twin Register. *Behav Genet, 48*(1), 1-11.

Boomsma, D. I., Van den Bree, M., Orlebeke, J., & Molenaar, P. (1989). Resemblances of parents and twins in sports participation and heart rate. *Behav Genet, 19*(1), 123-141.

Booth, F. W., Laye, M. J., Lees, S. J., Rector, R. S., & Thyfault, J. P. (2008). Reduced physical activity and risk of chronic disease: the biology behind the consequences. *Eur J Appl Physiol, 102*(4), 381-390.

Bouchard Jr, T. J., & McGue, M. (2003). Genetic and environmental influences on human psychological differences. *Journal of neurobiology, 54*(1), 4-45.

Boulé, N. G., Haddad, E., Kenny, G. P., Wells, G. A., & Sigal, R. J. (2001). Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. *Jama, 286*(10), 1218-1227.

Boutwell, B., Hinds, D., Agee, M., Alipanahi, B., Auton, A., Bell, R. K., . . . Furlotte, N. A. (2017). Replication and characterization of CADM2 and MSRA genes on human behavior. *Heliyon, 3*(7), e00349.

Breeze, C. E., Haugen, E., Reynolds, A., Teschendorff, A., van Dongen, J., Lan, Q., . . . Berndt, S. I. (2020). FORGE2: Integrative analysis of GWAS data.[manuscript submitted].

Brown, W., Bauman, A. E., Bull, F., & Burton, N. (2012). *Physical activity recommendations for adults (18–64 years).* Canberra, Australia: Department of Health.

Buford, T. W., Roberts, M. D., & Church, T. S. (2013). Toward exercise as personalized medicine. *Sports medicine, 43*(3), 157-165.

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., . . . Consortium, S. W. G. o. t. P. G. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics, 47*(3), 291-295.

Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., . . . Sollis, E. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research, 47*(D1), D1005-D1012.

Burton, D. (1988). The dropout dilemma in youth sports: Documenting the problem and identifying solutions. *Young athletes Biological, Psychological and Educational Perspectives. Champaign, IL: Human Kinetics*, 245-266.

Cai, G., Cole, S. A., Butte, N., Bacino, C., Diego, V., Tan, K., . . . Comuzzie, A. G. (2006). A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. *Obesity, 14*(9), 1596-1604.

Cai, N., Revez, J. A., Adams, M. J., Andlauer, T. F. M., Breen, G., Byrne, E. M., . . . Consortium, M. D. D. W. G. o. t. P. G. (2020). Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nature genetics, 52*(4), 437-447.

Carlsson, S., Andersson, T., Lichtenstein, P., Michaelsson, K., & Ahlbom, A. (2006). Genetic effects on physical activity: results from the Swedish Twin Registry. *Med Sci Sports Exerc, 38*(8), 1396.

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 & Science in Sports & Exercise, 32*(9), 1601-1609.

Caspi, A., & Herbener, E. S. (1993). Marital assortment and phenotypic convergence: Longitudinal evidence. *Social Biology, 40*(1-2), 48-60.

Cesarini, D., & Visscher, P. M. (2017). Genetics and educational attainment. *npj Science of Learning, 2*(1), 4.

Chan, D. K., Lonsdale, C., & Fung, H. H. (2012). Influences of coaches, parents, and peers on the motivational patterns of child and adolescent athletes. *Scandinavian journal of medicine & science in sports, 22*(4), 558-568.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*(1), 7.

Chapman, B. P., Hampson, S., & Clarkin, J. (2014). Personality-informed interventions for healthy aging: Conclusions from a National Institute on Aging work group. *Developmental psychology, 50*(5), 1426.

Chekroud, S. R., Gueorguieva, R., Zheutlin, A. B., Paulus, M., Krumholz, H. M., Krystal, J. H., & Chekroud, A. M. (2018). Association between physical exercise and mental health in 1· 2 million individuals in the USA between 2011 and 2015: a cross-sectional study. *The Lancet Psychiatry, 5*(9), 739-746.

Chen, L. S., Saccone, N. L., Culverhouse, R. C., Bracci, P. M., Chen, C. H., Dueker, N., . . . Kohno, T. (2012). Smoking and Genetic Risk Variation Across Populations of E uropean, A sian, and A frican A merican Ancestry—A Meta-Analysis of Chromosome 15q25. *Genetic epidemiology, 36*(4), 340-351.

Chen, M. H., Raffield, L. M., Mousas, A., Sakaue, S., Huffman, J. E., Moscati, A., . . . Lettre, G. (2020). Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell, 182*(5), 1198-1213.e1114.

Choh, A. C., Demerath, E. W., Lee, M., Williams, K. D., Towne, B., Siervogel, R. M., . . . Czerwinski, S. A. (2009). Genetic analysis of self-reported physical activity and adiposity: the Southwest Ohio Family Study. *Public health nutrition, 12*(8), 1052-1060.

Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., . . . MacArthur, D. G. (2020). A brief history of human disease genetics. *Nature, 577*(7789), 179-189.

Cole, S. A., Butte, N. F., Voruganti, V. S., Cai, G., Haack, K., Kent, J. W., Jr., . . . Gibbs, R. A. (2010). Evidence that multiple genetic variants of MC4R play a functional role in the regulation of energy expenditure and appetite in Hispanic children. *Am J Clin Nutr, 91*(1), 191-199.

Comuzzie, A. G., Cole, S. A., Laston, S. L., Voruganti, V. S., Haack, K., Gibbs, R. A., & Butte, N. F. (2012). Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PLoS One, 7*(12), e51954.

Costa Jr, P. T., McCrae, R. R., & Löckenhoff, C. E. (2019). Personality across the life span. *Annual review of psychology, 70*, 423-448.

Costa, P. T., & McCrea, R. R. (1992). Revised neo personality inventory (neo pi-r) and neo five-factor inventory (neo-ffi): Psychological Assessment Resources.

Courneya, K. S., & Hellsten, L.-A. M. (1998). Personality correlates of exercise behavior, motives, barriers and preferences: An application of the five-factor model. *Personality Individual differences, 24*(5), 625-633.

Coventry, W. L., & Keller, M. C. (2005). Estimating the extent of parameter bias in the classical twin design: a comparison of parameter estimates from extended twin-family and classical twin designs. *Twin Res Hum Genet, 8*(3), 214-223.

Craig, C. L., Marshall, A. L., Sjostrom, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., . . . Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc, 35*(8), 1381-1395.

D'Haese, S., Van Dyck, D., De Bourdeaudhuij, I., Deforche, B., & Cardon, G. (2014). The association between objective walkability, neighborhood socio-economic status, and physical activity in Belgian children. *Int J Behav Nutr Phys Act, 11*, 104-104.

da Silva, A. R. (1980). Structure and development of 11-to-19-years-old's motivation for physical activity. *Arquivos Brasileiros de Psicologia, 32*(4), 128-131.

Dalton, A., Mermier, C., & Zuhl, M. (2019). Exercise influence on the microbiome–gut–brain axis. *Gut Microbes, 10*(5), 555-568.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10*(2), giab008.

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., . . . Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nat Genet, 48*(10), 1284-1287.

Davey Smith, G., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol, 32*(1), 1-22.

de Geus, E. J. C. (2020). A genetic perspective on the association between exercise and mental health in the era of genome-wide association studies. *Mental Health and Physical Activity*, 100378.

de Geus, E. J. C., Bartels, M., Kaprio, J., Lightfoot, J. T., & Thomis, M. (2014). Genetics of regular exercise and sedentary behaviors. *Twin Research and Human Genetics, 17*(04), 262-271.

de Geus, E. J. C., Boomsma, D. I., & Snieder, H. (2003). Genetic correlation of exercise with heart rate and respiratory sinus arrhythmia. *Med Sci Sports Exerc, 35*(8), 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, 1*(2), 53-61.

de Geus, E. J. C., & de Moor, M. H. M. (2011). Genes, exercise, and psychological factors. *Genetic and molecular aspects of sport performance*, 294-305.

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. *Prev Med, 42*(4), 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. *Arch Gen Psychiatry, 65*(8), 897-905.

de Moor, M. H. M., & de Geus, E. J. C. (2012). Genetic influences on exercise behavior. In R. J.M. (Ed.), *Lifestyle Medicine* (pp. 1367-1378). Boca Raton, FL: Taylor & Francis Group, LLC.

de Moor, M. H. M., & de Geus, E. J. C. (2018). Causality in the associations between exercise, personality, and mental health. In Budde H. & M. Wegner (Eds.), *The Exercise Effect on Mental Health* (pp. 67-99). Boca Raton: CRC Press.

de Moor, M. H. M., Liu, Y. J., Boomsma, D. I., Li, J., Hamilton, J. J., Hottenga, J. J., . . . Deng, H. W. (2009). Genome-wide association study of exercise behavior in Dutch and American adults. *Med Sci Sports Exerc, 41*(10), 1887-1895.

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. *Eur J Hum Genet, 15*(12), 1252-1259.

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? *Eur J Epidemiol, 22*(1), 27-32.

de Moor, M. H. M., van den Berg, S. M., Verweij, K. J., Krueger, R. F., Luciano, M., Arias Vasquez, A., . . . Boomsma, D. I. (2015). Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA Psychiatry, 72*(7), 642-650.

de Moor, M. H. M., Willemsen, G., Rebollo-Mesa, I., Stubbe, J. H., de Geus, E. J. C., & Boomsma, D. I. (2011). Exercise participation in adolescents and their parents: evidence for genetic and generation specific environmental effects. *Behav Genet, 41*(2), 211-222.

den Hoed, M., Brage, S., Zhao, J. H., Westgate, K., Nessa, A., Ekelund, U., . . . Loos, R. J. (2013). Heritability of objectively assessed daily physical activity and sedentary behavior. *The American journal of clinical nutrition, 98*(5), 1317-1325.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2004). Effects of censoring on parameter estimates and power in genetic modeling. *Twin Research and Human Genetics, 7*(6), 659-669.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual review of psychology, 41*(1), 417-440.

Ding, D., Lawson, K. D., Kolbe-Alexander, T. L., Finkelstein, E. A., Katzmarzyk, P. T., van Mechelen, W., . . . Lancet Physical Activity Series 2 Executive Committee. (2016). The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet, 388*(10051), 1311-1324.

Dishman, R. K., Sallis, J. F., & Orenstein, D. R. (1985). The determinants of physical activity and exercise. *Public health reports, 100*(2), 158.

Doherty, A., Smith-Byrne, K., Ferreira, T., Holmes, M. V., Holmes, C., Pulit, S. L., & Lindgren, C. M. (2018). GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nature communications, 9*(1), 5257.

Dolan, C. V., de Kort, J. M., van Beijsterveldt, T. C., Bartels, M., & Boomsma, D. I. (2014). GE covariance through phenotype to environment transmission: an assessment in longitudinal twin data and application to childhood anxiety. *Behav Genet, 44*(3), 240-253.

Duncan, G. E., Cash, S. W., Horn, E. E., & Turkheimer, E. (2015). Quasi-causal associations of physical activity and neighborhood walkability with body mass index: a twin study. *Prev Med, 70*, 90-95.

Duncan, L. E., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., . . . Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications, 10*(1), 3328.

Dunham, I., Kulesha, E., Iotchkova, V., Morganella, S., & Birney, E. (2015). FORGE: A tool to discover cell specific enrichments of GWAS associated SNPs in regulatory regions. *F1000Research, 4*(18), 18.

Dunton, G. F., Liao, Y., Intille, S., Huh, J., & Leventhal, A. (2015). Momentary assessment of contextual influences on affective response during physical activity. *Health Psychology, 34*(12), 1145.

Eagleton, J. R., McKelvie, S. J., & de Man, A. (2007). Extra version and neuroticism in team sport participants, individual sport participants, and nonparticipants. *Journal of Perceptual motor skills, 105*(1), 265-275.

Eaves, L. J. (1979). The use of twins in the analysis of assortative mating. *Heredity, 43*(3), 399.

Eaves, L. J. (2009). Putting the 'human'back in genetics: modeling the extended kinships of twins. *Twin Research and Human Genetics, 12*(1), 1-7.

Eaves, L. J., Fulker, D., & Heath, A. (1989). The effects of social homogamy and cultural inheritance on the covariances of twins and their parents: A LISREL model. *Behav Genet, 19*(1), 113-122.

Eime, R. M., Harvey, J. T., Charity, M. J., & Payne, W. R. (2016). Population levels of sport participation: implications for sport policy. *BMC public health, 16*, 752-752.

Ekelund, U., Steene-Johannessen, J., Brown, W. J., Fagerland, M. W., Owen, N., Powell, K. E., . . . Group1, L. S. B. W. (2016). Does physical activity attenuate, or even eliminate, the detrimental association of sitting time with mortality? A harmonised meta-analysis of data from more than 1 million men and women. *The Lancet, 388*(10051), 1302-1310.

Ekelund, U., Ward, H. A., Norat, T., Luan, J. a., May, A. M., Weiderpass, E., . . . Tjønneland, A. (2015). Physical activity and all-cause mortality across levels of overall and abdominal adiposity in European men and women: the European Prospective Investigation into Cancer and Nutrition Study (EPIC). *The American journal of clinical nutrition, 101*(3), 613-621.

Enriquez, J., & Gullans, S. (2012). Genetically enhanced Olympics are coming. *Nature, 487*(7407), 297-297.

Eriksson, M., Rasmussen, F., & Tynelius, P. (2006). Genetic factors in physical activity and the equal environment assumption–the Swedish young male twins study. *Behav Genet, 36*(2), 238-247.

Eysenck, H. J., Nias, D. K., & Cox, D. (1982). Sport and personality. *Advances in behaviour research and therapy, 4*(1), 1-56.

Eysenck, S., & Eysenck, H. J. (1968). The measurement of psychoticism: a study of factor stability and reliability. *British Journal of Social Clinical Psychology, 7*(4), 286-294.

Falconer, D., & Mackay, T. (1996). Introduction to quantitative genetics. 1996. *Harlow, Essex, UK: Longmans Green, 3*.

Fedko, I. O., Hottenga, J.-J., Medina-Gomez, C., Pappa, I., van Beijsterveldt, C. E., Ehli, E. A., . . . Swertz, M. A. (2015). Estimation of genetic relationships between individuals across cohorts and platforms: application to childhood height. *Behav Genet, 45*(5), 514-528.

Fishman, E. I., Steeves, J. A., Zipunnikov, V., Koster, A., Berrigan, D., Harris, T. A., & Murphy, R. (2016). Association between objectively measured physical activity and mortality in NHANES. *Med Sci Sports Exerc, 48*(7), 1303.

Flemming, E. G. (1934). Personality and the athletic girl. *School and Society*

Fortier, M. D., Katzmarzyk, P. T., Malina, R. M., & Bouchard, C. (2001). Seven-year stability of physical activity and musculoskeletal fitness in the Canadian population. *Med Sci Sports Exerc, 33*(11), 1905-1911.

Francis, S. L., Morrissey, J. L., Letuchy, E. M., Levy, S. M., & Janz, K. F. (2013). Ten-year objective physical activity tracking: Iowa Bone Development Study. *Med Sci Sports Exerc, 45*(8), 1508.

Frederiksen, H., & Christensen, K. (2003). The influence of genetic factors on physical functioning and exercise in second half of life. *Scandinavian journal of medicine & science in sports, 13*(1), 9-18.

Freedson, P. S., Lyden, K., Kozey-Keadle, S., & Staudenmayer, J. (2011). Evaluation of artificial neural network algorithms for predicting METs and

activity type from accelerometer data: validation on an independent sample. *Journal of Applied Physiology, 111*(6), 1804-1812.

Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2015). minimac2: faster genotype imputation. *Bioinformatics, 31*(5), 782-784.

Fuentes, R. M., Perola, M., Nissinen, A., & Tuomilehto, J. (2002). ACE gene and physical activity, blood pressure, and hypertension: a population study in Finland. *J Appl Physiol (1985), 92*(6), 2508-2512.

Galligan, F. (2000). *Advanced pe for edexcel*: Heinemann.

Garber, C. E., Blissmer, B., Deschenes, M. R., Franklin, B. A., Lamonte, M. J., Lee, I. M., . . . American College of Sports Medicine. (2011). Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc, 43*(7), 1334-1359.

Genome of the Netherlands Consortium. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature genetics, 46*(8), 818-825.

Gidlow, C., Johnston, L. H., Crone, D., Ellis, N., & James, D. (2006). A systematic review of the relationship between socio-economic position and physical activity. *Health Education Journal, 65*(4), 338-367.

Gielen, M., Westerterp-Plantenga, M. S., Bouwman, F. G., Joosen, A. M. C. P., Vlietinck, R., Derom, C., . . . Westerterp, K. R. (2014). Heritability and genetic etiology of habitual physical activity: a twin study with objective measures. *Genes & Nutrition, 9*(4), 415.

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1· 9 million participants. *The Lancet Global Health, 6*(10), e1077-e1086.

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. *Prev Med, 39*(1), 182-190.

Hagstromer, M., Ainsworth, B. E., Oja, P., & Sjostrom, M. (2010). Comparison of a subjective and an objective measure of physical activity in a population sample. *Journal of Physical Activity and Health, 7*(4), 541-550.

Hagströmer, M., Troiano, R. P., Sjöström, M., & Berrigan, D. (2010). Levels and patterns of objectively assessed physical activity—a comparison between Sweden and the United States. *Am J Epidemiol, 171*(10), 1055-1064.

Hakanen, M., Raitakari, O. T., Lehtima¨ki, T., Peltonen, N., Pahkala, K., Sillanma¨ki, L., . . . Ro¨nnemaa, T. (2009). FTO Genotype Is Associated with Body Mass Index after the Age of Seven Years But Not with Energy Intake or Leisure-Time Physical Activity. *The Journal of Clinical Endocrinology & Metabolism, 94*(4), 1281-1287.

Hallal, P. C., Andersen, L. B., Bull, F. C., Guthold, R., Haskell, W., Ekelund, U., & Lancet Physical Activity Series Working Group. (2012). Global physical activity levels: surveillance progress, pitfalls, and prospects. *The Lancet, 380*(9838), 247-257.

Haller, N., Lorenz, S., Pfirrmann, D., Koch, C., Lieb, K., Dettweiler, U., . . . Jung, P. (2018). Individualized Web-Based Exercise for the Treatment of Depression: Randomized Controlled Trial. *JMIR Ment Health, 5*(4), e10698.

Hallgren, M., Vancampfort, D., Giesen, E. S., Lundin, A., & Stubbs, B. (2017). Exercise as treatment for alcohol use disorders: systematic review and meta-analysis. *British journal of sports medicine, 51*(14), 1058-1064.

Hampson, S. E., & Goldberg, L. R. (2006). A first large cohort study of personality trait stability over the 40 years between elementary school and midlife. *Journal of personality and social psychology, 91*(4), 763.

Hansen, B. H., Kolle, E., Dyrstad, S. M., Holme, I., & Anderssen, S. A. (2012). Accelerometer-determined physical activity in adults and older people. *Med Sci Sports Exerc, 44*(2), 266-272.

Hara, M., Hachiya, T., Sutoh, Y., Matsuo, K., Nishida, Y., Shimanoe, C., . . . Kawaguchi, T. (2018). Genomewide association study of leisure-time exercise behavior in Japanese adults. *Med Sci Sports Exerc, 50*(12), 2433.

Haskell, W. L., Lee, I. M., Pate, R. R., Powell, K. E., Blair, S. N., Franklin, B. A., . . . Bauman, A. E. (2007). Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc, 39*(8), 1423-1434.

Hasson, R. E., Brown, D. R., Dorn, J., Barkley, L., Torgan, C., Whitt-Glover, M., . . . Keith, N. (2017). Achieving Equity in Physical Activity Participation: ACSM Experience and Next Steps. *Med Sci Sports Exerc, 49*(4), 848-858.

Heath, A. C., & Eaves, L. J. (1985). Resolving the effects of phenotype and social background on mate selection. *Behav Genet, 15*(1), 15-30.

Heller, R., O'Connell, D., Roberts, D., Allen, J., Knapp, J., Steele, P., . . . Vogler, G. (1988). Lifestyle factors in monozygotic and dizygotic twins. *Genetic epidemiology, 5*(5), 311-321.

Hirvensalo, M., & Lintunen, T. (2011). Life-course perspective for physical activity and sports participation. *European Review of Aging and Physical Activity, 8*(1), 13-22.

Hoekstra, H. A., Ormel, J., & De Fruyt, F. d. (1996). *NEO persoonlijkheids vragenlijsten: NEO-PI-R: NEO-FFI*: Swets Test Services (STS).

Hoogendoorn, M., & Funk, B. (2018). Machine learning for the quantified self. *On the art of learning from sensory data*.

Horimoto, A. R., Giolo, S. R., Oliveira, C. M., Alvim, R. O., Soler, J. P., De Andrade, M., . . . Pereira, A. C. (2011). Heritability of physical activity traits in Brazilian families: the Baependi Heart Study. *BMC Med Genet, 12*(1), 155.

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., . . . Davies, N. M. (2021). Within-sibship GWAS improve estimates of direct genetic effects. *bioRxiv*, 2021.2003.2005.433935.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet, 44*(8), 955-959.

Hubáček, J. A., Pikhart, H., Peasey, A., Kubínová, R., & Bobák, M. (2011). FTO variant, energy intake, physical activity and basal metabolic rate in Caucasians. The HAPIEE study. *Physiol Res, 60*(1), 175-183.

Huppertz, C., Bartels, M., de Geus, E. J. C., van Beijsterveldt, C. E., Rose, R. J., Kaprio, J., & Silventoinen, K. (2017). The effects of parental education on exercise behavior in childhood and youth: a study in Dutch and Finnish twins. *Scandinavian journal of medicine & science in sports, 27*(10), 1143-1156.

Huppertz, C., Bartels, M., de Zeeuw, E. L., van Beijsterveldt, C. E., Hudziak, J. J., Willemsen, G., . . . de Geus, E. J. C. (2016). Individual Differences in Exercise Behavior: Stability and Change in Genetic and Environmental Determinants From Age 7 to 18. *Behav Genet, 46*(5), 665-679.

Huppertz, C., Bartels, M., Groen-Blokhuis, M. M., Dolan, C. V., de Moor, M. H. M., Abdellaoui, A., . . . Willemsen, G. (2014). The dopaminergic reward system and leisure time exercise behavior: a candidate allele study. *BioMed research international, 2014*.

Huppertz, C., Bartels, M., Jansen, I. E., Boomsma, D. I., Willemsen, G., de Moor, M. H. M., & de Geus, E. J. C. (2014). A twin-sibling study on the relationship between exercise attitudes and exercise behavior. *Behav Genet, 44*(1), 45-55.

Huppertz, C., Bartels, M., van Beijsterveldt, C., Boomsma, D. I., Hudziak, J. J., & de Geus, E. J. C. (2012). The impact of shared environmental factors on exercise behavior from age 7 to 12 years. *Med Sci Sports Exerc, 44*(10), 2025-2032.

Ibrahim-Verbaas, C., Bressler, J., Debette, S., Schuur, M., Smith, A., Bis, J., . . . Wolf, C. (2016). GWAS for executive function and processing speed suggests involvement of the CADM2 gene. *Mol Psychiatry, 21*(2), 189-197.

Ingledew, D. K., & Markland, D. (2008). The role of motives in exercise participation. *Psychology and health, 23*(7), 807-828.

Jang, K. L., Livesley, W. J., & Vemon, P. A. (1996). Heritability of the big five personality dimensions and their facets: a twin study. *Journal of personality, 64*(3), 577-592.

Janssen, I., & Leblanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *Int J Behav Nutr Phys Act, 7*, 40.

Jayakody, K., Gunadasa, S., & Hosker, C. (2014). Exercise for anxiety disorders: systematic review. *British journal of sports medicine, 48*(3), 187-196.

Joosen, A. M., Gielen, M., Vlietinck, R., & Westerterp, K. R. (2005). Genetic analysis of physical activity in twins. *The American journal of clinical nutrition, 82*(6), 1253-1259.

Jöreskog, K. G. (1970). A general method for estimating a linear structural equation system. *ETS Research Bulletin Series, 1970*(2), i-41.

Jozkow, P., Slowinska-Lisowska, M., Laczmanski, L., & Medras, M. (2013). DRD2 C313T and DRD4 48-bp VNTR polymorphisms and physical activity of healthy men in Lower Silesia, Poland (HALS study). *Ann Hum Biol, 40*(2), 186-190.

Karvinen, S., Waller, K., Silvennoinen, M., Koch, L. G., Britton, S. L., Kaprio, J., . . . Kujala, U. M. (2015). Physical activity in adulthood: genes and mortality. *Scientific reports, 5*(1), 1-9.

Kawai, T., Morita, K., Masuda, K., Nishida, K., Sekiyama, A., Teshima-Kondo, S., . . . Rokutan, K. (2007). Physical exercise-associated gene expression signatures in peripheral blood. *Clinical journal of sport medicine, 17*(5), 375-383.

Keller, M. C., Medland, S. E., Duncan, L. E., Hatemi, P. K., Neale, M. C., Maes, H. H., & Eaves, L. J. (2009). Modeling extended twin family data I: description of the Cascade model. *Twin Res Hum Genet, 12*(1), 8-18.

Kichaev, G., Bhatia, G., Loh, P. R., Gazal, S., Burch, K., Freund, M. K., . . . Price, A. L. (2019). Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet, 104*(1), 65-75.

Kim, J., Kim, J., Min, H., Oh, S., Kim, Y., Lee, A. H., & Park, T. (2014). Joint identification of genetic variants for physical activity in Korean population. *International journal of molecular sciences, 15*(7), 12407-12421.

Kirkcaldy, B., & Furnham, A. (1991). Extraversion, neuroticism, psychoticism and recreational choice. *Personality and Individual Differences, 12*(7), 737-745.

Kjønniksen, L., Torsheim, T., & Wold, B. (2008). Tracking of leisure-time physical activity during adolescence and young adulthood: a 10-year longitudinal study. *International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 69.

Klein, D. N., Kotov, R., & Bufferd, S. J. (2011). Personality and Depression: Explanatory Models and Review of the Evidence. *Annual Review of Clinical Psychology, 7*(1), 269-295.

Klimentidis, Y. C., Newell, M., van der Zee, M. D., Bland, V. L., May-Wilson, S., Arora, A., . . . Pirastu, N. (2021). *Genome-wide association study of liking of physical activity in the UK Biobank.* Paper presented at the BGA2021.

Klimentidis, Y. C., Raichlen, D. A., Bea, J., Garcia, D. O., Wineinger, N. E., Mandarino, L. J., . . . Going, S. B. (2018). Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *International journal of obesity, 42*(6), 1161-1176.

Knapen, J., Vancampfort, D., Moriën, Y., & Marchal, Y. (2015). Exercise therapy improves both mental and physical health in patients with major depression. *Disability and Rehabilitation, 37*(16), 1490-1495.

Kohl, H. W., 3rd, Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., . . . Lancet Physical Activity Series Working Group. (2012). The pandemic of physical inactivity: global action for public health. *Lancet, 380*(9838), 294-305.

Koopmans, J. R., van Doornen, L. J., & Boomsma, D. I. (1994). Smoking and sports participation. In *Genetic factors in coronary heart disease* (pp. 217-235): Springer.

Kostrzewa, E., Brandys, M. K., van Lith, H. A., & Kas, M. J. (2015). A candidate syntenic genetic locus is associated with voluntary exercise levels in mice and humans. *Behav Brain Res, 276*, 8-16.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychological bulletin, 136*(5), 768.

Köttgen, A., Pattaro, C., Böger, C. A., Fuchsberger, C., Olden, M., Glazer, N. L., . . . Fox, C. S. (2010). New loci associated with kidney function and chronic kidney disease. *Nat Genet, 42*(5), 376-384.

Kujala, U. M., Kaprio, J., & Koskenvuo, M. (2002). Modifiable risk factors as predictors of all-cause mortality: the roles of genetics and childhood environment. *Am J Epidemiol, 156*(11), 985-993.

Kukita, T., Wada, N., Kukita, A., Kakimoto, T., Sandra, F., Toh, K., . . . Matsusaki, H. (2004). RANKL-induced DC-STAMP is essential for osteoclastogenesis. *The Journal of experimental medicine, 200*(7), 941-946.

Kullo, I. J., Ding, K., Jouni, H., Smith, C. Y., & Chute, C. G. (2010). A genome-wide association study of red blood cell traits using the electronic medical record. *PLoS One, 5*(9).

Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., . . . Zhernakova, A. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nature genetics, 53*(2), 156-165.

Kyu, H. H., Bachman, V. F., Alexander, L. T., Mumford, J. E., Afshin, A., Estep, K., . . . Forouzanfar, M. H. (2016). Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ, 354*.

Lake, R. I., Eaves, L. J., Maes, H. H., Heath, A. C., & Martin, N. G. (2000). Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behav Genet, 30*(3), 223-233.

Lange, K., Papp, J. C., Sinsheimer, J. S., Sripracha, R., Zhou, H., & Sobel, E. M. (2013). Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics, 29*(12), 1568-1570.

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. *Med Sci Sports Exerc, 29*(8), 1062-1068.

Laurier, E., Amiable, N., Gagnon, E., Brown, J. P., & Michou, L. (2017). Effect of a rare genetic variant of TM7SF4 gene on osteoclasts of patients with Paget's disease of bone. *BMC Med Genet, 18*(1), 133.

Lawal, I. A., & Bano, S. (2020). Deep Human Activity Recognition With Localisation of Wearable Sensors. *IEEE Access, 8*, 155060-155070.

Ledingham, I. (1977). Factors influencing oxygen availability. Journal of Clinical Pathology. Supplement (Royal College of Pathologists). 11, 1.

Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., & Katzmarzyk, P. T. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet, 380*(9838), 219-229.

Lewis, B. A., Napolitano, M. A., Buman, M. P., Williams, D. M., & Nigg, C. R. (2017). Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. *Journal of behavioral medicine, 40*(1), 112-126.

Liao, Y., Chou, C.-P., Huh, J., Leventhal, A., & Dunton, G. (2017). Examining acute bi-directional relationships between affect, physical feeling states, and physical activity in free-living situations using electronic ecological momentary assessment. *Journal of behavioral medicine, 40*(3), 445-457.

Liao, Y., Shonkoff, E. T., & Dunton, G. F. (2015). The acute relationships between affect, physical feeling states, and physical activity in daily life: a review of current evidence. *Frontiers in Psychology, 6*, 1975.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics, 27*(12), 1739-1740.

Lightfoot, J. T. (2020). Chicken and the Egg: Physical Activity and Epigenetics. *Medicine & Science in Sports & Exercise, 52*(3), 588.

Lightfoot, J. T., de Geus, E. J. C., Booth, F. W., Bray, M. S., Den Hoed, M., Kaprio, J., . . . Thomis, M. A. (2018). Biological/genetic regulation of physical activity level: consensus from GenBioPAC. *Medicine & Science in Sports & Exercise, 50*(4), 863-873.

Ligthart, L., van Beijsterveldt, C. E., Kevenaar, S. T., de Zeeuw, E., van Bergen, E., Bruins, S., . . . Hottenga, J.-J. (2019). The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and Human Genetics*, 1-14.

Lin, X., Chan, K. K., Huang, Y. T., Luo, X. I., Liang, L., Wilson, J., . . . Liu, S. (2018). Genetic Determinants for Leisure-Time Physical Activity. *Med Sci Sports Exerc, 50*(8), 1620-1628.

Liu, E. Y., Li, M., Wang, W., & Li, Y. (2013). MaCH-Admix: genotype imputation for admixed populations. *Genetic epidemiology, 37*(1), 25-37.

Liu, G., Zhu, H., Lagou, V., Gutin, B., Stallmann-Jorgensen, I. S., Treiber, F. A., . . . Snieder, H. (2010). FTO variant rs9939609 is associated with body mass index and waist circumference, but not with energy intake or physical activity in European- and African-American youth. *BMC Med Genet, 11*, 57.

Liu, L., Shi, Y., Li, T., Qin, Q., Yin, J., Pang, S., . . . Wei, S. (2016). Leisure time physical activity and cancer risk: evaluation of the WHO's recommendation

based on 126 high-quality epidemiological studies. *Br J Sports Med, 50*(6), 372-378.

Lo, M.-T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C.-C., Wang, Y., . . . Kauppi, K. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature genetics, 49*(1), 152-156.

Lochbaum, M. R., Rhodes, R. E., Stevenson, S. J., Surles, J., Stevens, T., & Wang, C. K. J. (2010). Does gender moderate the exercising personality? An examination of continuous and stage-based exercise. *Psychology, Health & Medicine, 15*(1), 50-60.

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., . . . Yang, J. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature, 518*(7538), 197-206.

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., . . . Abecasis, G. R. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics, 48*(11), 1443.

Loos, R., 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*(4), 420-428.

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*(3), 257-261.

Machiela, M. J., & Chanock, S. J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics, 31*(21), 3555-3557.

Mäestu, J., Lätt, E., Rääsk, T., Sak, K., Laas, K., Jürimäe, J., & Jürimäe, T. (2013). Ace I/D polymorphism is associated with habitual physical activity in pubertal boys. *J Physiol Sci, 63*(6), 427-434.

Maher, J. P., Rhodes, R. E., Dzubur, E., Huh, J., Intille, S., & Dunton, G. F. (2017). Momentary assessment of physical activity intention-behavior coupling in adults. *Translational behavioral medicine, 7*(4), 709-718.

Maia, J., Gomes, T. N., Trégouët, D.-A., & Katzmarzyk, P. T. (2014). Familial resemblance of physical activity levels in the Portuguese population. *Journal of science and medicine in sport, 17*(4), 381-386.

Mailing, L. J., Allen, J. M., Buford, T. W., Fields, C. J., & Woods, J. A. (2019). Exercise and the Gut Microbiome: A Review of the Evidence, Potential Mechanisms, and Implications for Human Health. *Exercise and Sport Sciences Reviews, 47*(2).

Mairbäurl, H. (2013). Red blood cells in sports: Effects of exercise and training on oxygen supply by red blood cells. *Frontiers in Physiology, 4*(332).

Malina, R. M. (2001). Tracking of physical activity across the lifespan. Research Digest President's Council on Physical Fitness and Sports. Washington D.C. Series 3, 14.

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics, 26*(22), 2867-2873.

Martinez-Gonzalez, M. A., Varo, J. J., Santos, J. L., Irala, J. d., Gibney, M., Kearney, J., & Martinez, J. A. (2001). Prevalence of physical activity during leisure time in the European Union.

Matson-Koffman, D. M., Brownstein, J. N., Neiner, J. A., & Greaney, M. L. (2005). A site-specific literature review of policy and environmental

interventions that promote physical activity and nutrition for cardiovascular health: what works? *American Journal of Health Promotion, 19*(3), 167-193.

Matthews, C. E., Chen, K. Y., Freedson, P. S., Buchowski, M. S., Beech, B. M., Pate, R. R., & Troiano, R. P. (2008). Amount of time spent in sedentary behaviors in the United States, 2003–2004. *Am J Epidemiol, 167*(7), 875-881.

Maurano, M. T., Wang, H., Kutyavin, T., & Stamatoyannopoulos, J. A. (2012). Widespread site-dependent buffering of human regulatory polymorphism. *PLoS Genet, 8*(3), e1002599.

McCaffery, J. M., Papandonatos, G. D., Bond, D. S., Lyons, M. J., & Wing, R. R. (2009). Gene× environment interaction of vigorous exercise and body mass index among male Vietnam-era twins. *The American journal of clinical nutrition, 89*(4), 1011-1018.

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., . . . the Haplotype Reference, C. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet, 48*(10), 1279-1283.

McCrae, R. R., & Costa Jr, P. T. (2008). The five-factor theory of personality.

McCrae, R. R., Costa Jr, P. T., Ostendorf, F., Angleitner, A., Hřebíčková, M., Avia, M. D., . . . Woodfield, R. (2000). Nature over nurture: temperament, personality, and life span development. *Journal of personality and social psychology, 78*(1), 173.

McMurray, R. G., Harrell, J. S., Bangdiwala, S. i., & Hu, J. (2003). Tracking of physical activity and aerobic power from childhood through adolescence. *Med Sci Sports Exerc, 35*(11), 1914-1922.

Migueles, J. H., Aadland, E., Andersen, L. B., Brønd, J. C., Chastin, S. F., Hansen, B. H., . . . Ortega, F. B. (2021). GRANADA consensus on analytical approaches to assess associations with accelerometer-determined physical behaviours (physical activity, sedentary behaviour and sleep) in epidemiological studies. *British journal of sports medicine*, bjsports-2020-103604.

Milojev, P., & Sibley, C. G. (2014). The stability of adult personality varies across age: Evidence from a two-year longitudinal sample of adult New Zealanders. *Journal of Research in Personality, 51*, 29-37.

Minică, C. C., Boomsma, D. I., Dolan, C. V., de Geus, E. J. C., & Neale, M. C. (2020). Empirical comparisons of multiple Mendelian randomization approaches in the presence of assortative mating. *Int J Epidemiol*.

Minică, C. C., Dolan, C. V., Boomsma, D. I., de Geus, E. J. C., & Neale, M. C. (2018). Extending causality tests with genetic instruments: An integration of Mendelian randomization with the classical twin design. *Behav Genet, 48*(4), 337-349.

Minică, C. C., Dolan, C. V., Kampert, M. M., Boomsma, D. I., & Vink, J. M. (2015). Sandwich corrected standard errors in family-based genome-wide association studies. *European Journal of Human Genetics, 23*(3), 388.

Ministry of Sports Wellbeing and Health (Producer). (2016, 14-7-2018). Dertig meest beoefende takken van sport naar achtergrondkenmerken, bevolking 4 jaar en ouder. Retrieved from https://www.volksgezondheidenzorg.info/bestanden/documenten/top30sporten4jrenouder20152016xlsx

Morgan, M. D., Pairo-Castineira, E., Rawlik, K., Canela-Xandri, O., Rees, J., Sims, D., . . . Jackson, I. J. (2018). Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nat Commun, 9*(1), 5271.

Morris, J. N., Everitt, M. G., Pollard, R., Chave, S. P., & Semmence, A. M. (1980). Vigorous exercise in leisure-time: protection against coronary heart disease. *Lancet, 2*(8206), 1207-1210.

Mullin, B. H., Zhu, K., Brown, S. J., Mullin, S., Tickner, J., Pavlos, N. J., . . . Wilson, S. G. (2019). Genetic regulatory mechanisms in human osteoclasts suggest a role for the STMP1 and DCSTAMP genes in Paget's disease of bone. *Scientific reports, 9*(1), 1052.

Murakami, H., Iemitsu, M., Sanada, K., Gando, Y., Ohmori, Y., Kawakami, R., . . . Miyachi, M. (2011). Associations among objectively measured physical activity, fasting plasma homocysteine concentration, and MTHFR C677T genotype. *European journal of applied physiology, 111*(12), 2997-3005.

Mustelin, L., Joutsi, J., Latvala, A., Pietiläinen, K. H., Rissanen, A., & Kaprio, J. (2012). Genetic influences on physical activity in young adults: a twin study. *Med Sci Sports Exerc, 44*(7), 1293-1301.

Mustelin, L., Latvala, A., Pietiläinen, K. H., Piirilä, P., Sovijärvi, A., Kujala, U. M., . . . Kaprio, J. (2011). Associations between sports participation, cardiorespiratory fitness, and adiposity in young adult twins. *Journal of Applied Physiology, 110*(3), 681-686.

Mustelin, L., Silventoinen, K., Pietilainen, K., Rissanen, A., & Kaprio, J. (2009). Physical activity reduces the influence of genetic effects on BMI and waist circumference: a study in young adult twins. *Int J Obes (Lond), 33*(1), 29-36.

Muthén, B., Asparouhov, T., & Rebollo, I. (2006). Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Research and Human Genetics, 9*(3), 313-324.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., . . . Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika, 81*(2), 535-549.

Neale, M. C., Walters, E. E., Eaves, L. J., Maes, H. H., & Kendler, K. S. (1994). Multivariate genetic analysis of twin-family data on fears: Mx models. *Behav Genet, 24*(2), 119-139.

Nederend, I., Schutte, N. M., Bartels, M., ten Harkel, A. D., & de Geus, E. J. C. (2016). Heritability of heart rate recovery and vagal rebound after exercise. *European journal of applied physiology, 116*(11-12), 2167-2176.

Nelson, M. C., Gordon-Larsen, P., North, K. E., & Adair, L. S. (2006). Body mass index gain, fast food, and physical activity: effects of shared environments over time. *Obesity, 14*(4), 701-709.

Pahkala, K., Hernelahti, M., Heinonen, O. J., Raittinen, P., Hakanen, M., Lagström, H., . . . Simell, O. (2013). Body mass index, fitness and physical activity from childhood through adolescence. *Br J Sports Med*, bjsports-2011-090704.

Pandey, A., Garg, S., Khunger, M., Darden, D., Ayers, C., Kumbhani, D. J., . . . Berry, J. D. (2015). Dose Response Relationship Between Physical Activity and Risk of Heart Failure: A Meta-Analysis. *Circulation*.

Parsons, T. J., Power, C., & Manor, O. (2006). Longitudinal physical activity and diet patterns in the 1958 British Birth Cohort. *Med Sci Sports Exerc, 38*(3), 547-554.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet, 2*(12), e190.

Payne, N., Jones, F., & Harris, P. R. (2005). The impact of job strain on the predictive validity of the theory of planned behaviour: An investigation of exercise and healthy eating. *Br J Health Psychol, 10*(1), 115-131.

Pellegrini, A. D., & Smith, P. K. (1998). Physical activity play: The nature and function of a neglected aspect of play. *Child development, 69*(3), 577-598.

Pérusse, 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*(3), 201-205.

Pérusse, L., Tremblay, A., Leblanc, C., & Bouchard, C. (1989). Genetic and environmental influences on level of habitual physical activity and exercise participation. *Am J Epidemiol, 129*(5), 1012-1022.

Physical Activity Guidelines for Americans. Retrieved from https://health.gov/paguidelines/

Picavet, H. S. J., Wendel-Vos, G. W., Vreeken, H. L., Schuit, A. J., & Verschuren, W. M. M. (2011). How stable are physical activity habits among adults? The Doetinchem Cohort Study. *Medicine & Science in Sports & Exercise, 43*(1), 74-79.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. (2013). *Behavioral genetics*. New York: Worth Publishers.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics, 38*(8), 904-909.

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Connor Gorber, S., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act, 5*, 56.

Purcell, S. M., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The american journal of human genetics, 81*(3), 559-575.

Python Software Foundation. (2014). Python 3. Retrieved from https://www.python.org/

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., . . . Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature, 464*(7285), 59-65.

R Core Team. (2013). R: A language and environment for statistical computing.

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2018). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research, 47*(D1), D886-D894.

Reynolds, C. A., Baker, L. A., & Pedersen, N. L. (2000). Multivariate models of mixed assortment: Phenotypic assortment and social homogamy for education and fluid ability. *Behav Genet, 30*(6), 455-476.

Reynolds, C. A., Barlow, T., & Pedersen, N. L. (2006). Alcohol, tobacco and caffeine use: spouse similarity processes. *Behav Genet, 36*(2), 201.

Rhodes, R. E., Courneya, K. S., & Jones, L. W. (2003). Translating exercise intentions into behavior: Personality and social cognitive correlates. *Journal of Health Psychology, 8*(4), 447-458.

Rhodes, R. E., & de Bruijn, G. J. (2013). How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. *Br J Health Psychol, 18*(2), 296-309.

Rhodes, R. E., & Smith, N. E. I. (2006). Personality correlates of physical activity: a review and meta-analysis. *British journal of sports medicine, 40*(12), 958-965.

Richards, R., Williams, S., Poulton, R., & Reeder, A. I. (2007). Tracking club sport participation from childhood to early adulthood. *Research quarterly for exercise and sport, 78*(5), 413-419.

Richert, L., Chevalley, T., Manen, D., Bonjour, J.-P., Rizzoli, R., & Ferrari, S. (2007). Bone Mass in Prepubertal Boys Is Associated with a Gln223Arg Amino Acid Substitution in the Leptin Receptor. *The Journal of Clinical Endocrinology & Metabolism, 92*(11), 4380-4386.

Rico-Sanz, J., Rankinen, T., Rice, T., Leon, A., Skinner, J., Wilmore, J., . . . Bouchard, C. (2004). Quantitative trait loci for maximal exercise capacity phenotypes and their responses to training in the HERITAGE Family Study. *Physiological genomics, 16*(2), 256-260.

Ridley, K., Ainsworth, B. E., & Olds, T. S. (2008). Development of a compendium of energy expenditures for youth. *International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 45.

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: a quantitative review of longitudinal studies. *Psychological bulletin, 126*(1), 3.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., . . . Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature, 555*(7695), 210-215.

Rowland, T. (2016). Biologic regulation of physical activity: Human Kinetics.

Ryan, R. M., & Deci, E. L. (2007). Active human nature: Self-determination theory and the promotion and maintenance of sport, exercise, and health. *Intrinsic motivation and self-determination in exercise and sport, 1*, 19.

Sallis, J., Zacarian, J., Howell, M., & Hofstetter, R. (1996). Ethnic, socioeconomic, and sex differences in physical activity among adolescents. *J Clin Epidem, 49*.

Samitz, G., Egger, M., & Zwahlen, M. (2011). Domains of physical activity and all-cause mortality: systematic review and dose-response meta-analysis of cohort studies. *Int J Epidemiol, 40*(5), 1382-1400.

Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. *Journal of psychiatric research, 77*, 42-51.

Schumann, G., Coin, L. J., Lourdusamy, A., Charoen, P., Berger, K. H., Stacey, D., . . . Amin, N. (2011). Genome-wide association and genetic functional studies identify autism susceptibility candidate 2 gene (AUTS2) in the regulation of alcohol consumption. *Proceedings of the National Academy of Sciences, 108*(17), 7119-7124.

Schutte, N. M., Bartels, M., & de Geus, E. J. C. (2018). Genetics of physical activity and physical fitness. In N. Armstrong & W. van Mechelen (Eds.), *Oxford Textbook of Children's Sport and Exercise Medicine* (Vol. 3, pp. 293-302). Oxford: Oxford University Press.

Schutte, N. M., Nederend, I., Bartels, M., & de Geus, E. J. C. (2019). A twin study on the correlates of voluntary exercise behavior in adolescence. *Psychology of Sport and Exercise, 40*, 99-109.

Schutte, N. M., Nederend, I., Hudziak, J. J., Bartels, M., & de Geus, E. J. C. (2016). Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiological genomics, 48*(3), 210-219.

Schutte, N. M., Nederend, I., Hudziak, J. J., Bartels, M., & de Geus, E. J. C. (2017). Heritability of the affective response to exercise and its correlation to

exercise behavior. *Psychology of Sport and Exercise, 31*(Supplement C), 139-148.

Schutte, N. M., Nederend, I., Hudziak, J. J., de Geus, E. J. C., & Bartels, M. (2016). Differences in Adolescent Physical Fitness: A Multivariate Approach and Meta-analysis. *Behav Genet, 46*(2), 217-227.

Seabra, A. F., Mendonça, D. M., Göring, H. H., Thomis, M. A., & Maia, J. A. (2008). Genetic and environmental factors in familial clustering in physical activity. *European journal of epidemiology, 23*(3), 205-211.

Seabra, A. F., Mendonça, D. M., Göring, H. H., Thomis, M. A., & Maia, J. A. (2014). Genetic influences of sports participation in Portuguese families. *European journal of sport science, 14*(5), 510-517.

Seefeldt, V., Malina, R. M., & Clark, M. A. (2002). Factors affecting levels of physical activity in adults. *Sports medicine, 32*(3), 143-168.

Sherwood, N. E., & Jeffery, R. W. (2000). The behavioral determinants of exercise: implications for physical activity interventions. *Annual review of nutrition, 20*(1), 21-44.

Silfee, V. J., Haughton, C. F., Jake-Schoffman, D. E., Lopez-Cepero, A., May, C. N., Sreedhara, M., . . . Lemon, S. C. (2018). Objective measurement of physical activity outcomes in lifestyle interventions among adults: A systematic review. *Preventive medicine reports, 11*, 74-80.

Simonen, R. L., Levälahti, E., Kaprio, J., Videman, T., & Battie, M. C. (2004). Multivariate genetic analysis of lifetime exercise and environmental factors. *Med Sci Sports Exerc, 36*(9), 1559-1566.

Simonen, R. L., Rankinen, T., Pérusse, L., Leon, A. S., Skinner, J. S., Wilmore, J. H., . . . Bouchard, C. (2003). A dopamine D2 receptor gene polymorphism and physical activity in two family studies. *Physiology & behavior, 78*(4), 751-757.

Simonen, R. L., Rankinen, T., Perusse, L., Rice, T., Rao, D., Chagnon, Y., & Bouchard, C. (2003). Genome-wide linkage scan for physical activity levels in the Quebec Family study. *Med Sci Sports Exerc, 35*(8), 1355-1359.

Simonen, R. L., Videman, T., Kaprio, J., Levälahti, E., & Battie, M. C. (2003). Factors associated with exercise lifestyle-A study of monozygotic twins. *International journal of sports medicine, 24*(07), 499-505.

Siva, N. (2008). 1000 Genomes project. In: Nature Publishing Group.

Slunecka, J. L., van der Zee, M. D., Beck, J. J., Johnson, B. N., Finnicum, C. T., Pool, R., . . . Ehli, E. A. (2021). Implementation and Implications for Polygenic Risk Scores in Healthcare. *Manuscript under review.*

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics, 14*(7), 483-495.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: the impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of personality and social psychology, 101*(4), 862.

Spinath, F. M., Wolf, H., Angleitner, A., Borkenau, P., & Riemann, R. (2002). Genetic and environmental influences on objectively assessed activity in adults. *Personality and Individual Differences, 33*(4), 633-645.

Srivastava, S., Mishra, R. K., & Dhawan, J. (2010). Regulation of cellular chromatin state: insights from quiescence and differentiation. *Organogenesis, 6*(1), 37-47.

Stamler, J. S., Jia, L., Eu, J. P., McMahon, T. J., Demchenko, I. T., Bonaventura, J., . . . Piantadosi, C. A. (1997). Blood flow regulation by S-

nitrosohemoglobin in the physiological oxygen gradient. *Science, 276*(5321), 2034-2037.

Stanton, R., & Reaburn, P. (2014). Exercise and the treatment of depression: A review of the exercise program variables. *Journal of science and medicine in sport, 17*(2), 177-182.

Staudenmayer, J., Pober, D., Crouter, S., Bassett, D., & Freedson, P. (2009). An artificial neural network to estimate physical activity energy expenditure and identify physical activity type from an accelerometer. *Journal of Applied Physiology*.

Steca, P., Baretta, D., Greco, A., D'Addario, M., & Monzani, D. (2018). Associations between personality, sports participation and athletic success. A comparison of Big Five in sporting and non-sporting adults. *Personality and Individual Differences, 121*, 176-183.

Stefan, N., Vozarova, B., Del Parigi, A., Ossowski, V., Thompson, D. B., Hanson, R. L., . . . Tataranni, P. A. (2002). The Gln223Arg polymorphism of the leptin receptor in Pima Indians: influence on energy expenditure, physical activity and lipid metabolism. *Int J Obes Relat Metab Disord, 26*(12), 1629-1632.

Stephan, Y., Sutin, A. R., & Terracciano, A. (2014). Physical activity and personality development across adulthood and old age: Evidence from two longitudinal studies. *Journal of Research in Personality, 49*, 1-7.

Steptoe, A., Wardle, J., Cui, W., Baban, A., Glass, K., Pelzer, K., . . . Vinck, J. (2002). An international comparison of tobacco smoking, beliefs and risk awareness in university students from 23 countries. *Addiction, 97*(12), 1561-1571.

Steptoe, A., Wardle, J., Fuller, R., Holte, A., Justo, J., Sanderman, R., & Wichstrøm, L. (1997). Leisure-time physical exercise: prevalence, attitudinal correlates, and behavioral correlates among young Europeans from 21 countries. *Prev Med, 26*(6), 845-854.

Stubbe, J. H., Boomsma, D. I., & de Geus, E. J. C. (2005). Sports participation during adolescence: a shift from environmental to genetic factors. *Medicine & Science in Sports & Exercise, 37*(4), 563-570.

Stubbe, J. H., Boomsma, D. I., Vink, J. M., Cornes, B. K., Martin, N. G., Skytthe, A., . . . Kaprio, J. (2006). Genetic influences on exercise participation in 37.051 twin pairs from seven countries. *PLoS One, 1*(1), e22.

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. *Prev Med, 44*(2), 148-152.

Stubbs, B., Vancampfort, D., Hallgren, M., Firth, J., Veronese, N., Solmi, M., . . . Kahl, K. G. (2018). EPA guidance on physical activity as a treatment for severe mental illness: a meta-review of the evidence and Position Statement from the European Psychiatric Association (EPA), supported by the International Organization of Physical Therapists in Mental Health (IOPTMH). *European Psychiatry, 54*, 124-144.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545-15550.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine, 12*(3), e1001779.

Sun, B. B., Maranville, J. C., Peters, J. E., Stacey, D., Staley, J. R., Blackshaw, J., . . . Butterworth, A. S. (2018). Genomic atlas of the human plasma proteome. *Nature, 558*(7708), 73-79.

Sutin, A. R., Stephan, Y., Luchetti, M., Artese, A., Oshio, A., & Terracciano, A. (2016). The five-factor model of personality and physical inactivity: A meta-analysis of 16 samples. *Journal of Research in Personality, 63*, 22-28.

Tachmazidou, I., Süveges, D., Min, J. L., Ritchie, G. R. S., Steinberg, J., Walter, K., . . . Zeggini, E. (2017). Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *Am J Hum Genet, 100*(6), 865-884.

Tammelin, R., Yang, X., Leskinen, E., Kankaanpaa, A., Hirvensalo, M., Tammelin, T., & Raitakari, O. (2014). Tracking of physical activity from early childhood through youth into adulthood. *Med. Sci. Sports Exerc, 46*, 955-962.

Tammelin, T., Näyhä, S., Hills, A. P., & Järvelin, M.-R. (2003). Adolescent participation in sports and adult physical activity. *American Journal of Preventive Medicine, 24*(1), 22-28.

Telama, R. (2009). Tracking of physical activity from childhood to adulthood: a review. *Obesity facts, 2*(3), 187-195.

Telama, R., Leskinen, E., & Yang, X. (1996). Stability of habitual physical activity and sport participation: a longitudinal tracking study. *Scandinavian journal of medicine & science in sports, 6*(6), 371-378.

Telama, R., Yang, X., Hirvenssalo, M., & Raitakari, O. (2006). Participation in organized youth sport as a predictor of adult physical activity: A 21-year longitudinal Study. *Pediatr Exerc Sci, 18*(1), 76-88.

Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., & Raitakari, O. (2005). Physical activity from childhood to adulthood – a 21–year tracking study. *American Journal of Preventive Medicine, 28*(3), 267-273.

Terracciano, A., McCrae, R. R., & Costa Jr, P. T. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality, 44*(1), 31-37.

Thompson, P. D., Tsongalis, G. J., Seip, R. L., Bilbie, C., Miles, M., Zoeller, R., . . . Pescatello, L. (2004). Apolipoprotein E genotype and changes in serum lipids and maximal oxygen uptake with exercise training. *Metabolism, 53*(2), 193-202.

Tin, A., Colantuoni, E., Boerwinkle, E., Kottgen, A., Franceschini, N., Astor, B. C., . . . Kao, W. H. (2013). Using multiple measures for quantitative trait association analyses: application to estimated glomerular filtration rate. *J Hum Genet, 58*(7), 461-466.

Troiano, R. P., Berrigan, D., Dodd, K. W., Masse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc, 40*(1), 181.

Tudor-Locke, C., Giles-Corti, B., Knuiman, M., & McCormack, G. (2008). Tracking of pedometer-determined physical activity in adults who relocate: results from RESIDE. *International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 39.

Twisk, J. W. R., Kemper, H. C. G., & van Mechelen, W. (2002). The Relationship Between Physical Fitness and Physical Activity During Adolescence and Cardiovascular Disease Risk Factors at Adult Age. The Amsterdam Growth and Health Longitudinal Study. *Int J Sports Med, 23*(S1), 8-14.

van Beijsterveldt, C. E., Groen-Blokhuis, M., Hottenga, J. J., Franić, S., Hudziak, J. J., Lamb, D., . . . Schutte, N. M. (2013). The Young Netherlands Twin

Register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics, 16*(1), 252-267.

van den Berg, S. M., de Moor, M. H. M., McGue, M., Pettersson, E., Terracciano, A., Verweij, K. J., . . . van Grootheest, G. (2014). Harmonization of Neuroticism and Extraversion phenotypes across inventories and cohorts in the Genetics of Personality Consortium: an application of Item Response Theory. *Behav Genet, 44*(4), 295-313.

van der Aa, N., de Geus, E. J. C., van Beijsterveldt, T. C., Boomsma, D. I., & Bartels, M. (2010). Genetic influences on individual differences in exercise behavior during adolescence. *International Journal of Pediatrics, 2010*.

van der Mee, D. J., Fedko, I. O., Hottenga, J. J., Ehli, E. A., van der Zee, M. D., Ligthart, L., . . . de Geus, E. J. C. (2017). Dopaminergic Genetic Variants and Voluntary Externally Paced Exercise Behavior. *Med Sci Sports Exerc, 50*(5), 700-708.

van der Zee, M. D. (2021). *Summary statistics for GWAS of voluntary exercise behaviors in the Netherlands*. Retrieved from: https://doi.org/10.34894/ZLSPFG

van der Zee, M. D., & de Geus, E. J. C. (2019). Is Physical Activity Regulated by Genetics? Evidence From Studies in Humans. In J. T. Lightfoot, M. J. Hubal, & S. M. Roth (Eds.), *Routledge Handbook of Sport and Exercise Systems Genetics*: Routledge.

van der Zee, M. D., Ehli, E. A., Pool, R., Hottenga, J. J., van der Mee, D. J., Ligthart, L., . . . de Geus, E. J. C. (2021). Genome-wide association study of voluntary exercise behaviors in the Netherlands. *Manuscript submitted for publication.*

van der Zee, M. D., Helmer, Q., Boomsma, D. I., Dolan, C. V., & de Geus, E. J. C. (2020). An Extended Twin-Pedigree Study of Different Classes of Voluntary Exercise Behavior. *Behav Genet, 50*(2), 94-104.

van der Zee, M. D., Schutte, N. M., de Moor, M. H. M., & de Geus, E. J. C. (2019). Genetic influences on regular exercise behavior. In *Lifestyle Medicine, 3rd Edition* (3 ed.).

van der Zee, M. D., van der Mee, D. J., Bartels, M., & de Geus, E. J. C. (2019). Tracking of voluntary exercise behaviour over the lifespan. *International Journal of Behavioral Nutrition and Physical Activity, 16*(1), 17.

van Grootheest, D., van den Berg, S. M., Cath, D., Willemsen, G., & Boomsma, D. (2008). Marital resemblance for obsessive–compulsive, anxious and depressive symptoms in a population-based sample. *Psychol Med, 38*(12), 1731-1740.

van Roekel, E. H., Dugué, P.-A., Jung, C.-H., Joo, J. E., Makalic, E., Wong, E. M., . . . Lynch, B. M. (2019). Physical activity, television viewing time, and DNA methylation in peripheral blood. *Med Sci Sports Exerc, 51*(3), 490-498.

van Rossum, G., & Drake Jr, F. L. (1995). *Python reference manual*: Centrum voor Wiskunde en Informatica Amsterdam.

van Rossum, G., & Drake Jr, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Varo, J. J., Martínez-González, M. A., de Irala-Estévez, J., Kearney, J., Gibney, M., & Martínez, J. A. (2003). Distribution and determinants of sedentary lifestyles in the European Union. *Int J Epidemiol, 32*(1), 138-146.

Vernon, P. A., Martin, R. A., Schermer, J. A., & Mackie, A. (2008). A behavioral genetic investigation of humor styles and their correlations with the Big-5 personality dimensions. *Personality and Individual Differences, 44*(5), 1116-1125.

Vilhjálmsson, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., . . . Do, R. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The american journal of human genetics, 97*(4), 576-592.

Vink, J. M., Boomsma, D. I., Medland, S. E., de Moor, M. H. M., Stubbe, J. H., Cornes, B. K., . . . Rose, R. J. (2011). Variance components models for physical activity with age as modifier: a comparative twin study in seven countries. *Twin Research and Human Genetics, 14*(1), 25-34.

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *The american journal of human genetics, 90*(1), 7-24.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: biology, function, and translation. *The american journal of human genetics, 101*(1), 5-22.

Visscher, P. M., & Yang, J. (2016). A plethora of pleiotropy across complex traits. *Nature genetics, 48*(7), 707-708.

Voisin, S., Eynon, N., Yan, X., & Bishop, D. (2015). Exercise training and DNA methylation in humans. *Acta Physiologica, 213*(1), 39-59.

Waller, K., Vähä-Ypyä, H., Törmäkangas, T., Hautasaari, P., Lindgren, N., Iso-Markku, P., . . . Sievänen, H. (2018). Long-term leisure-time physical activity and other health habits as predictors of objectively monitored late-life physical activity–A 40-year twin study. *Scientific reports, 8*(1), 9400.

Watanabe, K., Taskesen, E., van Bochoven, A., & Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. *Nature communications, 8*(1), 1826.

Watanabe, K., Umićević Mirkov, M., de Leeuw, C. A., van den Heuvel, M. P., & Posthuma, D. (2019). Genetic mapping of cell type specificity for complex traits. *Nature communications, 10*(1), 3222.

Weggemans, R. M., Backx, F. J. G., Borghouts, L., Chinapaw, M., Hopman, M. T. E., Koster, A., . . . Committee Dutch Physical Activity, G. (2018). The 2017 Dutch Physical Activity Guidelines. *International Journal of Behavioral Nutrition and Physical Activity, 15*(1), 58.

Weiss, K. M. (1993). Genetic variation and human disease: principles and evolutionary approaches (Vol. 11): Cambridge University Press.

Westerterp, K. R. (2013). Physical activity and physical activity induced energy expenditure in humans: measurement, determinants, and effects. *Frontiers in Physiology, 4*, 90.

Wheeler, E., Leong, A., Liu, C. T., Hivert, M. F., Strawbridge, R. J., Podmore, C., . . . Meigs, J. B. (2017). Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med, 14*(9), e1002383.

Wilde, G. J. S. (1970). Neurotische labiliteit gemeten met de vragenlijstmethode (2nd edition) Amsterdam: van Rossen.

Wilkinson, A. V., Gabriel, K. P., Wang, J., Bondy, M. L., Dong, Q., Wu, X., . . . Spitz, M. R. (2013). Sensation-seeking genes and physical activity in youth. *Genes Brain Behav, 12*(2), 181-188.

Willemsen, G., Vink, J. M., Abdellaoui, A., den Braber, A., van Beek, J. H., Draisma, H. H., . . . van Lien, R. (2013). The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Research and Human Genetics, 16*(1), 271-281.

Wilson, K. E., & Dishman, R. K. (2015). Personality and physical activity: A systematic review and meta-analysis. *Personality and Individual Differences, 72*, 230-242.

Winnicki, M., Accurso, V., Hoffmann, M., Pawlowski, R., Dorigatti, F., Santonastaso, M., . . . Pessina, A. C. (2004). Physical activity and angiotensin-converting enzyme gene polymorphism in mild hypertensives. *American Journal of Medical Genetics Part A, 125*(1), 38-44.

Wojczynski, M. K., Li, M., Bielak, L. F., Kerr, K. F., Reiner, A. P., Wong, N. D., . . . Reilly, M. P. (2013). Genetics of coronary artery calcification among African Americans, a meta-analysis. *BMC Med Genet, 14*, 75.

World Anti-Doping Agency. (2021). The prohibited list. https://www.wada-ama.org/sites/default/files/resources/files/2021list_en.pdf

World Health Organization. (2013). Global Recommendations on Physical Activity for Health. Geneva; 2010. In: ISBN 978-92-4-159-997-9.

World Health Organization. (2014). *Global status report on noncommunicable diseases 2014*: World Health Organization.

Yagi, M., Miyamoto, T., Sawatani, Y., Iwamoto, K., Hosogane, N., Fujita, N., . . . Miyamoto, K. (2005). DC-STAMP is essential for cell–cell fusion in osteoclasts and foreign body giant cells. *The Journal of experimental medicine, 202*(3), 345-351.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics, 88*(1), 76-82.

Yen, C.-T., & Lin, J.-D. (2020). Human body activity recognition using wearable inertial sensors integrated with a feature extraction–based machine-learning classification algorithm. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 0954405420937894.

Zhernakova, D. V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., . . . Franke, L. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. *Nature genetics, 49*(1), 139-145.

Zhu, X., Yin, S., Lang, M., He, R., & Li, J. (2016). The more the better? A meta-analysis on effects of combined cognitive and physical intervention on cognition in healthy older adults. *Ageing Research Reviews, 31*, 67-79.

Other Publications

van der Mee, D. J., I. Fedko, I.O, Hottenga, J.J., Ehli, E.A., **van der Zee, M.D.**, Ligthart, L., van Beijsterveldt, T., Davies, G.E., Bartels, M., Landers J.G. & de Geus E.J.C. (2018). Dopaminergic Genetic Variants and Voluntary Externally Paced Exercise Behavior. *Medicine and Science in Sports and Exercise,* 50(4):700-708.

Mbarek, H., van de Weijer, M.P., **van der Zee, M.D.**, Ip, H.F., Beck, J.J., Abdellaoui, A., Ehli, E.A., Davies, G.E., Baselmans, B.M.L., Nivard, M.G., Bartels, M., de Geus E.J. & Boomsma, D.I. (2019). Biological insights into multiple birth: genetic findings from UK Biobank. *European Journal of Human Genetics,* 27(6):970-979.

Pasman, J.A., Verweij, K.J.H., Gerring, Z., Stringer, S., Sanchez-Roige, S., Treur, J.L., Abdellaoui, A., Nivard, M.G., Baselmans, B.M.L., Ong, J.S., Ip, H.F., **van der Zee, M.D.**, Bartels, M., ... Posthuma, D., Munafò, M.R., Gillespie, N.A., Derks, E.M. & Vink, J.M. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature Neuroscience*. 2018 Sep;21(9):1161-1170.

Abdellaoui, A., Sanchez-Roige, S., Sealock, J., Treur, J.L., Dennis, J., Fontanillas, P., Elson, S., The 23andme Research Team, Nivard, M.G., Ip, H.F., **van der Zee, M.D.**, ... & Boomsma, D.I. (2019). Phenome-wide investigation of health outcomes associated with genetic predisposition to loneliness. *Human molecular genetic*s, 28(22), 3853-3865.

Pirastu, N., Cordioli, M., Nandakumar, P., Mignogna, G., Abdellaoui, A., Hollis, B., ..., **van der Zee M.D.**, ... & iPSYCH Consortium. (2021). Genetic analyses identify widespread sex-differential participation bias. *Nature Genetics, 53,* 663-671.

Arends, R. M., Pasman, J. A., Verweij, K. J., Derks, E. M., Gordon, S. D., Hickie, I., Thomas, N.S., Aliev F., Zietsch B.P., **van der Zee M.D.**, ... & Vink, J. M. (2021). Associations between the CADM2 gene, substance use, risky sexual behavior, and self-control: A phenome-wide association study. *Addiction biology*, e13015.

Beck, J.J., Pool, R., van de Weijer, M., Chen, X., Krapohl, E., Gordon, S.D., Nygaard, M., Debrabant, B., Palvianen, T, **van der Zee, M.D.**, ... & Hottenga, J. J. (2021). Genetic Meta-Analysis of Twin Birth Weight Shows High Genetic Correlation with Singleton Birth Weight. *Human Molecular Genetics, 2021,* ddab121.

McCartney, D. L., Min, J. L., Richmond, R. C., Lu, A. T., Sobczyk, M. K., Davies, G., ..., **van der Zee, M.D.**, ... & DNA Methylation Consortium. (2021). Genome-wide association studies identify 137 loci for DNA methylation biomarkers of ageing. *Genome Biology, 22,* 194.

Howe, L.J., Nivard. M.G., Morris, T.T., Hansen, A.F., Rasheed, H., Cho, Y., Chittoor, G., Lind, P.A., Palviainen, T., **van der Zee, M.D.**, ... Brumpton, B., Hemani, G. & Davies, N.M. (2021, *under review*). Within-sibship GWAS improve estimates of direct genetic effects. *Nature Genetics*

Slunecka, J.L., **van der Zee, M.D.**, Beck, J.J., Johnson, B.N., Finnicum, C.T., Pool, R., Hottenga, J.J., de Geus, E.J.C. & Ehli, E. (2021). Implementation and Implications for Polygenic Risk Scores in Healthcare. *Human Genomics, 15,* 46.

Huider, F., Milaneschi, Y., **van der Zee, M.D.**, de Geus, E.J.C., Helmer, Q., Penninx, B.W.J.H. & Boomsma D.I. (2021). Major Depressive Disorder and Lifestyle: Correlated Genetic Effects in Extended Twin Pedigrees.

Baselmans, B., Hammerschlag, A.R., Noordijk, S., Ip, H.F., **van der Zee, M.D.**, de Geus, E.J.C., Abdellaoui, A., Treur, J.L. & van 't Ent, D. (2021, *in review*). The Genetic and Neural Substrates of Externalizing Behavior.

van de Weijer, M.P., Pelt, D.H.M. de Vries, L.P., Huider, F., **van der Zee, M.D.**, Lighthart, L., Willemsen, G., Boomsma, D.I., de Geus, E.J.C., Bartels, M. (2021, *submitted*). Genetic and Environmental Influences on Quality of Life: the COVID-19 Pandemic as a Natural Experiment.

Klimentidis, Y.C., Newell, M., **van der Zee, M.D.**, Bland, V.L., May-Wilson, S., Menni, C., Mangino, M., Arora, A., Raichlen, D.A., Alexander, G., Wilson, J.F., Hottenga, J.J. de Geus, E.J.C. & Pirastu, N. (2021, *submitted*). Genome-wide association study of liking of physical activity in the UK Biobank.

Supplementary

Note: Some supplementary tables and figures do not fit standard paper size. These supplementary materials are available online at:

https://drive.google.com/drive/folders/1i8IUQJcMmzj0QTPtGTxs-FuTn8tpKIKZr?usp=sharing

Supplementary methods

Supplementary Methods Chapter 7. Genotyping and QC in the NTR

Genotyping in the NTR was done on multiple platforms, following respective manufacturer protocols: Perlegen Affymetrix, Affymetrix 6.0, Affymetrix Axiom, Illumina Human Quad Bead 660, Illumina Omni 1M and Illumina GSA. For each genotype platform, samples were removed if DNA-based sex did not match self-reported sex, if the heterozygosity F statistic as estimated by was below -0.10 or above 0.10, or if the genotyping call rate was below 0.90. SNPs were excluded if the minor allele frequency (MAF) was below 10^{-6}, if the Hardy-Weinberg equilibrium (HWE) p-value was below 10^{-6}, and/or if the call rate was below 0.95. These statistics were calculated using PLINK (Chang et al., 2015). The genotype data was then aligned with the 1000 Genomes (Siva, 2008) reference panel using the HRC (Loh et al., 2016) and 1000 Genomes checking tool, testing and filtering for SNPs with allele frequency differences larger than 0.20 as compared to the CEU population, palindromic SNPs and DNA strand issues. The data of the different platforms was then merged into a single dataset, and one platform was chosen for each individual. Based on the ~10.8k SNPs that all platforms have in common, DNA identity-by-descent state was estimated for all individual pairs using PLINK (Chang et al., 2015) and King (Manichaikul et al., 2010) software. Samples were excluded if these estimates did not correspond to expected familial relationships. CEU population outliers, based on 1000 Genomes PC projection per platform with the SmartPCA software (Patterson et al., 2006; Price et al., 2006), were removed from the data. Then, per platform, the data was phased using Eagle (Loh et al., 2016) and then imputed to 1000 Genomes and Topmed reference databases using Minimac (Fuchsberger et al., 2015; Howie et al., 2012) following the Michigan imputation server protocols (Das et al., 2016). Post-imputation, the resulting per-platform VCF files were merged with Bcftools (Danecek

et al., 2021) into a single file per chromosome for each reference, for SNPs present on all platforms. For the polygenic scoring the imputed data were converted to best guess data and were filtered to include only ACGT SNPs, SNPs with MAF above 0.01, HWE p-value above 10^{-5} and a genotype call rate above 0.98, and to exclude non-biallelic SNPs. All mendelian errors were set to missing. Twenty PCs were calculated with Smartpca using LD-pruned 1000 Genomes–imputed SNPs genotyped on at least one platform, having MAF above 0.05 and not present in the long-range LD regions.

Supplementary tables

	Total		Team		Sol		Comp		Ncomp		EP		IP	
Age	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
8	527	517	170	374	357	409	232	423	294	366	230	421	194	336
10	766	667	383	596	382	464	487	628	279	386	486	636	124	304
12	872	798	427	704	445	556	556	742	316	486	549	741	157	396
14	984	908	528	820	456	606	663	865	321	514	659	870	160	394
16	1057	1047	540	857	518	764	665	912	392	676	658	906	225	529
18	1046	1186	479	891	567	915	584	956	462	825	578	960	317	702
20	912	1108	373	808	539	830	444	866	468	757	444	890	345	635
22	907	1066	314	737	594	853	391	807	516	800	391	826	413	701
24	921	1116	287	702	633	941	360	768	561	877	359	779	475	812
26	834	962	206	549	628	847	276	608	558	804	272	605	488	762
28	762	903	171	506	591	785	251	607	511	713	248	606	453	678
30	719	932	163	510	556	785	225	587	494	725	221	574	434	706
32	701	900	124	415	577	798	189	486	512	751	191	502	451	698
34	619	780	99	352	520	717	171	458	448	665	169	454	403	631
36	551	728	63	275	488	677	123	364	428	643	125	380	374	583
38	552	734	63	265	489	695	128	370	424	657	128	369	378	628
40	552	716	57	253	495	670	131	363	422	633	130	362	377	603
42	549	693	56	238	493	659	145	380	403	597	143	376	364	570
44	543	712	51	239	492	674	147	393	395	602	145	389	355	582
46	544	696	46	196	497	672	149	383	395	600	147	380	358	577
48	562	743	41	183	520	725	145	381	417	652	143	379	380	634
50	576	761	37	173	539	743	134	374	441	672	132	372	402	639
52	586	833	31	171	555	810	123	376	463	744	119	359	420	705
54	602	899	34	170	569	871	133	382	469	806	131	380	428	778
56	598	877	35	213	562	833	138	403	460	761	133	398	427	731
58	559	784	20	134	539	770	118	359	441	674	114	353	403	650
60	556	825	18	131	537	799	121	382	434	709	116	372	395	672
62	560	911	21	151	539	881	116	352	444	830	110	347	409	795
64	597	837	13	114	584	824	132	427	465	704	125	414	435	680
66	543	868	13	100	530	864	118	366	425	779	112	357	403	767
68	448	628	10	82	439	623	101	327	347	553	95	317	313	522
70	469	756	15	92	454	753	123	402	345	656	114	388	319	637
72	399	576	12	88	387	557	119	378	280	465	94	338	283	475
74	433	740	6	55	428	734	114	374	319	582	101	365	311	583
76	319	549	18	161	301	488	106	390	213	383	90	354	198	350
78	336	643	0	0	336	643	48	287	288	597	42	283	279	582
80	267	516	4	37	263	514	56	234	211	485	43	221	224	487
82	190	315	0	0	190	315	7	50	183	307	0	0	181	313

Supplementary table 2.1. *Mean, and standard deviation (SD) of MET-minutes in the 7 exercise dimensions over age. Comp: Competitive; Ncomp: non-competitive; EP: externally paced; IP: internally paced.*

	2-year		8-year		14-year		20-year	
	Range	Median	Range	Median	Range	Median	Range	Median
Within domain								
Total								
Males	0.4-0.77	0.56	0.21-0.78	0.42	0.13-0.72	0.36	0.09-0.59	0.36
Females	0.37-0.71	0.53	0.23-0.64	0.39	0.2-0.65	0.32	0.26-0.5	0.35
Team								
Males	0.38-0.81	0.63	0.16-0.67	0.46	0.24-0.58	0.42	0.06-0.41	0.28
Females	0.52-0.98	0.7	0.14-0.79	0.48	0.15-1	0.65	0.03-0.72	0.23
Competitive								
Males	0.48-0.84	0.63	0.28-0.74	0.47	0.26-0.61	0.45	0.19-0.42	0.26
Females	0.53-0.83	0.69	0.36-0.7	0.49	0.26-0.91	0.45	0.21-0.66	0.37
Externally paced								
Males	0.47-0.88	0.64	0.29-0.72	0.51	0.28-0.6	0.45	0.21-0.43	0.26
Females	0.52-0.83	0.67	0.35-0.7	0.52	0.25-0.8	0.42	0.21-0.67	0.36
Solitary								
Males	0.24-0.77	0.54	0.11-0.8	0.4	0.08-0.82	0.42	0.18-0.75	0.4
Females	0.3-0.72	0.51	0.2-0.66	0.35	0.11-0.68	0.26	0.07-0.53	0.26
Non-competitive								
Males	0.26-0.78	0.56	0.06-0.93	0.37	0.05-0.92	0.36	0.07-0.81	0.43
Females	0.28-0.7	0.46	0.17-0.69	0.32	0.07-0.57	0.24	0.04-0.5	0.16
Internally Paced								
Males	0.22-0.75	0.55	0.12-0.87	0.39	0.04-0.84	0.34	-0.09-0.8	0.43
Females	0.26-0.77	0.44	0.15-0.73	0.33	0.06-0.42	0.23	0.01-0.48	0.1
Between domains								
Team								
Males	0.12-0.52	0.25	0.06-0.33	0.22	0.06-0.35	0.16	0.03-0.22	0.12
Females	0.01-0.5	0.24	0.03-0.38	0.22	0.03-0.35	0.17	0.03-0.4	0.19
Competitive								
Males	0.24-0.51	0.36	0.14-0.4	0.28	0.13-0.38	0.22	0.04-0.32	0.21
Females	0.27-0.49	0.33	0.18-0.42	0.29	0.11-0.4	0.2	0.13-0.35	0.24
Externally paced								
Males	0.23-0.52	0.37	0.14-0.39	0.28	0.13-0.37	0.26	-0.46-0.31	0.21
Females	0.26-0.47	0.33	0.18-0.42	0.29	0.11-0.41	0.2	0.13-0.37	0.24
Solitary								
Males	0.07-0.72	0.46	0.01-0.77	0.33	0.06-1.23	0.56	0.12-0.67	0.32
Females	0.14-0.7	0.48	0.03-0.64	0.31	0.03-0.66	0.28	0.02-0.52	0.26
Non-competitive								
Males	0.03-0.66	0.39	0.05-0.72	0.29	0.02-0.74	0.37	0.1-0.63	0.41
Females	0.11-0.58	0.4	0.03-0.71	0.26	0.03-0.46	0.2	0-0.44	0.2
Internally Paced								
Males	0.06-0.64	0.4	0.03-0.74	0.27	-0.01-0.84	0.41	-0.11-0.62	0.36
Females	0.04-0.57	0.37	0.02-0.76	0.25	0.08-0.44	0.2	0-0.41	0.22

Supplementary Table 2.2 *95-percentile ranges, and medians of tracking coefficient*

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
Aerobics	Aerobics	Aerobics		no	no	4
	High impact	Combi-fit		no	no	4
		Grit		no	no	4
		High impact		no	no	4
		HIIT		no	no	4
		Insanity		no	no	4
		Labooca		no	no	4
		Tae Bo		no	no	4
	Low impact	Aeronatics		no	no	4
		BBB		no	no	4
		Callanetics		no	no	4
		Low impact		no	no	4
	Steps	Body Steps		no	no	4
		Steps		no	no	4
Acrobatics/gymnastics	Acrobatics	Acrobatic Rock-n-Roll		Can be	Can be	3
		Acrobatics		Can be	Can be	3
		Tumbling		Can be	Can be	3
	Cheerleading	Cheerleading		yes	yes	3
	Color Guard	Color Guard		Can be	Can be	3
	Figure Scating	Figure Scating		Can be	yes	3
	Gymnastics	Cesar therapy		no	no	4
		Competitive		Can be	yes	4
		Gymnastics		no	no	4
		Heart Disease		no	no	4
		Mensendieck		no	no	4
		Pregnancy		no	no	4
		Recovery		no	no	4
		Rhythmic		Can be	no	3
		Voltige		Can be	Can be	3
	Platform diving	Platform diving		no	yes	3
Aiming based	Archery	Archery	yes	no	Can be	4
	Billiards	Billiards	yes	no	yes	4
	Bowling	Bowling	yes	Can be	yes	4
		Jeu de boules	yes	Can be	yes	4

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
		Petanque	yes	Can be	yes	4
	Curling	Curling	yes	yes	yes	4
	Darts	Darts	yes	no	yes	4
	Frisbee	Frisbee		yes	Can be	3
		Ultimate frisbee		yes	Can be	3
	Golf	Golf		no	Can be	4
	Shooting	Gun	yes	no	Can be	4
		Rifle	yes	no	Can be	4
Air based sports		Hang gliding		no	no	3
		Sky diving		no	no	3
Ball games	Baseball	Baseball		yes	yes	1
		Softball		yes	yes	1
Ball	Basketball	Basketball		yes	yes	1
		Basketball		yes	yes	1
		Basketball		yes	yes	1
	Cricket	Cricket		yes	yes	1
	Handball	Handball		yes	yes	1
	Hockey	Hockey		yes	yes	1
		Hockey		yes	yes	1
		Ice Hockey		yes	yes	1
		In-line Hockey		yes	yes	1
		Water Hockey		yes	yes	1
	Korfball play	Korfball play		yes	yes	1
	Polo	Polo		yes	yes	1
	Rugby	Rugby		yes	yes	1
		Rugby		yes	yes	1
	Soccer	Beach Soccer		yes	yes	1
		Floorball		yes	yes	1
		Hall sports		yes	yes	1
		Hall sports		yes	yes	1
		Indoor soccer		yes	yes	1
		Indoor soccer		yes	yes	1
		Soccer		yes	yes	1
	Volleyball	Beach Volleyball		yes	yes	1
		Competitive		yes	yes	1

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
		Volleyball		yes	yes	1
		Volleyball (sitting)		yes	yes	1
	Various	Ball games (other)		yes	yes	1
Cardiotraining/ edndurance	Athlons	Biathlon		no	yes	3
		Pentathlon		no	yes	3
		Triathlon		no	yes	3
	Cardiotraining	Cardiotraining		no	no	4
		Elliptical trainer		no	no	4
		Rowing machine		no	no	4
		Spinning		no	no	4
		Water Spinning		no	no	4
	Cycling	Cycling (race)		no	Can be	4
		Handbiking		no	no	4
		Mountainbiking		no	Can be	4
	Endurance	Competitive		no	yes	4
		Endurance		no	no	4
		Fast walking		no	Can be	4
		Freerunning		no	Can be	3
		Jogging		no	no	4
		Jumping rope		no	no	4
		Running		no	Can be	4
		Treadmill		no	no	4
		Trampoline		no	no	4
Cardiotraining/ edndurance (cont.)		Water jogging		no	no	4
	Skating	In-line skating		no	Can be	4
		Rollarskating		no	no	4
		Skateboarding		no	Can be	4
	Stationairy	Cycling (tours)		no	no	4
		Hometrainer		no	no	4
		Hypoxi		no	no	4
Climbing	Mountain	Mountain		no	no	3
	Wall Climbing	Wall Climbing		no	no	3
		Canyoning		no	no	3
		Potholing		no	no	3
Dancing	Ballet (like)	Ballet		no	Can be	3

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
Dancing (cont.)	Ballroom dancing	Jazz dancing		no	no	3
		Modern dancing		no	Can be	3
		Ballroom dancing		no	Can be	3
		Competitive		no	yes	3
		Latin dancing		no	Can be	3
		Tango dancing		no	Can be	3
	Cultural dancing	Belly dancing		no	no	3
		Bio dancing		no	no	3
		Body by Dance		no	no	3
		Breakdance		no	no	3
		Flamengo		no	no	3
		Hiphop dancing		no	no	3
		Salsa dancing		no	no	3
		Showdance		no	no	3
		Streetdance		no	no	3
	Dancing	Country dancing		no	no	3
		Dancing		no	no	3
		Dancing		no	no	3
		Disco dancing		no	no	3
		Folk dancing		no	no	3
		Line dancing		no	no	3
	Dancing aerobics	Aerobic dance		no	no	3
		Rock & Roll		no	no	3
		Shabam		no	no	3
		Zumba		no	no	3
	Gogo dancing	Gogo dancing		no	no	3
	Tapdancing	Tapdancing		no	no	3
	Twirling	Twirling		no	Can be	3
Equestrian		Carriage driving	yes	no	Can be	3
		Dressage		no	Can be	3
		Horse training		no	no	3
		Horseback riding		no	Can be	3
Fighting sports	Boxing	Boxing		no	yes	2
	Fencing	Fencing		no	yes	2
Fighting sports	Martial arts	Aiki jitsu		no	yes	2

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
(cont.)		Barokai		no	yes	2
		Capoeira		no	Can be	3
		Choi Kwan Do		no	yes	2
		Eskrima		no	yes	2
		Iaido		no	yes	2
		Judo		no	yes	2
		Jujitsu		no	yes	2
		Karate		no	yes	2
		Kempo		no	yes	2
		Kickboxing		no	yes	2
		Kickfun		no	yes	2
		Kobudo		no	yes	2
		Krav Maga		no	yes	2
		Kung Fu		no	yes	2
		Martial arts		no	yes	2
		Ninjatsa		no	yes	2
		Tae kwan do		no	yes	2
		Tang soo do		no	yes	2
		Thai boxing		no	yes	2
		Wing Chun		no	yes	2
	Sword fighting	Sword fighting		no	yes	2
	Wrestling	Wrestling		no	yes	2
Fitness training	Bootcamp	Bootcamp		no	no	4
	Boxing, punching	Boxing, punching		no	no	4
	Calisthenics	Calisthenics		no	no	4
		Calisthenics (back		no	no	4
		Calisthenics		no	no	4
		Curves		no	no	4
		Powerplate		no	no	4
		Slender you		no	no	4
		TMF heat cabine		no	no	4
	Fitness	55+ Sport		no	no	4
		Fitness		no	no	4
		Fysiofitness		no	no	4
		Nederlands on		no	no	4

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
Fitness training (cont.)	Group fitness	7-min workout		no	no	4
		Body Attack		no	no	4
		Body Boost		no	no	4
		Body Combat		no	no	4
		Body Fit		no	no	4
		Body Five		no	no	4
		Body Heat		no	no	4
		Body Jam		no	no	4
		Body Kick		no	no	4
		Body Power		no	no	4
		Body Pump		no	no	4
		Body Shape		no	no	4
		Body Styling		no	no	4
		Bodyline		no	no	4
		Crossfit		no	no	4
		Fat attack/burn		no	no	4
		Femme Fit		no	no	4
		Group fitness		no	no	4
		Keep-fit		no	no	4
		Piloxing		no	no	4
		X-CO		no	no	4
	Home exercises	Home exercises		no	no	4
		Video workout		no	no	4
	Pole Fitness	Pole Fitness		no	Can be	4
	Water aerobics	Water aerobics		no	no	4
	Weight lifting	Bodybuilding		no	yes	4
		Weight lifting		no	Can be	4
	Weight training	Circuit training		no	no	4
		CRXworx		no	no	4
		Dry training		no	no	4
		Total body		no	no	4
		Weight training		no	no	4
Mind games		Bridge	yes	yes	yes	4
		Checkers	yes	no	yes	4
		Chess	yes	no	yes	4

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
		Mind Games	yes	no	Can be	4
Motor sports		Drag Racing	yes	no	yes	3
		Go carting	yes	no	yes	3
		Motor-cross		no	yes	3
Swimming	Diving	Diving		no	no	3
		Scubadiving		no	no	3
	Lifeguard	Lifeguard		no	no	4
Swimming	Swimming	Competitive		no	yes	4
(cont.)		Pregnancy		no	no	4
		Relaxation		no	no	4
		Rheumatic		no	no	4
		Swimming		no	Can be	4
		Swimming (laps)		no	no	4
		Swimming		no	no	4
		Swimming (sport)		no	Can be	4
		Swimming		no	no	4
		Synchronized		no	yes	3
	Water polo	Water polo		yes	yes	1
Racket sports	Badminton	Badminton		Can be	yes	2
		Badminton		Can be	yes	2
		Competitive		Can be	yes	2
Racket sports	Ping pong	Ping pong		no	yes	2
(cont.)	Squash	Racquetball		no	yes	2
		Squash		no	yes	2
	Tennis	Tennis		Can be	yes	2
		Tennis		Can be	yes	2
Rowing	Canoeing	Canoeing		Can be	Can be	3
		Kayaking		Can be	Can be	3
	Rowing	Competitive		yes	yes	3
		Rowing		yes	Can be	3
Tai chi/yoga	Pilates	Body Balance		no	no	4
		Pilates		no	no	4
	Tai chi	Aikido		no	no	3
		Chu Yu Do		no	no	3
		Hapkido		no	no	3

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
		Pentjak Silat		no	no	3
		Qi gong		no	no	3
		Tai Chi		no	no	3
Thai chi/yoga	Yoga	Baby Balance		no	no	4
(cont.)		Ismakogie		no	no	4
		Stretching		no	no	4
		Tacoyo		no	no	4
		Yoga (hatha)		no	no	4
		Yoga		no	no	4
		Yoga (power)		no	no	4
		Yoga (Surya		no	no	4
Track-and-field	Jumping	Hurdles		no	Can be	3
		High jump		no	Can be	3
		Fierljeppen		no	Can be	3
		Long jump		no	Can be	3
		Pole vault		no	Can be	3
	Running	Sprinting		no	Can be	4
	Throwing	Discus Throwing		no	yes	3
Walking based	Hiking	Fitwalking		no	no	4
		Hiking		no	no	4
		Mountain hiking		no	no	4
		Marching		no	no	4
		Nordic Walking		no	no	4
		Orienteering		no	no	4
		Survival trail		no	Can be	3
	Referee/coach	Referee		no	no	4
		Sport coach		no	no	4
Water activities	Fishing	Fishing	yes	no	Can be	3
	Jet-skiing	Jet-skiing		no	no	3
		Water skiing		no	Can be	4
	Kiting	Kite flying	yes	no	no	3
		Kitesurfing		no	Can be	3
Water activities	Sailing	Competitive		no	yes	3
(cont.)		Sailing		no	Can be	3
		Yachting		no	no	3

Exercise domain	Subdomain	Reported activities	Excluded	Team sport	Competitive sport	Internal/ External pacing
	Surfing	Flowboarding		no	no	4
		Paddleboarding		no	no	4
		Surfing		no	no	4
		Wakeboarding		no	Can be	4
		Windsurfing		no	Can be	3
Winter activities	Cross Country	Cross Country		no	Can be	4
	Ice Skating	Ice Skating		Can be	Can be	4
		Ice Skating (long		Can be	yes	4
		Ice Skating		Can be	yes	4
		Ice skating (short		yes	yes	4
	Skiing	Ski-gym		no	no	4
		Skiing (indoor)		no	Can be	4
		Snowboarding		no	Can be	4

Supplementary Table 5.1*. Classification of various types of exercise This is a complete list of the exercise activities that were reported by the NTR participants (NB: minor variations of the same activity not listed to exhaustion). Domains and subdomains are not used in the analyses and just serve to help the reader navigate this table. Exclude means that the activity was considered below the threshold for moderate-to-vigorous intensity.* ***Team sports*** *means any sport that is or can be performed in a team-setting, where multiple individuals work together to achieve a shared objective. If a sport can be a team sport is counted as a team sport over a solitary sport.* ***Competitive sport*** *means a type of sport where a team or individual (can) competes against another team or individual to win. If a sport can be a competitive sport is counted as a team sport over a solitary sport.* ***Internal/external pacing*** *is coded as follows: 1 Highly externally paced, influences from both teammates and opponent; 2 Intermediate extenally paced, influenced by opponent or teammates only; 3 Low externally paced, influenced by elements (wind/water/ music/synchronic team movements); 4 Internally paced for the purposes of this thesis, categories 1 and 2 are considered externally paced, and category 4 is considered internally paced. Category 3 is excluded.*

		Age	Birth year	Total	Team	Competitive	Externally paced	Solitary	Non-competitive	Internally paced
Male parents	Mean	48.21	1959	673.50	137.03	244.52	234.82	536.47	428.98	414.69
	SD	7.41	9.95	1067.79	408.11	549.25	532.82	981.62	912.07	901.12
Female parents	Mean	45.89	1962	502.92	39.95	118.27	116.63	462.97	384.65	345.91
	SD	7.68	10.10	720.78	215.11	357.05	357.31	689.37	630.01	600.38
Male children	Mean	27.03	1978	1210.77	559.08	710.40	690.19	651.70	500.38	460.05
	SD	12.00	12.68	1384.59	945.37	1055.70	1044.40	1090.23	959.64	917.58
Female children	Mean	29.48	1977	780.26	223.30	305.69	301.97	556.96	474.57	391.08
	SD	13.06	13.70	1022.47	627.34	707.39	711.73	854.72	783.18	712.11
MZM	Mean	27.26	1978	1253.26	565.40	734.03	712.84	687.86	519.22	478.73
	SD	12.41	12.82	1421.74	928.63	1078.47	1054.75	1146.86	970.73	918.54
MZF	Mean	29.88	1977	796.28	229.72	325.25	318.82	566.56	471.03	392.72
	SD	13.63	13.96	1068.76	641.64	739.33	725.18	910.94	818.66	755.74
DZM	Mean	24.89	1979	1258.62	646.01	788.97	769.86	612.61	469.65	421.32
	SD	11.17	12.22	1375.40	1039.40	1109.98	1099.72	1027.20	922.06	891.88
DZF	Mean	27.20	1979	824.07	254.63	332.94	330.97	569.43	491.13	399.36
	SD	12.00	13.07	1071.69	686.06	747.68	763.81	860.50	796.93	720.05

Supplementary table 5.2*. Exercise and age descriptives per sex and family relation. MZM: Monozygotic male twins; MZF: Monozygotic female twins; DZM: Males part of a dyzigotic twin pair; DZF: Females part of a dizygotic twin pair; Total: Total volume of exercise (METminutes /week) ; Team: volume of exercise in team-based activities; Comp: volume of exercise in competitive activities; Externally paced: volume of exercise in externally paced activities; Solitary: volume of exercise in solitary activities; Non-competitive: volume of exercise in non-competitive activities; Internally paced: volume of exercise in internally paced activities.*

		Total		Team		Solitary		Competitive		Non-Competitive		Ext. Paced		Int. Paced	
		ACE	ACDE	ACE	ACDE	ACE	ACDE	ACE	ACDE	ACE	ACDE	ACE	ACDE	ACE	ACDE
Household	A	0.28	0.19	0.32	0.14	0.23	0.15	0.3	0.17	0.21	0.13	0.31	0.18	0.2	0.13
	C	0.2	0.2	0.09	0.11	0.19	0.2	0.19	0.2	0.17	0.17	0.19	0.2	0.17	0.17
	D		0.21		0.32		0.18		0.26		0.17		0.26		0.16
	E	0.52	0.4	0.59	0.42	0.58	0.48	0.51	0.37	0.62	0.52	0.5	0.36	0.63	0.54
	Log Likelihood	13317	13394	32470	32698	11870	11925	22595	22731	12312	12361	23165	23308	12604	12647
SibHousehold	A	0.25	0.22	0.25	0.17	0.22	0.19	0.27	0.23	0.19	0.17	0.29	0.23	0.19	0.16
	C	0.1	0.04	0.11	0.03	0.07	0.02	0.08	0.01	0.06	0.01	0.08	0	0.05	0
	D		0.15		0.26		0.14		0.21		0.14		0.23		0.14
	E	0.65	0.59	0.64	0.53	0.72	0.65	0.64	0.55	0.74	0.68	0.63	0.54	0.76	0.69
	Log Likelihood	12996	13018	32512	32615	11574	11593	22349	22407	12092	12111	22907	22974	12386	12405
SpouseHousehold	A	0.34	0.26	0.36	0.19	0.28	0.22	0.36	0.26	0.25	0.19	0.37	0.26	0.24	0.18
	C	0.25	0.24	0.19	0.16	0.25	0.25	0.26	0.25	0.22	0.22	0.27	0.26	0.21	0.21
	D		0.15		0.28		0.12		0.19		0.12		0.19		0.12
	E	0.42	0.35	0.46	0.36	0.47	0.41	0.37	0.31	0.52	0.47	0.36	0.29	0.55	0.49
	Log Likelihood	13411	13445	32543	32709	11994	12016	22720	22783	12408	12431	23295	23361	12676	12697
TwinHousehold	A	0.24	0.22	0.24	0.18	0.21	0.19	0.26	0.23	0.19	0.17	0.27	0.23	0.18	0.16
	C	0.14	0.08	0.16	0.07	0.11	0.06	0.14	0.07	0.1	0.05	0.14	0.06	0.09	0.04
	D		0.12		0.23		0.1		0.16		0.1		0.17		0.11
	E	0.61	0.58	0.6	0.53	0.68	0.65	0.6	0.55	0.71	0.68	0.59	0.54	0.73	0.69
	Log Likelihood	13010	13022	32554	32624	11590	11598	22389	22417	12106	12115	22949	22982	12397	12407

Supplementary Table 5.3. *Estimates from all models tested with the Mendel software package.*

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
r1 > 0	1,608	0.24>0	61.88	1	3.66×10^{-15}	√			
r4 > 0	10,711	0.71>0	2335.71	1	$< 10^{-200}$	√			
r1 > r2	1,420	0.24>0.12	16.61	2	2.48×10^{-4}		√	-	
r2 > r3	419	0.12>0.31	8.12	2	0.017		-	√	
$r2_{mz} > r2_{dz}$	465	0.12>0.12	0	1	>0.99		-	√	
$r3_{mz} > r3_{dz}$	157	0.36>0.14	2.02	1	0.16		-	√	
r4 > r1	1,608	0.71>0.24	656.74	1	7.66×10^{-145}				√

Supplementary Table 5.4*. Spousal resemblance and its sources in team-based sports activities. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is r1 > r2 > r3, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is (r1 = r2 ≥ r3), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$. Under marital interaction we expect r4 > r1; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with* $\alpha = 0.01$.

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
r1 > 0	1,608	0.23>0	81.91	1	1.43×10^{-19}	√			
r4 > 0	10,711	0.6>0	1525.94	1	$< 10^{-200}$	√			
r1 > r2	1,420	0.23>0.07	36.25	2	1.35×10^{-8}		√	-	
r2 > r3	419	0.07>0.2	4.16	2	0.12		-	√	
$r2_{mz} > r2_{dz}$	465	0.06>0.11	1.47	1	0.22		-	√	
$r3_{mz} > r3_{dz}$	157	0.18>0.23	0.19	1	0.66		-	√	
r4 > r1	1,608	0.6>0.23	326.41	1	5.80×10^{-73}				√

Supplementary Table 5.5*. Spousal resemblance and its sources in competitive sports activities. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is r1 > r2 > r3, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is (r1 = r2 ≥ r3), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$. Under marital interaction we expect r4 > r1; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with $\alpha = 0.01$.*

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
r1 > 0	1,608	0.26>0	93.04	1	5.12×10^{-22}	√			
r4 > 0	10,711	0.61>0	1664.68	1	$< 10^{-200}$	√			
r1 > r2	1,420	0.26>0.11	31.94	2	1.16×10^{-7}		√	-	
r2 > r3	419	0.11>0.2	2.58	2	0.28		-	√	
$r2_{mz} > r2_{dz}$	465	0.1>0.13	0.37	1	0.54		-	√	
$r3_{mz} > r3_{dz}$	157	0.17>0.24	0.4	1	0.53		-	√	
r4 > r1	1,608	0.61>0.26	314.86	1	1.91×10^{-70}				√

Supplementary Table 5.6*. Spousal resemblance and its sources in externally paced sports activities. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is r1 > r2 > r3, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is (r1 = r2 ≥ r3), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$. Under marital interaction we expect r4 > r1; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with* $\alpha = 0.01$.

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
r1 > 0	1,608	0.16>0	69.15	1	9.12×10^{-17}	√			
r4 > 0	10,711	0.28>0	692.13	1	1.54×10^{-152}	√			
r1 > r2	1,420	0.16>0.08	18.55	2	9.37×10^{-5}		√	-	
r2 > r3	419	0.08>0.06	0.71	2	0.70		-	√	
$r2_{mz} > r2_{dz}$	465	0.07>0.1	0.34	1	0.56		-	√	
$r3_{mz} > r3_{dz}$	157	0.07>0.03	0.18	1	0.67		-	√	
r4 > r1	1,608	0.28>0.16	27.8	1	1.35×10^{-7}				√

Supplementary Table 5.7*. Spousal resemblance and its sources in solitary paced sports activities. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is r1 > r2 > r3, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is (r1 = r2 ≥ r3), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$ Under marital interaction we expect r4 > r1; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with* $\alpha = 0.01$.

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
r1 > 0	1,608	0.14>0	47.22	1	6.33×10^{-12}	√			
r4 > 0	10,711	0.23>0	535.41	1	1.88×10^{-118}	√			
r1 > r2	1,420	0.14>0.06	13.19	2	1.37×10^{-3}		√	-	
r2 > r3	419	0.06>0.07	0.61	2	0.74		-	√	
$r2_{mz} > r2_{dz}$	465	0.06>0.09	0.43	1	0.51		-	√	
$r3_{mz} > r3_{dz}$	157	0.08>0.03	0.25	1	0.61		-	√	
r4 > r1	1,608	0.23>0.14	19.98	1	7.82×10^{-6}				√

Supplementary Table 5.8*. Spousal resemblance and its sources in non-competitive sports activities. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is r1 > r2 > r3, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is (r1 = r2 ≥ r3), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$. Under marital interaction we expect r4 > r1; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with* $\alpha = 0.01$.

Hypothesis	N	Estimates	Δ-2LL	Δdf	p	Spousal Resemblance	Phenotypic Assortment	Social Homogamy	Marital interaction
r1 > 0	1,608	0.14>0	48.93	1	2.66×10^{-12}	√			
r4 > 0	10,711	0.22>0	496.66	1	5.06×10^{-110}	√			
r1 > r2	1,420	0.14>0.07	15.32	2	4.71×10^{-4}		√	-	
r2 > r3	419	0.07>0.06	0.89	2	0.64		-	√	
$r2_{mz} > r2_{dz}$	465	0.05>0.1	1.1	1	0.30		-	√	
$r3_{mz} > r3_{dz}$	157	0.08>0.03	0.23	1	0.63		-	√	
r4 > r1	1,608	0.22>0.14	12.1	1	5.04×10^{-4}				√

Supplementary Table 5.9*. Spousal resemblance and its sources in internally paced sports activities. Hypothesis: expectations for the patterns of twin-spouse correlations (r1), co-twin spouse-correlations(r2), spouse1-spouse2 correlations (r3), and parent-parent correlations (r4) under phenotypic assortment, social homogamy, and marital interaction. Under phenotypic assortment the expected pattern of correlations is r1 > r2 > r3, $r2_{MZ} > r2_{DZ}$, $r3_{MZ} > r3_{DZ}$. Under social homogamy the expected pattern is (r1 = r2 ≥ r3), $r2_{MZ} = r2_{DZ}$, $r3_{MZ} = r3_{DZ}$. Under marital interaction we expect r4 > r1; N: Number of complete pairs in the data (if two correlation coefficients are compared the lowest N is presented); Δ-2LL: Difference in -2 log-likelihood compared to the base model; Δdf: Difference in degrees of freedom compared to the base model, and used in the chi-squared test; p: p-value of the chi-squared difference test comparing the base model to the constrained model; √: Hypothesis supported with $\alpha = 0.01$.*

The following supplementary tables do not fit standard paper size and can be found online at

https://drive.google.com/drive/folders/1i8lUQJcMmzj0QTPtGTxsFuTn8tpKlKZr

Supplementary table 6.1
Supplementary table 6.2
Supplementary table 6.3
Supplementary table 6.4
Supplementary table 6.5
Supplementary table 6.6
Supplementary table 6.7
Supplementary table 6.8

	Factor1	Factor2
	Competitive/team activities	Non-competitive solitary activities
Competitive exercise	0.994	0.035
Team-based exercise	0.802	-0.053
Externally paced exercise	0.998	0.028
Non-competitive exercise	-0.064	0.978
Solitary exrecise	0.111	0.921
Internally paced exercise	-0.049	0.942

Supplementary table 7.1. *Factor Analysis on the weekly METminutes spent on the six different exercise activities yields two main Factors: Competitive/team based and Non-competitive/solitary.*

The following supplementary tables do not fit standard paper size and can be found online at

https://drive.google.com/drive/folders/1i8lUQJcMmzj0QTPtGTxsFuTn8tpKlKZr

Supplementary table 7.2

Supplementary table 7.3

	SNP-based heritability (SNP-h^2) (SE)	LD-reg based genetic correlation (Rg)					
		Neuroticism	NeuroticismMeta	Extraversion	Openness	Agreeableness	Conscientiousness
Neuroticism	0.12 (0.02)**						
NeuroticismMeta	0.10 (0.01)**	0.86 (0.03)**					
Extraversion	0.18 (0.01)**	-0.37 (0.05)**	-0.20 (0.04)**				
Openness	0.10 (0.01)**	-0.16 (0.08)	-0.06 (0.05)	0.34 (0.05)**			
Agreeableness	0.08 (0.01)**	-0.39 (0.07)**	-0.30 (0.05)**	0.22 (0.06)**	0.12 (0.08)		
Conscientiousness	0.09 (0.01)**	-0.16 (0.06)	-0.15 (0.04)**	0.15 (0.05)	-0.21 (0.07)*	0.24 (0.07)*	
StrenuousSports OrOtherExercises	0.06 (0.00)**	-0.20 (0.04)**	-0.22 (0.03)**	0.14 (0.04)*	0.15 (0.04)*	0.10 (0.06)	0.03 (0.05)

Supplementary Table 7.4*. SNP heritability (left column), and genetic correlations (right matrix) of all summary statistics used to generate polygenic risk scores. *: p < 0.0033; **: p < 0.0003; SNP-h2 (SE): SNP heritability as estimated by Ldscore regression, standard error included in parentheses; LD-reg based genetic correlation (Rg): Genetic correlation as estimated by Ldscore regression, standard error included in parentheses.*

		0.01		0.2		0.5		inf		max(Prior)		
Phenotype	**PRS**	**R2**	**p**	**R2**	**p**	**R2**	**p**	**R2**	**p**	**Prior**	**R2**	**p**
Neuroticism	Neuroticism	0.008	2.31×10^{-16}	0.008	1.09×10^{-16}	0.008	1.07×10^{-16}	0.006	6.98×10^{-13}	0.5	0.008	1.07×10^{-16}
	MetaNeuroticism	0.004	3.20×10^{-9}	0.020	9.63×10^{-40}	0.020	2.92×10^{-40}	0.016	1.61×10^{-31}	0.5	0.020	2.92×10^{-40}
Extraversion	Extraversion	0.010	1.47×10^{-20}	0.010	2.38×10^{-20}	0.010	2.45×10^{-20}	0.006	4.91×10^{-13}	0.01	0.010	1.47×10^{-20}
Openness	Openness	0.008	1.08×10^{-14}	0.008	7.27×10^{-15}	0.008	7.21×10^{-15}	0.006	7.34×10^{-12}	0.5	0.008	7.21×10^{-15}
Agreeableness	Agreeableness	0.004	2.27×10^{-9}	0.004	3.36×10^{-9}	0.004	3.41×10^{-9}	0.003	4.23×10^{-8}	0.01	0.004	2.27×10^{-9}
Conscientiousness	Conscientiousness	0.005	6.05×10^{-11}	0.005	7.87×10^{-11}	0.005	7.99×10^{-11}	0.003	8.37×10^{-8}	0.01	0.005	6.05×10^{-11}
Total exercise	StrenuousSports OtherExercises	0.001	1.33×10^{-3}	0.003	2.72×10^{-7}	0.003	2.54×10^{-7}	0.003	1.34×10^{-7}	inf	0.003	1.34×10^{-7}
Competitive/ Team activities	StrenuousSports OtherExercises	0.001	2.11×10^{-3}	0.002	2.69×10^{-5}	0.002	2.68×10^{-5}	0.002	2.20×10^{-5}	0.01	0.001	2.11×10^{-3}
Non-competitive/ Solitary activities	StrenuousSports OtherExercises	0.002	1.41×10^{-5}	0.004	8.91×10^{-10}	0.004	9.19×10^{-10}	0.004	1.64×10^{-9}	0.2	0.004	8.91×10^{-10}

Supplementary Table 7.5*. Predictions of the personality traits and exercise phenotypes in NTR from polygenetic risk based on the leave-NTR-out summary statistics obtained from meta-analysis on these phenotypes. LDpred was used with different priors to compute the Polygenic Risk Scores (PRS) and the prior that gave the highest prediction (last column) was used in the reported MR-DoC analyses. Phenotype: Personality or exercise phenotype predicted by PRS; PRS: Summary statistics used to generate PRS; The following columns are included once per prior: R2: R squared: variance in the phenotype explained by the PRS; p: p-value of the regression coefficient of the PRS; The last three columns represent the best-fitting (highest R2) prior: Prior: Prior where the highest R2 was found; R2: R squared of the highest prior with the highest R2; p: p-value of the regression coefficient of the PRS with the highest R2.*

Note: this print-friendly version of **Supplementary Table 7.5** is incomplete. Some priors (0.05, 0.1 and 0.3) have been removed to allow this table to fit on paper. The full table is available online at

https://drive.google.com/drive/folders/1i8IUQJcMmzj0QTPtGTxsFuTn8tpKIKZr

The following supplementary tables do not fit standard paper size and can be found online at

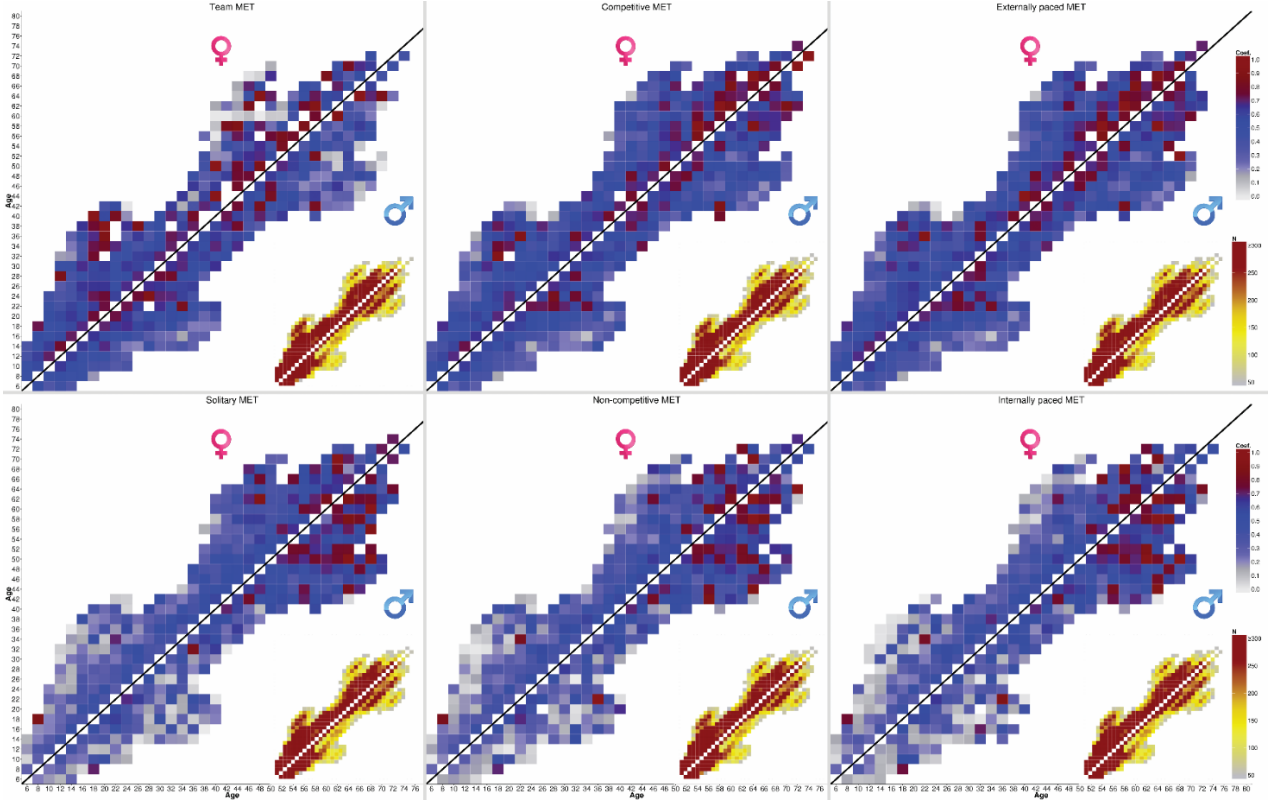
https://drive.google.com/drive/folders/1i8lUQJcMmzj0QTPtGTxsFuTn8tpKIKZr

Supplementary table 7.6

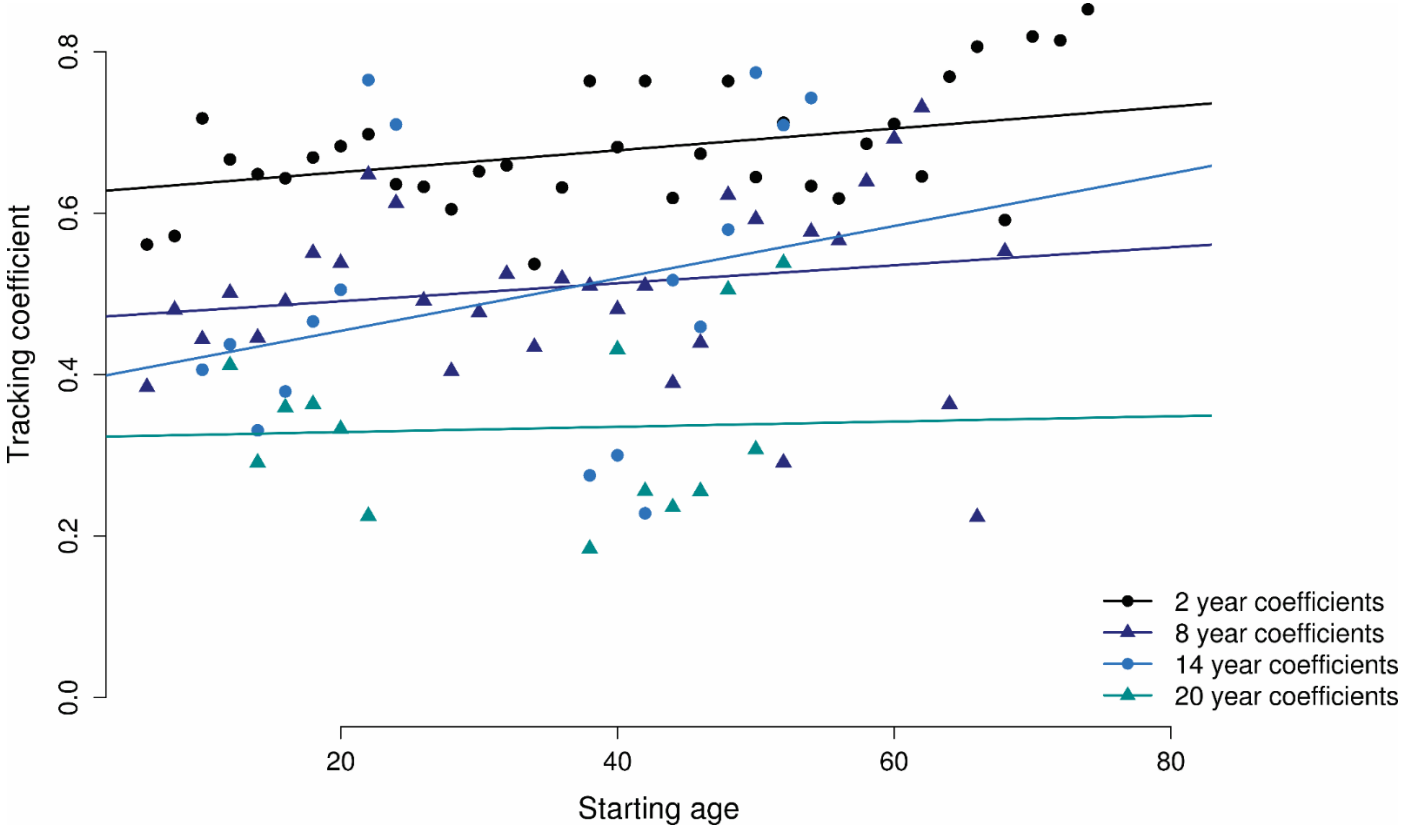
Supplementary table 7.7

Supplementary table 7.8

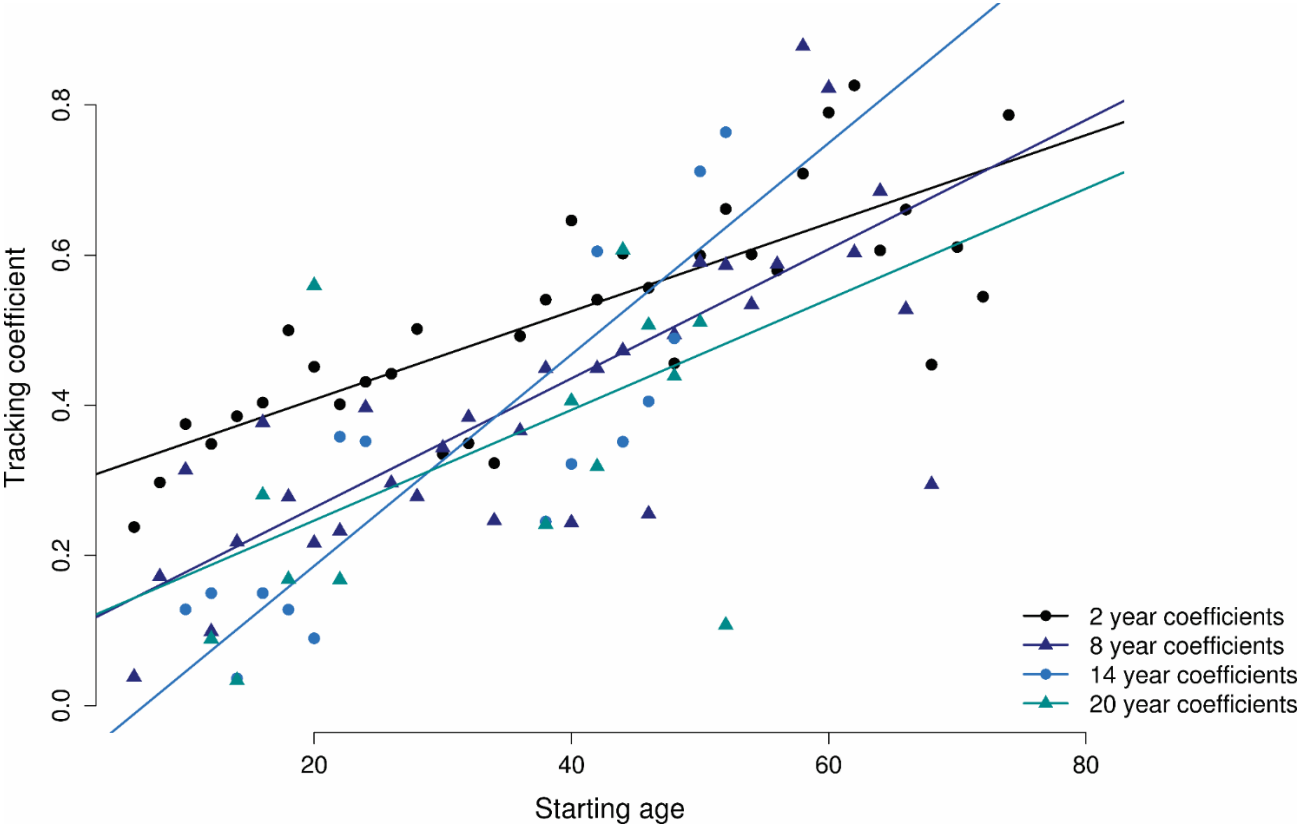
Supplementary figures



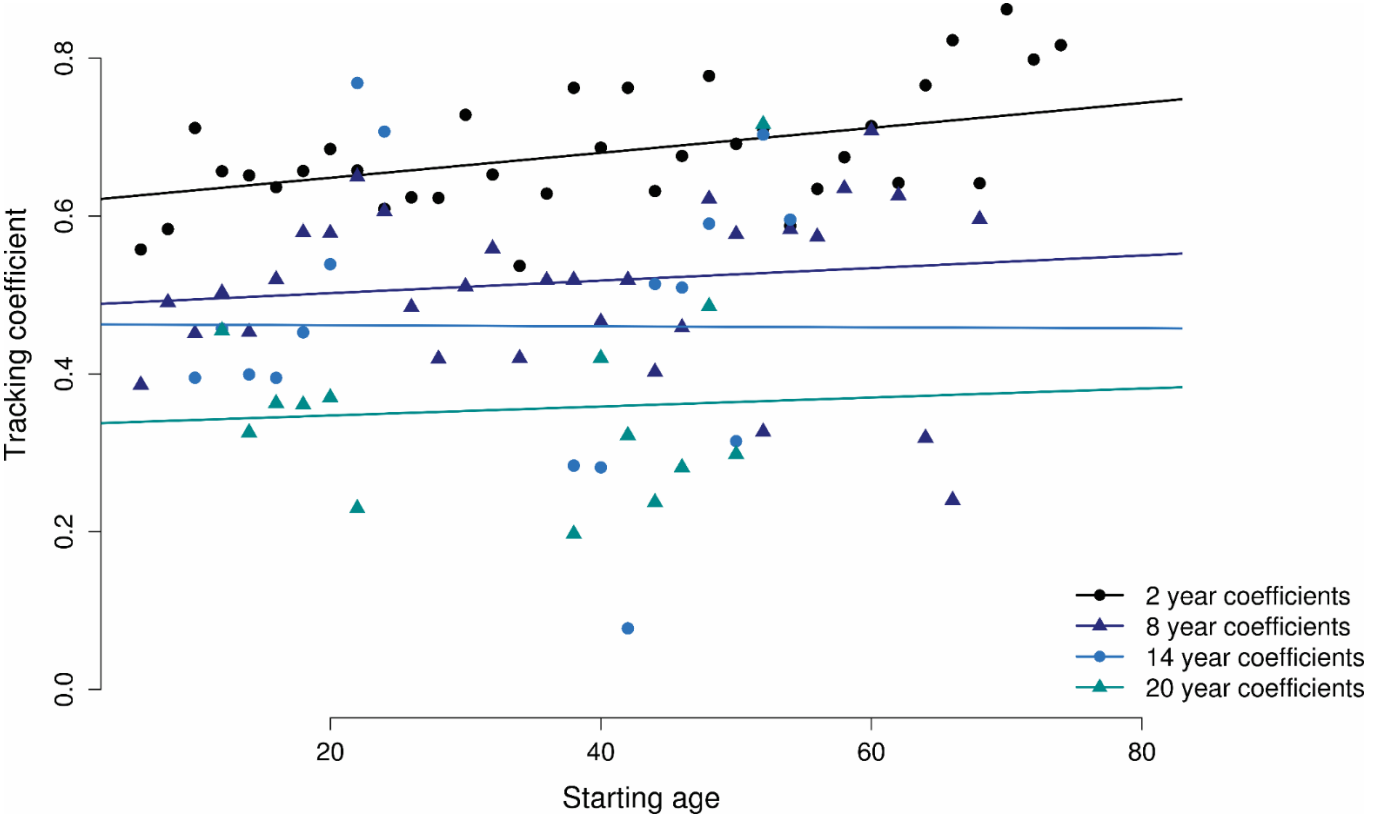
Supplementary Figure 2.1. *Tracking coefficient, and sample size heatmaps of exercise domains, similar to Figure 2.3 of the main manuscript. High-resolution version available online at* *https://drive.google.com/drive/folders/1i8IUQJcMmzj0QTPtGTxsFuTn8tpKIKZr*



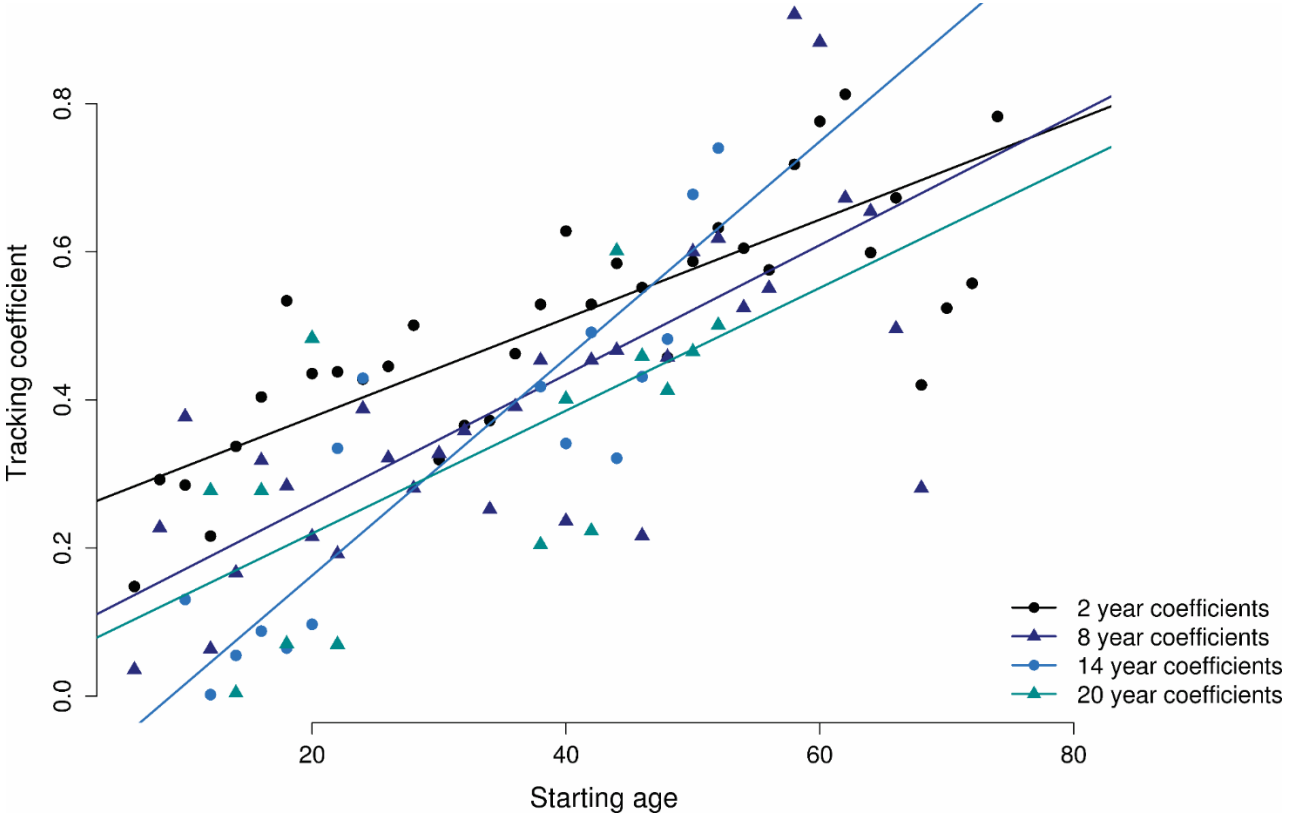
Supplementary Figure 2.2. *Tracking coefficients for volume of competitive exercise behavior as a function of age at baseline.*



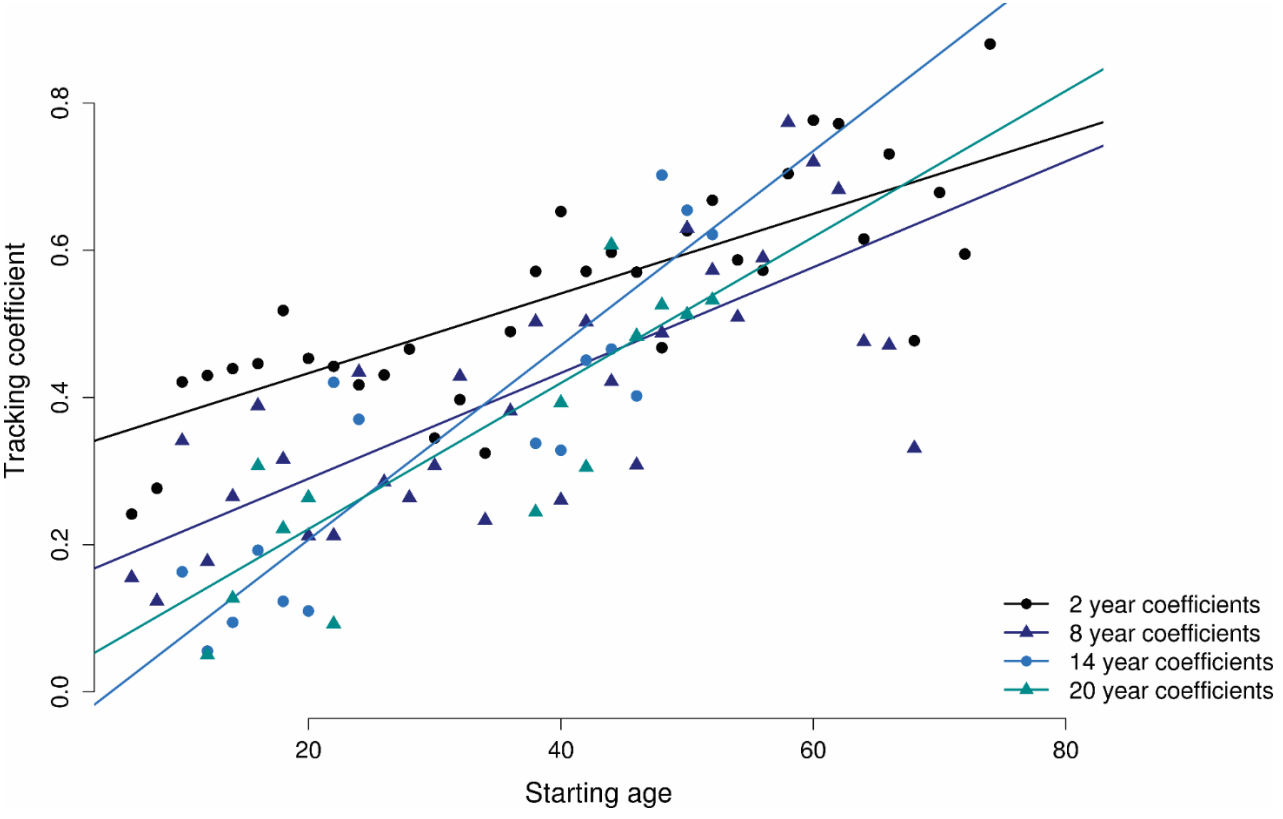
Supplementary Figure 2.3. *Tracking coefficients for volume of non-competitive exercise behavior as a function of age at baseline.*



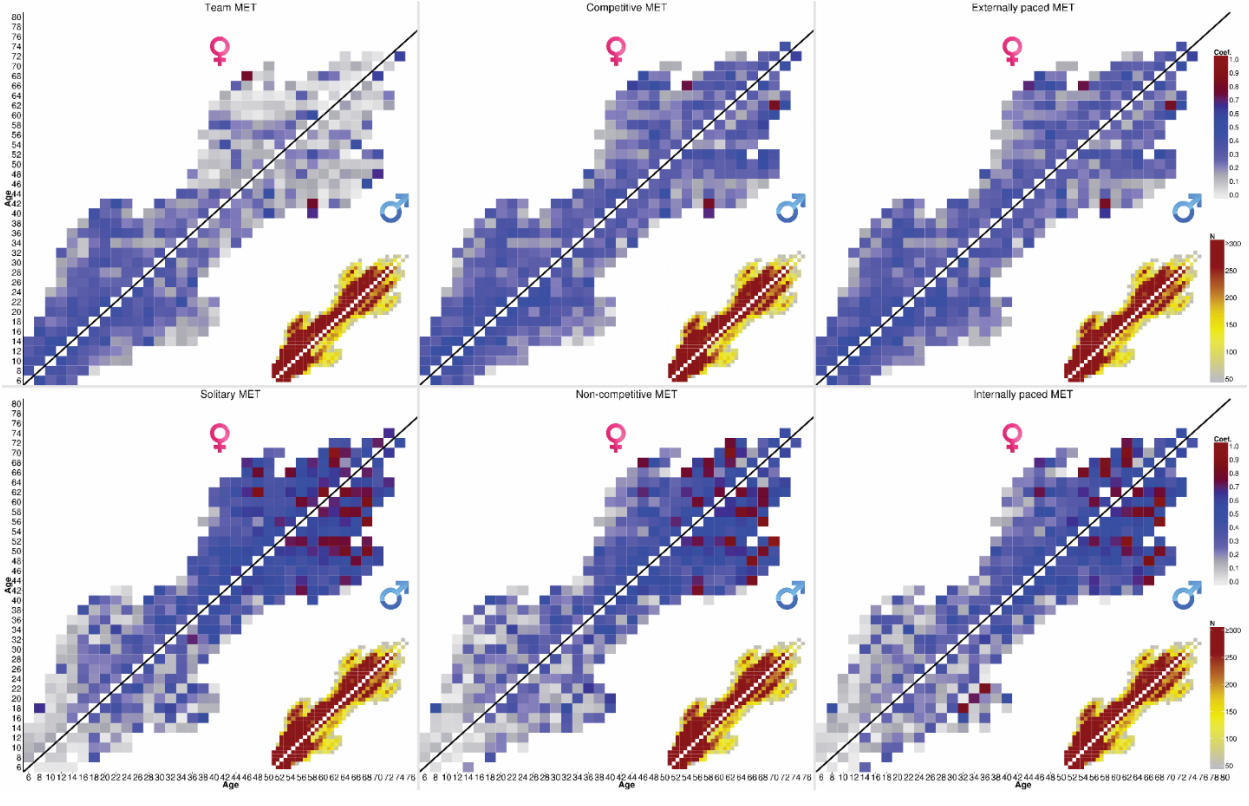
Supplementary Figure 2.4. *Tracking coefficients for volume of externally paced exercise behavior as a function of age at baseline.*



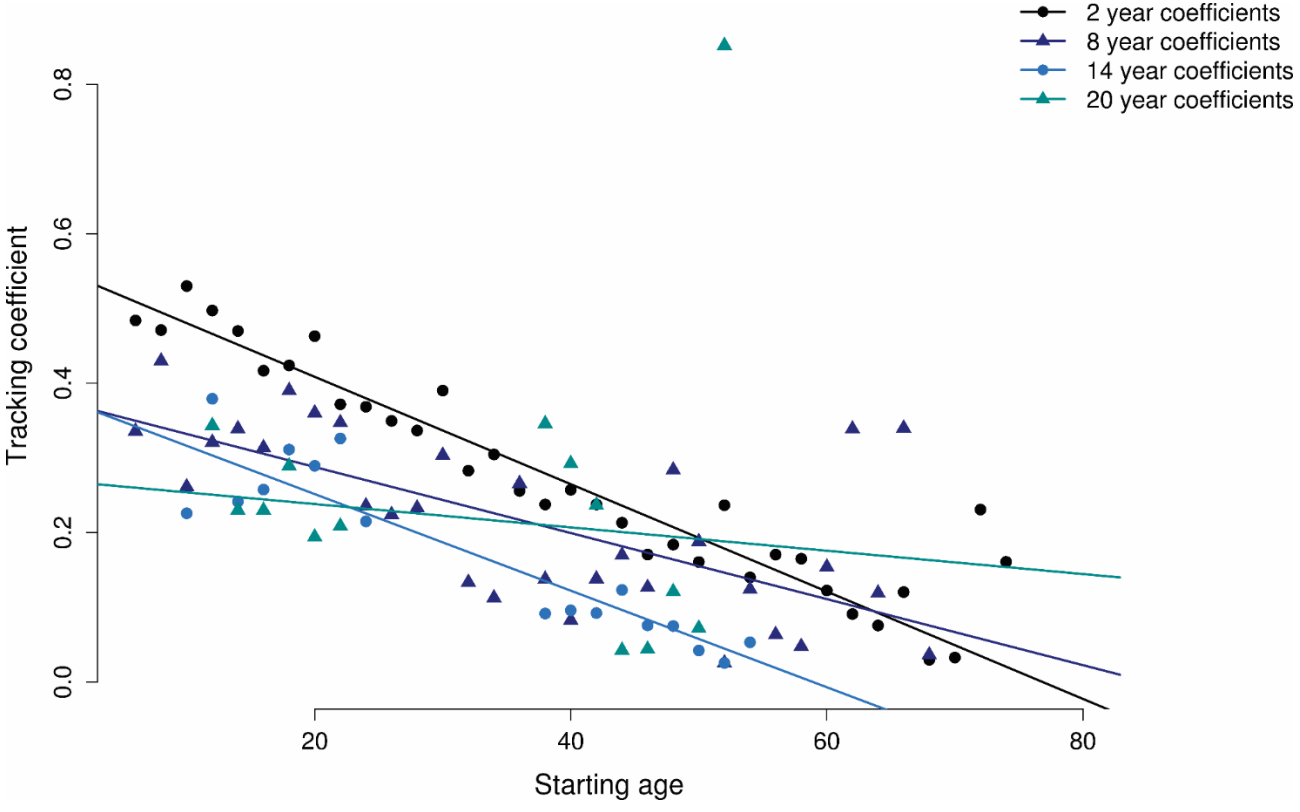
Supplementary Figure 2.5. *Tracking coefficients for volume of internally paced exercise behavior as a function of age at baseline.*



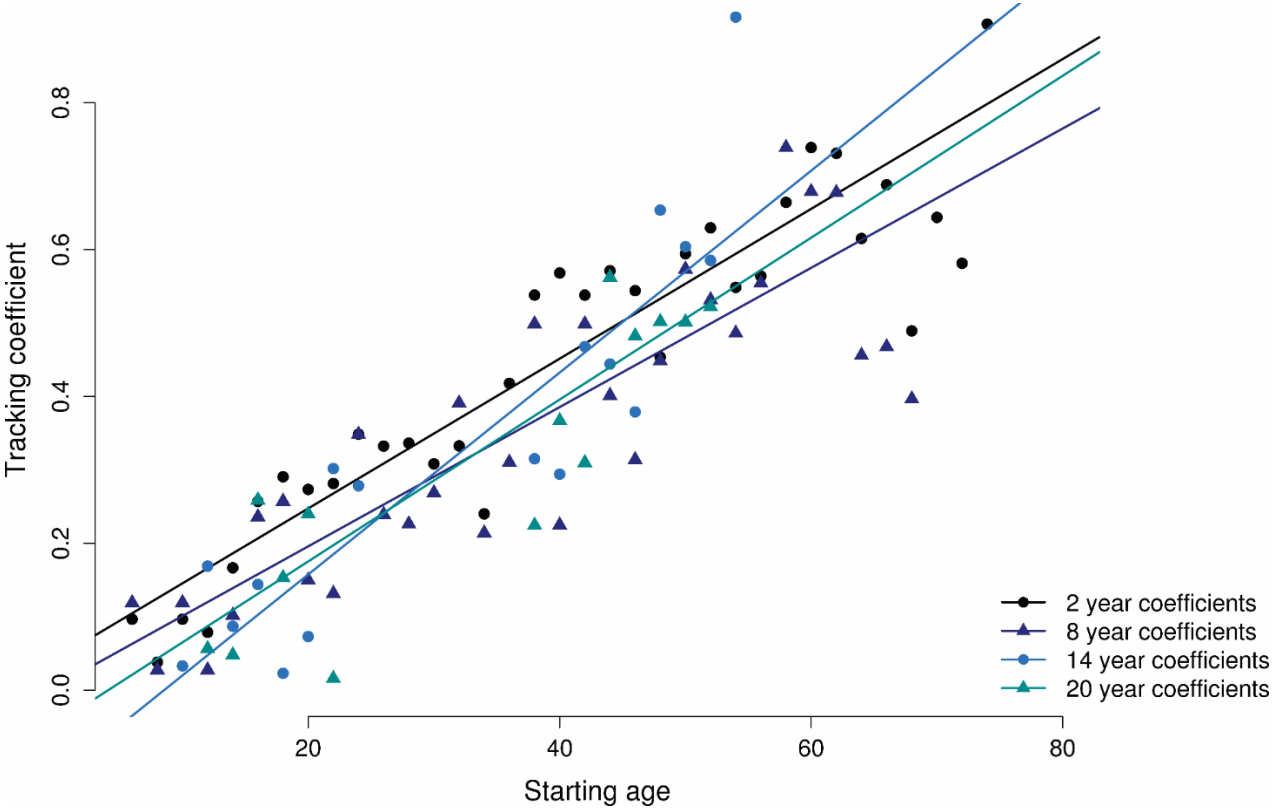
Supplementary Figure 2.6. *Tracking coefficients for volume of solitary exercise behavior as a function of age at baseline.*



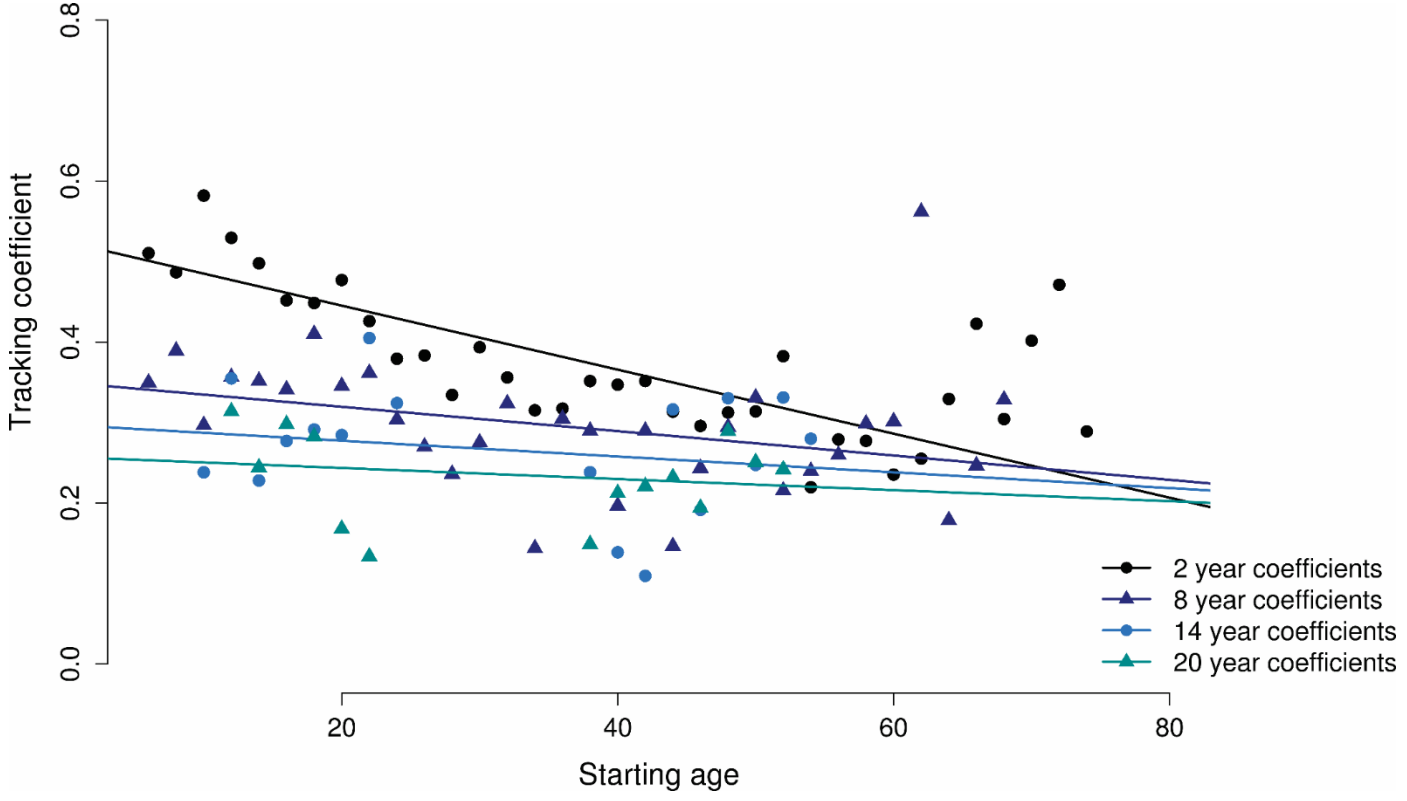
Supplementary Figure 2.7. *Between-domain tracking coefficient heatmaps of exercise domains at baseline and total volume of exercise at follow-up. High-resolution version available online at https://drive.google.com/drive/folders/1i8lUQJcMmzj0QTPtGTxsFuTn8tpKlKZr?*



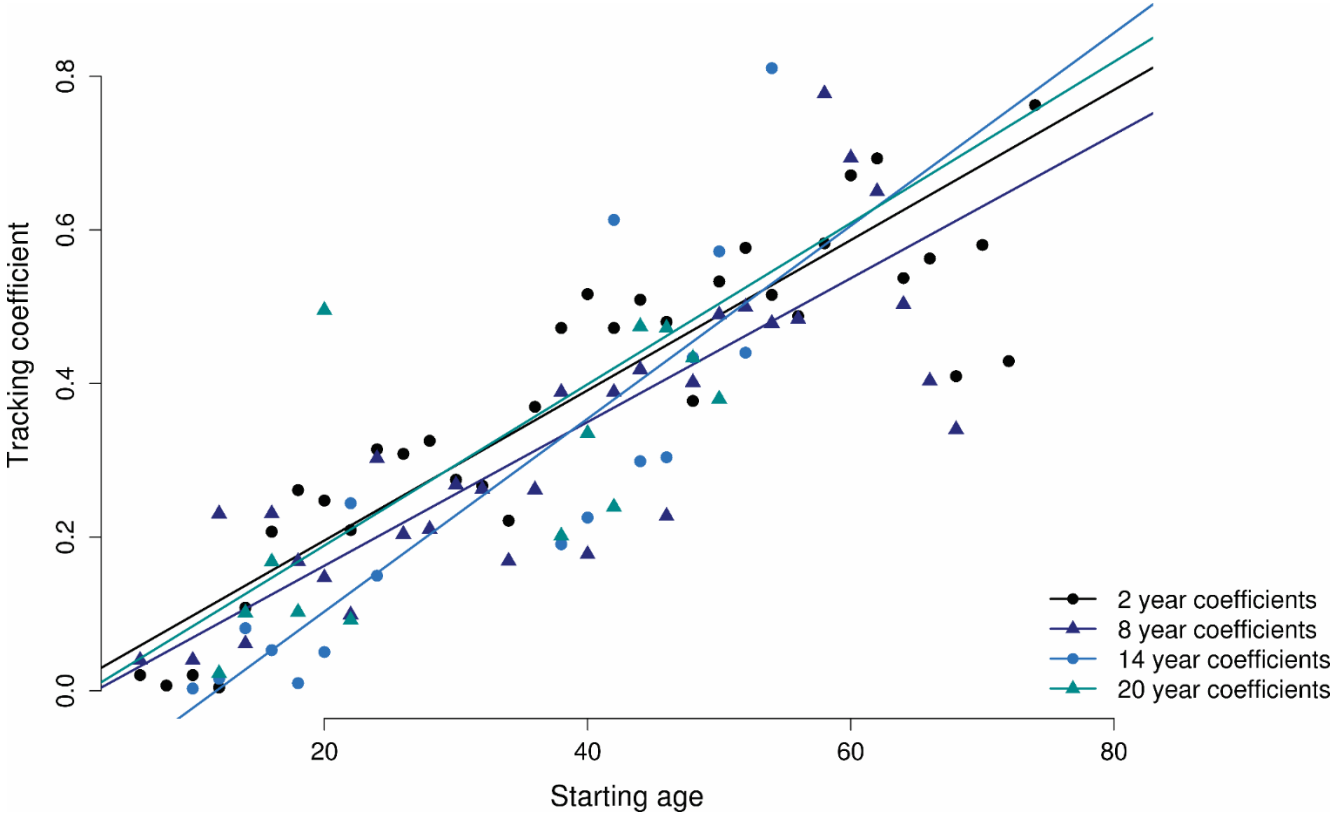
Supplementary Figure 2.8. *Tracking coefficients for volume of team-based exercise behavior at baseline and total volume of exercise at follow-up, as a function of age at baseline.*



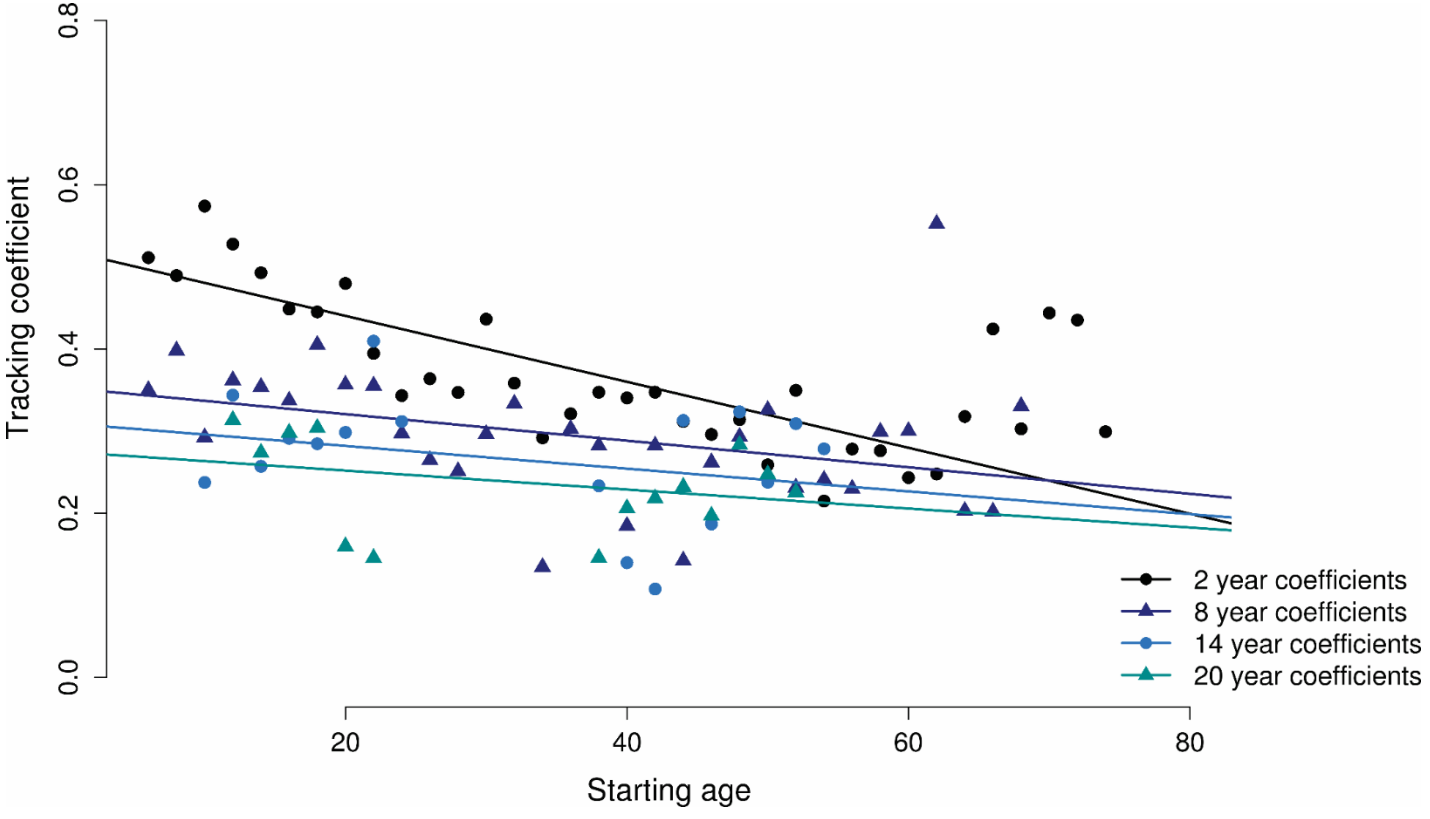
Supplementary Figure 2.9*. Tracking coefficients for volume of solitary exercise behavior at baseline and total volume of exercise at follow-up, as a function of age at baseline.*



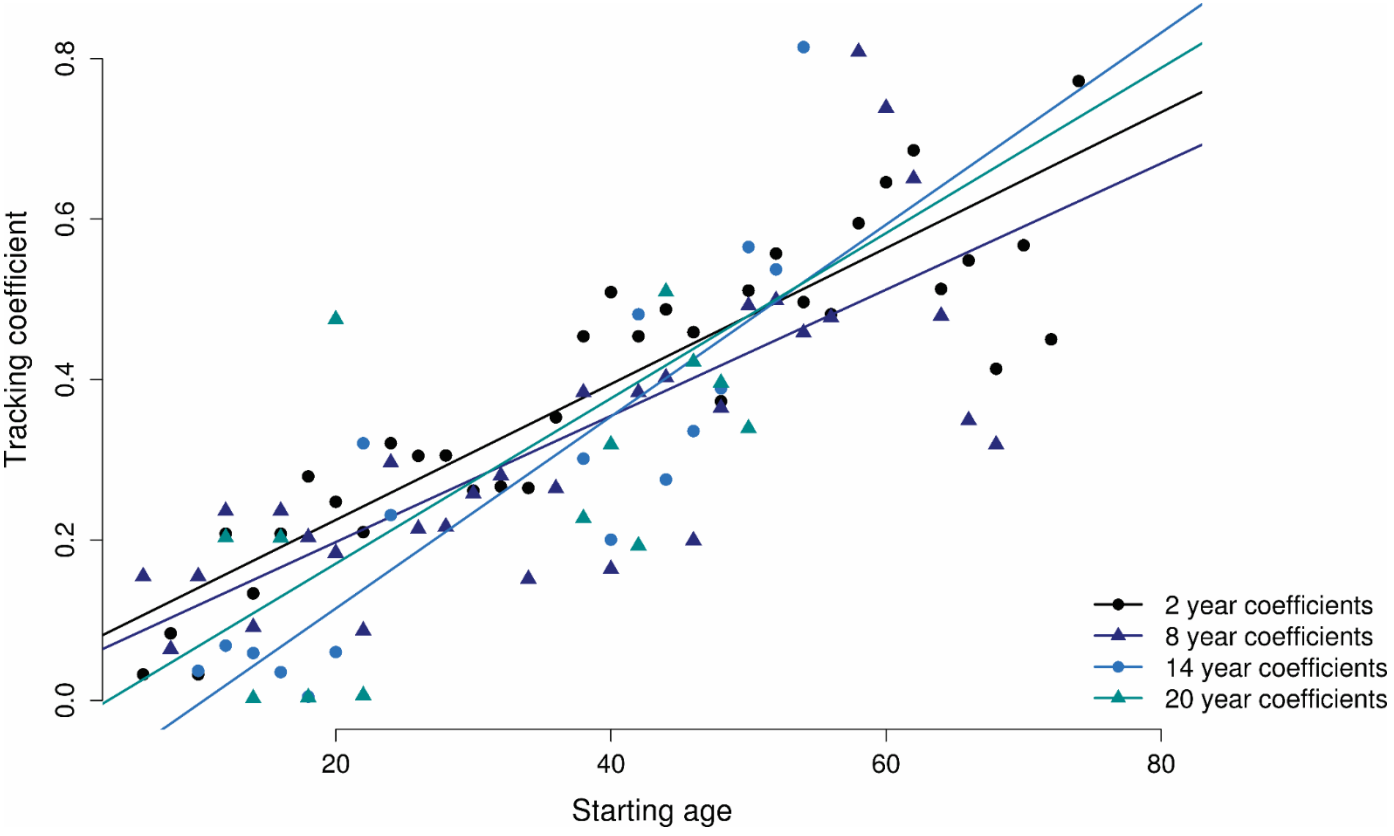
Supplementary Figure 2.10. *Tracking coefficients for volume of competitive exercise behavior at baseline and total volume of exercise at follow-up, as a function of age at baseline.*



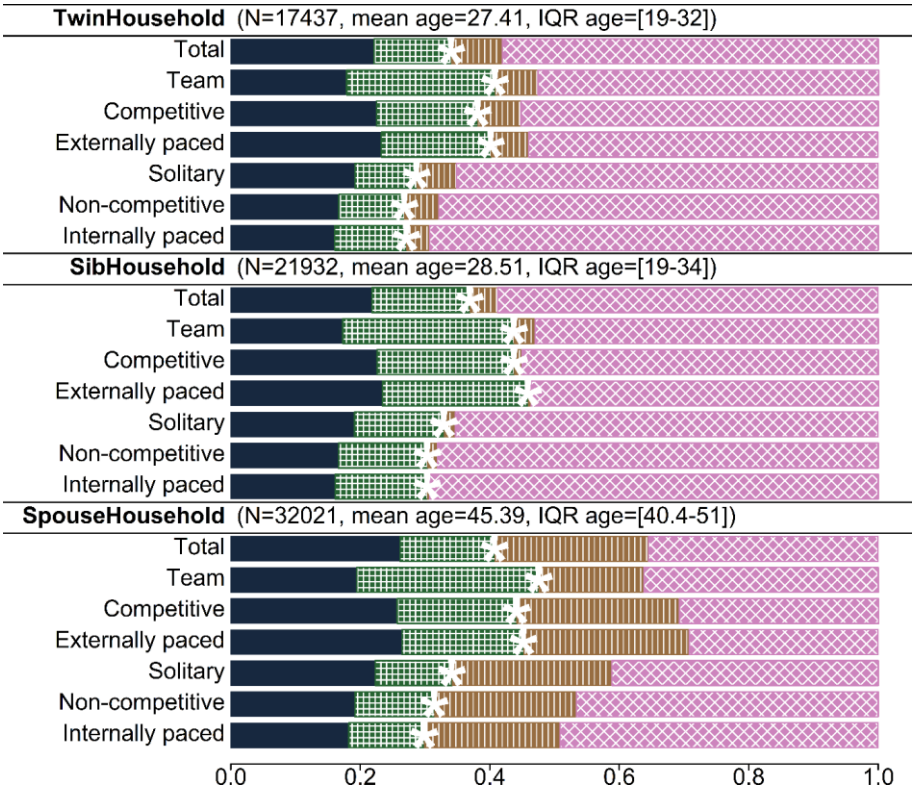
Supplementary Figure 2.11*. Tracking coefficients for volume of non-competitive exercise behavior at baseline and total volume of exercise at follow-up, as a function of age at baseline.*



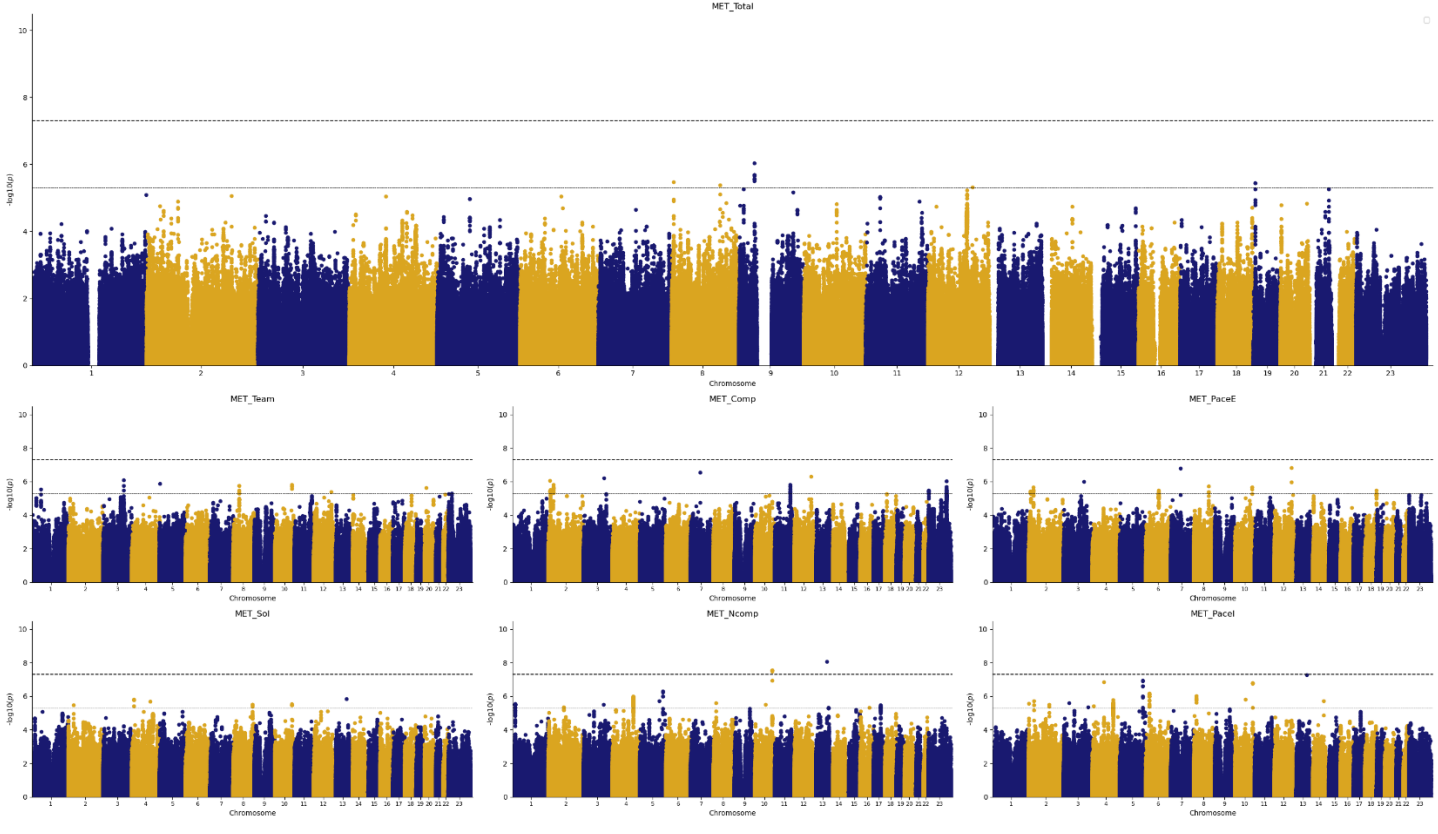
Supplementary Figure 2.12. *Tracking coefficients for volume of externally paced exercise behavior at baseline and total volume of exercise at follow-up, as a function of age at baseline.*



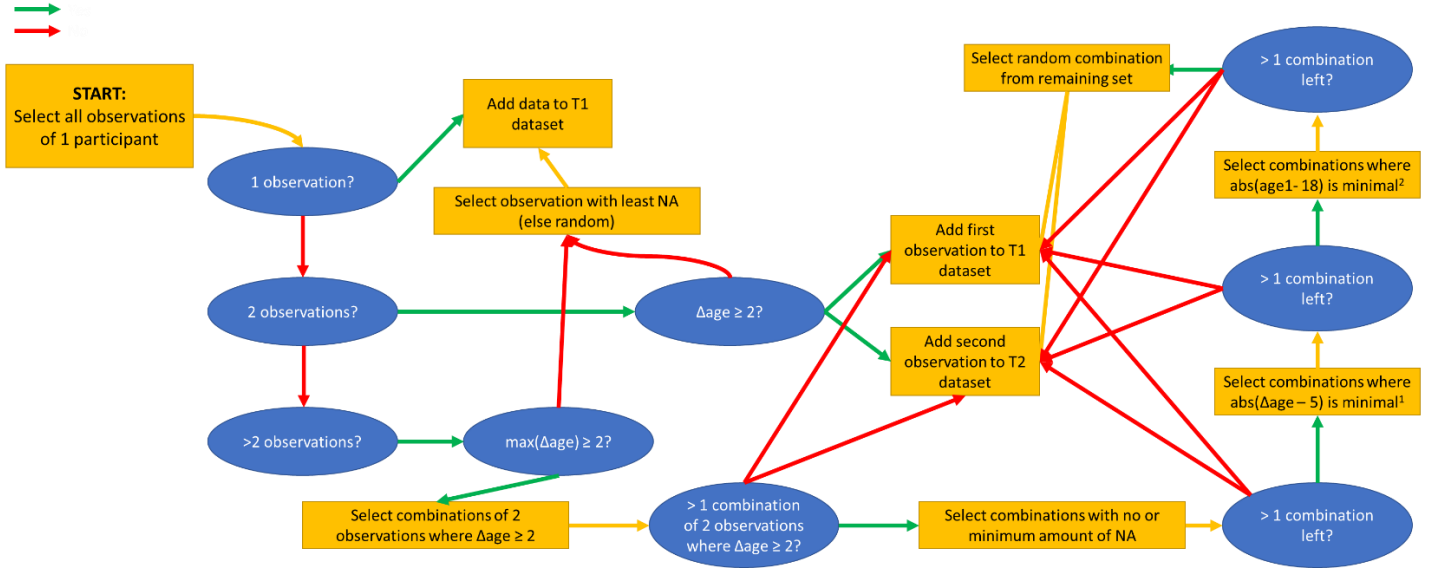
Supplementary Figure 2.13. *Tracking coefficients for volume of internally paced exercise behavior at baseline and total volume of exercise at follow-up, as a function of age at baseline.*



Supplementary Figure 5.1. *ACDE estimates for other household definitions: Proportion of variance explained by* ***additive genetic (A)****,* ***non-additive genetic (D)****,* ***common household environmental (C)****,* ***unique environmental (E)*** *factors and broad sense heritability(*) for different classes of exercise using various household definitions as shared environment.*



Supplementary Figure 6.1*. Manhattan plots of GWAS results from the various exercise domains. High-resolution version available online at https://drive.google.com/drive/folders/1i8lUQJcMmzj0QTPtGTxsFuTn8tpKlKZr*



Supplementary Figure 7.1. *Longitudinal data selection. [1]: Due to the biennial nature of the NTR surveys, abs(Δage – 5) will generally result in equal preference of surveys 4 and 6 years apart. [2]: Selects T1 closest to 18. High-resolution version available online at https://drive.google.com/drive/folders/1i8lUQJcMmzj0QTPtGTxsFuTn8tpKlKZr*

Dankwoord

To the majority of readers, welcome to my thesis. To those of you who have actually read everything and ended up here, thank you very much, I do hope you enjoyed the content. Just a heads up before I begin, this dankwoord will switch between English and Dutch often and mid-paragraph. I tried to create some semblance of structure in what's English and what's Dutch but to no avail.

Allereerst wil ik graag alle **deelnemers aan het NTR** ontzettend bedanken voor hun bijdrage. Dit proefschrift was niet mogelijk geweest zonder de voortdurende vrijwillige inzet van de vele deelnemers. Dus bij deze, aan alle deelnemers en iedereen die geholpen heeft bij de dataverzameling, ontzettend bedankt!

Next I would like to thank my promotion committee. **Peter Beek, Yannis Pitsiladis, Jaakko Kaprio, Hidde van der Ploeg** and **Nienke Schutte**. I would like to thank you all sincerely for all the time and effort you have spent on my thesis. **Nienke**, ik vind het een bijzondere eer dat je op deze manier aan het einde van mijn PhD betrokken bent, aangezien je ook aan de wieg van mijn wetenschappelijke carrière hebt gestaan. Jouw enthousiasme en motivatie heeft destijds een belangrijke rol gespeeld in mijn beslissing om een PhD te gaan doen, dus ook daarvoor heel erg bedankt.

I would also like to thank my co-promoters. **Erik Ehli**, thank you so much for your mentorship and guidance, especially during my time at AIHG. It has always been a joy for me to learn from your expertise. **Jenny van Dongen**, ook jij ontzettend bedankt voor alles wat je voor mijn proefschrift hebt betekend en bedankt voor alles wat ik van jou heb mogen leren tijdens mijn PhD, zowel over (epi-)genetica als over het goed begeleiden van studenten.

Uiteraard kan in mijn dankwoord ook mijn promotor, **Eco de Geus**, niet ontbreken. Beste Eco, ik kan jou nooit genoeg bedanken voor wat je voor dit proefschrift en voor mij persoonlijk hebt betekend. Voor je intensieve hulp en steun, bij de laatste loodjes in het bijzonder, ben ik je ongelofelijk dankbaar voor. Dit geldt eigenlijk evengoed voor het gehele project. Met jouw passie en interesse wist je mij keer op keer te motiveren voor nieuwe projecten, uitgebreide analyses en noem maar op. Dus voor alles: je hulp bij het schrijven, bij analyses, je openheid voor nieuwe ideeën, voor dat ik (na wat aandringen) toch Python mocht gebruiken, voor je steun en zo veel meer, ontzettend bedankt!

I would also like to thank everyone at the **Avera Institute for Human Genetics** for the continued collaboration throughout my PhD and for making me feel at home in Sioux Falls during my time there. **Ryan**, if it wasn't for your comforting words leading up to my first visit to AIHG, odds are none of this would have happened, so thank you very much for that and all the rest you have done to help make this PhD a reality. **John, Brandon** and **Austin**, thank you so much for the awesome collaboration throughout my PhD. I have enjoyed working with all of you very much and I'm looking forward to a continued collaboration in the future. Noah, your humor and helpfulness made my time at AIHG even more enjoyable than it already was, so thank you for making my time at AIHG even more fun. **Jeff**, though it pained me to be named 'New-Jeff' when I came back and you knew Wonu and Eshim better than me, I consider it a compliment. It has been awesome to work with you, but more importantly I consider you as a good friend and I am grateful for all the time we spent together, looking forward to much more. **Casey**, be it the microbiome lectures on a road trip, or beers and (Dutch) nachos, I have not only learned a lot from you, but more importantly I have enjoyed every second of our great friendship.

Of course, I would also like to thank all my colleagues at **BioPsy**, for all their help writing my thesis and equally important for their support throughout. **Meike**, jij diende als mijn "gateway-drug" voor de afdeling tijdens mijn bachelor state in 2013 en sindsdien ben ik er ook niet meer weggegaan. Dankzij jou zit ik hier en mede dankzij jou blijf ik hier ook. **Dennis**, om nog maar even bij mijn begin op de afdeling te blijven. Niet alleen was jij de beste teaser voor de afdeling in het bachelorvak, jouw vriendelijkheid en humor hebben ervoor gezorgd dat ik me snel op mijn gemak voelde op de afdeling. **Dorret**, zonder jou waren veel papers in dit proefschrift niet mogelijk geweest en was mijn fijne samenwerking met AIHG misschien ook wel nooit tot stand gekomen, dus ontzettend bedankt voor al je hulp! Over hulp gesproken, **Conor**, zonder jou had ik veel van de OpenMx en R dingen in dit proefschrift nooit kunnen doen, dus nogmaals ontzettend bedankt. **Jouke-Jan**, mijn hele these is gebouwd op genetische data en niets hiervan was mogelijk zonder jou. Niet alleen vanwege je werk aan MRG maar ook door de honderden keren waarop je mij hebt geholpen, dank daarvoor. Ditzelfde geldt overigens voor **René Pool**, niet alleen bedankt voor je hulp met PRS en data, maar het is vooral ook fijn om niet de enige Pythonista op de afdeling te zijn. **Michiel**, zonder jouw hulp was er nog veel meer dan alleen wat analyses niet gelukt, dus ontzettend bedankt! **Martin**, bedankt voor het zelfgemaakte kokosnootbier en de bijzonder fijne samenwerking tijdens Analyses Toolbox het onderwijs. **Michel**, dankzij jou ben ik op vele vlakken, zowel als wetenschapper als persoon, enorm gegroeid. Ontzettend bedankt daarvoor (en voor de PostDoc positie natuurlijk). En tot slot, **Natascha**, jij maakt de hele afdeling voor mij (en ik denk voor iedereen) altijd leuker, zelfs in corona tijden. Zonder jou en zonder de sfeer had ik me in al die jaren niet zo vermaakt en zonder jouw steun was deze PhD ook niet mogelijk, dus nogmaals ontzettend bedankt!

Wonu and **Eshim**, I am so very grateful to have had both of you as roommates from the very beginning, for support and just some good laughs, or a Pokémon-walk every now and then. **Lianne V**, ik kan je niet vaak genoeg bedanken voor alle hulp bij mijn laatste hoofdstuk, maar bij deze dus een poging. Daarnaast ook bedankt voor al je steun. **Perline, Denise, Camiel, Sofieke, Zenab**, as (former) roommates I would also like to thank all of you for your help and support throughout, even during Corona times. To all the other PhDs **Anne L, Bodine, Fiona, Floris, Laura, Nicole, Nikki, Selim, Susanne, Veronika** and **Zoey**, thank you all very much for the supportive PhD meetings, the encouraging words, and the help. En tot slot, **Margot**, voor al je steun van het begin tot eind, al het bier, dat je mijn paranimf wil zijn, pyramid M&M en zo veel meer, ontzettend bedankt!

Ook bedankt aan alle ex-collega's van BioPsy die de afdeling tijdens mijn PhD helaas hebben verlaten. **Pieter-Bas**, bedankt voor de tucjes en de bijzondere sta-bureau strategieën. **Anne H** en **Yayouk**, en Bart bedankt voor al jullie hulp en steun, jullie hebben mij niet alleen veel geleerd, maar mij ook door een groot deel van mijn PhD gesleept. **Laura**, ontzettend bedankt voor alle hulp, en dat je mijn drinking-buddy wilde zijn, helaas is deze traditie verder nooit een succes geworden. Tot slot, **Hill**, uiteraard bedankt voor alle hulp, maar belangrijker nog, bedankt voor de honderden keren dat jij anderen met R hebt geholpen omdat iedereen maar bleef geloven dat ik het niet kon.

Lianne R, heel erg bedankt voor alle gezelligheid, alle hulp en de oneindige steun van de master tot een PhD vinden tot het afsluiten. **Denise, Corné, Marijke** en **Mirjam**, bedankt voor alle steun vanaf de bachelor tot nu. **Isaac, Leroy, Michiel, Michiel, Tim** en **Rudy**, voor de gezelligheid, vakanties, alle potjes boerenbridge, hartenjagen, alle vakanties, en natuurlijk alle steun, heel erg bedankt. This gratitude for support and all the rest also extends to you, **Kyra** and **Valerija**.

Rowdy, niet alleen bedankt voor al het werk dat jij voor het ontwerpen van mijn dissertatie hebt verricht, maar ook heel erg bedankt voor alle steun. **Conny** en **Ron**, heel erg bedankt voor de roti, de gezelligheid en jullie gastvrijheid. **Marsha**, heel erg bedankt voor alle gezelligheid, voor alle steun en alle liefdevolle aanmoediging. En tot slot, **Tamara**, voor alles: dat je mijn paranimf bent, alles wat je voor de verdedigings-dag hebt geregeld, de gezelligheid, de steun -of het nou voor mijn PhD was of iets anders-, onze goede vriendschap en nog zo veel meer ontzettend, ontzettend bedankt!

Babette, ook jij heel erg bedankt voor alle steun en alle liefde. **Oma**, nogmaals bedankt voor alles, het heeft dan wel iets te lang geduurd, maar ik ben eindelijk klaar met studeren. **Mama**, waar moet ik beginnen... Zonder jouw hulp en steun, en soms aandringen, had ik mijn VWO waarschijnlijk niet eens gehaald, laat staan deze PhD. Lieve mama, voor alles wat je voor mij gedaan en betekend hebt (en dat is nogal wat) ontzettend bedankt!

En tot slot, **Myrthe**, ik heb het je, zeker in het laatste stuk, niet altijd makkelijk gemaakt, maar ondanks dat ben je me altijd blijven steunen, stimuleren en opvangen. Daarvoor, en voor alle liefde, hulp en andere steun, enorm bedankt!