Chapter 9

SUMMARY AND GENERAL DISCUSSION

This thesis used genetically informative designs to study the determinants and correlates of regular voluntary exercise behavior in children (ages 7, 10, 12), adolescents (ages 14 and 16) and adults (18+). In the first part of this thesis, the relative contribution of genes and the environment to individual differences in exercise behavior, including their contribution to stability over time and gene-by-environment interaction, was examined in twins aged approximately 7 to 18 years. The second part examined the nature of the association between exercise behavior and two commonly studied correlates, namely 1) the perceived benefits of and barriers towards exercising and 2) body mass index (BMI). The third part aimed to shed light on the molecular basis of the heritability of exercise behavior by means of a candidate gene study with genetic variants that play a role in the dopaminergic system. In this final chapter, the results of each study will be summarized in the order of appearance, followed by a discussion of their main implications and suggestions for future research. In addition, the fundamental assumptions of the classical twin design, exercise "omics", "deep" phenotyping of exercise behavior and the placement of exercise behavior into the broader context of "physical activity" are discussed. Finally, some overall conclusions will be drawn.

HERITABILITY OF REGULAR VOLUNTARY EXERCISE BEHAVIOR

Previous studies of the NTR have shown that both genes and the environment contribute to individual differences in exercise behavior both in adolescence (Boomsma et al., 1989; de Geus et al., 2003; de Moor et al., 2011; Koopmans et al., 1994; Stubbe et al., 2005; van der Aa et al., 2010) and adulthood (de Geus et al., 2003; de Moor & de Geus, 2013; de Moor et al., 2011; Stubbe et al., 2006; Stubbe & de Geus, 2009; Vink et al., 2011). Chapter 3 contains the first twin study on exercise behavior in childhood. Univariate ACE models were fitted to data of twins aged 7 (N= 3,966 individuals), 10 (3,562) and 12 years (8,687). With the exception of 10-year-old boys (A= 66%, C= 25%), most of the variance in exercise behavior could be explained by shared environmental factors (C= 50-72%). There were significant qualitative sex differences for the ages 7 and 12. At age 10, a drop in heritability was reported, which was likely a result of random fluctuation, as it had disappeared in a new and more powerful analysis of this age group in Chapter 4.

On several accounts, Chapter 4 is an extension of Chapter 3. It is based on a larger dataset and more age groups, namely on data of twins aged 7 (N= 7,331 individuals), 10 (8,007), 12 (14,629), 14 (9,030), 16 (6,019) and 18 years (2,759).

This allowed us to carry out longitudinal analyses on the same twins. As in Chapter 3, we found that the relative contribution of genes increased and the relative contribution of the shared environment decreased from childhood to adolescence. By also estimating the absolute variances, we noted that the absolute shared environmental variance was largely stable across ages, whereas the absolute genetic variance strongly increased. The phenotypic correlations across surveys were moderate to high with larger correlations for surveys that were in closer proximity to each other. In males, stability was mainly caused by transmission of the same genetic effects across ages with an increasing impact on the total variance. In females, genetic effects were transmitted from previous time points as well, but genetic innovation also contributed consistently to the increase in genetic variance of exercise behavior with age. Shared environmental effects were dominated by innovation in both males and females, meaning that different aspects of the shared environment played a role at different ages.

It is striking that the relative contribution of genes and the environment to exercise behavior changes vastly across childhood and youth. The generally large influence of the shared environment in childhood has also been found in previous small-scale studies on total physical activity measured with accelerometers (Fisher et al., 2010), respiratory gas exchange and doubly labeled water (Franks et al., 2005), and pedometers (Plomin & Foch, 1980). The relatively large heritability in adolescence has been found in data of the NTR (e.g., van der Aa et al., 2010) and other cohorts (Beunen & Thomis, 1999; Maia et al., 2002), especially in males. Most recently, Aaltonen et al. (2013) found heritability estimates of around 44-52% in approximately 16- to 18-year-old twins, with a shared environmental influence of 19-26%. It should be noted that in many studies (e.g., Beunen & Thomis, 1999; van der Aa et al., 2010), the shared environmental components were dropped when they were non-significant, meaning that any familial variance was modelled as genetic variance. The C-component is relatively hard to pick up with classical twin studies (Neale & Cardon, 1992; Posthuma & Boomsma, 2000; Visscher, Gordon & Neale, 2008), however, and non-significance of this component does not necessarily mean that it is absent. This is illustrated by Chapter 4 that reports somewhat lower heritability estimates compared to our earlier analyses of these age groups (e.g., van der Aa et al., 2010). This is because we decided, in contrast to our previous analyses, not to drop any components but to model the full ACE model.

Three overall conclusions can be drawn based on Chapter 3, Chapter 4 and

similar studies: 1) There are large individual differences in exercise behavior across the ages of 7 to 18 years, 2) these are mainly due to shared environmental factors in childhood and 3) mainly due to genetic factors in adolescence. In Chapter 3, we hypothesize that the most important shared environmental factor underlying exercise behavior in children might be parental support (which is for a part qualitatively different for boys and girls). This might gradually give way to the impact of peers and the school environment. De Moor et al. (2011) indeed show that generation specific environmental factors largely explain the shared environmental component in adolescence. The shared environmental component, however, is largely overwhelmed by genetic effects in this age group. The underlying factors driving heritability across childhood and adolescents are more stable than the shared environmental factors, but still there is innovation across ages, especially in girls, suggesting that there is no single set of genes that drives heritability in this age range. We suggested in Chapter 4 that exercise ability and trainability, the acute psychological response to exercise and a homeostatic need to be active or personality factors might underlie the high heritability of exercise behavior in adolescence (also see Rowland, 1998, and Eisenmann and Wickel, 2009). It is reasonable to assume that the contribution of these factors differs for younger versus older individuals. For instance, the homeostatic need to be active might be more relevant in children (Saudino, 2012), whereas exercise ability might play a major role in adolescents.

An important limitation of the twin studies on exercise behavior that have been conducted so far, is that they have not explicitly modeled gene-by-environment (GxE) interaction with exercise behavior as the outcome variable. Genes and environmental effects, however, may not act in splendid isolation and the expression of genetic variance may depend on the environment (Purcell, 2002). A more facilitating environment, for instance, might increase genetic variance, whereas a more restrictive environment might suppress genetic effects. In order to shed some light on potential effects of moderating variables, Chapter 5 assesses the effects of parental education on the means, total variances and variance components of children's exercise behavior in twins of the NTR and two Finnish cohorts. Consistent trends of a higher level of exercise behavior in children of high educated parents in both datasets point towards a role for parental education in offspring exercise behavior, which is in line with much previous research (Ferreira et al., 2006; Hanson & Chen, 2007; Singh et al., 2008). The tendency for lower total variances in this group has rarely been explored before (Johnson et al., 2010). In our study, we find only weak evidence for moderation of genetic or environmental variance of exercise

behavior by parental education. The lower genetic variance in exercise behavior in Dutch daughters of high educated parents suggests that genetic effects that would act against exercise behavior might be suppressed in this group.

Taken together, our studies highlight large individual differences in exercise behavior and a complex shift in the genetic and environmental determinants of exercise behavior from childhood to adolescence. This hopefully gives a better understanding of why there is no intervention that works for everyone. Interventions should be tailored to the specific needs of individuals and take into account causes of individual differences across ages and sex. One size surely does not fit all.

CAUSALITY TESTING USING TWIN DATA

Researchers have long been interested in correlates of exercise behavior, both as potential determinants and consequences of the behavior. Twin studies can shed more light on the nature of an association between two traits by assessing whether there could be a causal relationship or not. De Moor and colleagues (2008) have outlined and applied methods to test causality both in cross-sectional and longitudinal twin data. We have investigated the association of exercise behavior with a frequently studied potential determinant, namely the perceived benefits of and barriers towards exercise activities ("attitudes"), and a potential consequence, namely BMI.

Many social cognitive models of health behavior propose exercise attitudes to be important *determinants* of exercise behavior and highlight them as promising targets for interventions (Ajzen, 1985; Becker, 1974; Biddle & Nigg, 2000; Hagger et al., 2002; King et al., 1992; Schwarzer, 1992). In Chapter 6, we aimed 1) to unite this with the finding that exercise behavior is a heritable trait by showing - for the first time - that the perceived benefits of and barriers towards exercise behavior (summarized under the heading "exercise attitudes") are heritable themselves and 2) to test whether exercise attitudes and exercise behavior could be causally related. We ran a principal component analysis on questionnaire items assessing the perceived benefits of and barriers towards exercising in adult twins. Six components emerged, namely "Perceived benefits", "Lack of skills, support and/or resources", "Time constraints", "Lack of energy", "Lack of enjoyment" and "Embarrassment", and we showed that all of these were heritable, with heritability estimates ranging from 21% to 49%. Bivariate models revealed that the phenotypic correlations between the

attitude components and exercise behavior were all significant and ranged between -0.44 (“Lack of enjoyment”) and +0.32 (“Perceived benefits”), and together the attitude components explained 28% of the variance in exercise behavior. Moreover, the largest part of these correlations was due to overlapping genetic factors between the two traits.

It was tested whether the traits could be causally related by two means: bivariate genetic models and the monozygotic (MZ) twin intrapair differences model. The two approaches are fully described in Chapter 6 and will shortly be outlined in the following. It should be noted that both approaches make it possible to falsify, but not to prove, causality.

The rationale behind testing for causality with a bivariate genetic model is that if exercise attitudes causally influence exercise behavior, then everything that influences those attitudes will also, through the causal chain, influence exercise behavior (if A causes B, and B causes C, then A causes C). Therefore, if exercise attitudes are affected by both genes and the environment, then the genetic and environmental cross-trait correlations between the attitude components and exercise behavior need to be significant under the assumption of causality. Based on this rationale and the finding of significant genetic and environmental correlations in the bivariate models, a causal relationship could not be falsified for most attitude components in Chapter 6, with the exception of “Perceived benefits” and “Embarrassment” in males, where the unique environmental correlations were non-significant. However, the unique environmental factors contain both true environment, which is expected to be correlated between two causally linked traits, and measurement error, which is not expected to be correlated. The “true” unique environmental correlations might be very small and thus might have been non-significant due to a lack of power.

The results of the bivariate models were fully confirmed in a second approach to causality testing, the MZ twin intrapair differences model. If there is a causal association between attitudes and behavior, the twin with more positive attitudes should exercise more compared to his or her genetically identical co-twin with more negative attitudes. Therefore, within-pair differences in exercise attitudes should be associated with within-pair differences in exercise behavior. The within-pair difference scores were calculated for the attitude components and for exercise behavior in MZ twins. Next, the difference score of each attitude component was correlated with the difference score of exercise behavior. Significant correlations are compatible with a causal effect and these were found for 10 of the 12 correlations (6 attitude components in

both sexes). Non-significant correlations would imply that the phenotypic associations are caused by underlying genes influencing both phenotypes in the absence of causality ("genetic pleiotropy"). Again, this was found for "Perceived benefits" and "Embarrassment" in males only.

Overall, the results revealed that 1) exercise attitudes are heritable, 2) exercise attitudes are significantly related to exercise behavior and 3) it is likely that this relationship is causal. As this study was based on cross-sectional data, we could not draw any conclusions on possible directions of causation, however. The heritably of attitudes reveals that no one is born as tabula rasa when it comes to the future perceived benefits of and barriers towards exercise behavior. There are innate differences that make individuals more or less likely to have a positive attitude towards exercising. This makes intuitive sense. As stated above, personality factors, in part through a link with a homeostatic need to be active or the acute psychological response to exercise, as well as exercise ability and trainability are hypothesized to underlie the heritability of exercise behavior. Attitudes might be closely related to these. For instance, "Embarrassment" and "Lack of skills, support and/or resources" are probably related to exercise ability and trainability, "Lack of enjoyment" and "Perceived benefits" to the acute psychological response to exercise, "Lack of energy" to a homeostatic need to be active and "Time constraints" to personality factors such as neuroticism. Health promotion strategies often aim to change the populace's attitudes towards exercise behavior by educating people on the health benefits of regular exercise and ways to reduce barriers to engage in exercise activities. Although the largest part of variance in exercise attitudes was explained by the non-shared environment, such strategies should take into account innate differences between individuals. It does not make sense to try to convince someone that exercising will make him or her "feel energetic" when this effect does not apply to that specific person. This is all speculative, however, and replication in longitudinal studies is first needed to more firmly establish the direction of causality.

Next, we aimed to apply the same causality testing approach to investigate a very intensively studied potential *consequence* of exercise behavior, namely BMI. One of the main rationales for research on exercise behavior is that it has an effect on body weight and could therefore be a way to curb the obesity epidemic. A wealth of studies on the relationship between energy expenditure and body composition in childhood and youth has focused on daily physical activity, with rather mixed outcomes (Bleich et al., 2011; Jiménez-Pavón et al., 2010; Must & Tybor, 2005; Wareham et al., 2005; Wilks et al., 2011). The

majority of studies have methodological flaws such as a cross-sectional design that prohibits conclusions on cause-effect relationships, small sample sizes and suboptimal measurement instruments (Must & Tybor, 2005; Wareham et al., 2005). Systematic evidence for the often cited link between daily physical activity, let alone exercise behavior, and body composition in young people is still lacking.

In Chapter 7, we aimed 1) to test the widespread assumption of the presence of a significant negative association between exercise behavior and BMI and 2) to investigate in how far this association reflects a causal effect of exercise behavior on BMI. In contrast to many previous studies, our study was based on a large longitudinal dataset with detailed measures of exercise behavior. Based on the hypothesis that regular exercise behavior is a causal determinant of obesity, higher levels of exercise behavior in childhood and adolescence were expected to be associated with lower levels of BMI at all ages and changes in exercise behavior with time were expected to predict opposite changes in BMI.

Contrary to our expectations, we found no evidence for a cross-sectional or longitudinal association between exercise behavior and BMI at all in a (partly) longitudinal dataset of 7-, 10-, 12-, 14-, 16- and 18-year-old individuals. Alternative determinants of BMI such as basal metabolic rate, other aspects of daily physical activity and sedentary behavior, but prominently also energy intake are likely to be more important, meaning that weight loss programs that are based on increasing regular exercise only, i.e., without an accompanying dietary intervention, may not be successful. Importantly, this does not detract from the value of encouraging regular exercise behavior in childhood and youth, as this behavior has been shown to have many other favorable effects on health, even in the absence of an effect on body weight, and should thus still be promoted (Melanson et al., 2013). Claiming a primary role for exercise behavior in the variation of children's and adolescents' BMI, however, may foster false expectations. No one should be led to believe that every person can lose significant amounts of weight by exercising. It is better to stress the multiple potential health benefits of this behavior, such that even in the absence of a change in weight, advantageous changes may still take place in other risk factors.

More generally spoken, our null finding should remind us of the importance of understanding innate individual differences in the response to exercise before applying interventions (Bouchard et al., 2015; de Geus et al., 2014). The HERITAGE study and related work have brought about important findings on

the effects of training on a variety of fitness traits such as maximal oxygen uptake or skeletal muscle strength (Bouchard, 2012; Bouchard & Rankinen, 2001; Bouchard, Rankinen & Timmons, 2011). These kinds of studies will improve the ultimate effectiveness of interventions by making it possible to target so-called "responders" - individuals who will actually benefit. For instance, it does not make sense to motivate an overweight individual to exercise on a regular basis *in order to lose weight* if that individual's genetic makeup is not sensitive to such effects of exercise. Although currently not yet feasible, it would ultimately be more meaningful to focus on the health benefits that can be realistically expected given an individual's genetic makeup.

The increasing availability of DNA data for a large number of individuals makes it possible to not only test causality based on latent genetic variance components, but to directly test causality based on the effects of specific genes. The *Mendelian randomization technique* has gained popularity in the past years as a means to causality testing that is in principle similar to the approach that we have used (Davey Smith & Ebrahim, 2004; Davey Smith & Hemani, 2014; Lawlor, Harbord, Sterne, Timpson & Davey Smith, 2008). Instead of calculating the correlation between latent genetic and environmental factors that are thought to influence two traits, it is based on measured genetic variants. More specifically, a genetic variant that influences an exposure variable (such as exercise behavior) should also, through the causal chain, predict an outcome variable (such as BMI). The application of Mendelian randomization with exercise behavior as a predictor of any phenotype is challenging, as solid associations with genetic markers would first have to be identified, which is currently not the case (see p.190). However, the technique can already be applied to test potential reversed causal effects, for instance with BMI as the predictor. Large international consortia have yielded genetic risk scores for BMI that can be used as genetic instrumental variables. Richmond et al. (2014) found evidence that BMI causally influences physical activity in 4,296 children aged 11 years by regressing an allelic risk score for high BMI ("the genetic instrument") on physical activity as assessed by accelerometers. They also tried to test whether physical activity had a causal effect on BMI but found no evidence for this effect. In all fairness, the instrumentation of physical activity, namely a genome-wide prediction score based on their own study, provided a very weak genetic instrument and the authors indicate that their results should be interpreted with caution.

The big advantage of the Mendelian randomization technique is that it is based on measured genetic variants and can be applied to any large population-based

samples, whereas our methods for causality testing rely on latent (unmeasured) genetic and environmental factors and need large twin samples. A shortcoming with studies that are based on twins only is that the shared environmental variance (C) cannot be estimated simultaneously with the dominant genetic variance (D). Therefore, one of the two is usually selected and included into genetic models, based on twin correlations and/or fit statistics. When the dizygotic (DZ) twin correlations are larger than half the MZ correlations, shared environmental influence is assumed, whereas a lower DZ twin correlation compared to half the MZ correlation implies dominant genetic effects. However, in reality, both might be in place at the same time and it is even possible that they cancel each other out in the twin correlations. Therefore, a model estimating all four components (A, E, C and D) - which is only possible when including for instance data of parents - would be especially desirable in the context of causality testing using twin data.

In sum, we have shown that attitudes are heritable and that they might be causally related to exercise behavior, but we found no evidence for a (causal) association between exercise behavior and BMI. The genome-wide association era has yielded novel and powerful approaches to test for causality between traits and these should fully be exploited to better understand the nature of the relationship between exercise behavior and relevant correlates. Most importantly, researchers should not underestimate the relevance of studying individual differences in the response to exercise as targeting non-responders might not only lead to disappointment of the participants, but is actually not the optimal use of scarce public health resources.

ASSUMPTIONS OF THE CLASSICAL TWIN DESIGN

I hope that the outlined chapters have conveyed the beauty and versatility of twin studies to the reader. A number of critical assumptions have to be met, however, to obtain valid results. Most obviously, it is assumed that *MZ twins share all of their segregating genes, whereas DZ twin share on average 50% of their segregating genes identity-by-descent (IBD)*. Although there are MZ twins that are not entirely genetically identical, no systematic differences between the DNA sequences of MZ twins have been found so far (Baranzini et al., 2010; Veenma et al., 2012). The question of genetic similarity between DZ twins has been answered based on genome-wide marker data. In a sample of 11,214 sibling pairs, Visscher et al. (2007) have shown that the true proportion of IBD sharing lies within the range of 31% and 64%, with a mean of 50%. They calculated the heritability of height based on the empirical IBD sharing between

siblings and came to virtually the same heritability as previous twin studies. Thus, the assumption of the amount of genetic overlap between MZ and DZ twins is very likely met, at least under the *assumption of random mating*.

Non-random (or assortative) mating refers to spousal resemblance on a phenotype. Previous studies have found spousal correlations on exercise behavior-related traits ranging from 0.16 to 0.60 (Aarnio et al., 1997; Boomsma et al., 1989; Perusse et al., 1988; Perusse et al., 1989; Seabra et al., 2008). There are several possible mechanisms causing significant spousal correlations, including phenotypic assortment, social homogamy and social interaction (de Moor et al., 2011; Heath & Eaves, 1985). The most problematic of the three mechanisms for twin modelling is phenotypic assortment, meaning that initial partner selection is based on the phenotype under study (or a correlated phenotype), e.g., when exercisers are attracted to other exercisers. If genes are implicated in the phenotype, this would inflate the DZ twin correlation and thus result in an overestimation of shared environmental effects if assortment is not taken into account. Alternatively, spouses might resemble each other simply because individuals from similar social backgrounds are more likely to meet and might therefore start a relationship (social homogamy). Finally, spouses might not necessarily be alike in the first place, but might become more similar in the course of their relationship (social interaction). De Moor et al. (2011) were the first to explicitly investigate these three mechanisms in the context of exercise behavior using twin models and concluded that the observed spouse resemblance was best explained by phenotypic assortment. Unfortunately, a rather rough measure of exercise behavior (namely the dichotomy “regular exerciser” versus “non-exerciser”) was used. A replication of their results based on a more precise measure of exercise behavior would be desirable and would imply that future studies might want to correct for assortative mating - although if different (uncorrelated) genes determine exercise behavior at different points in life, this might not be necessary or even make any sense. Assortment would then only affect genes that influence exercise behavior in early adulthood (when most individuals select their mating partners). Applying any corrections was deemed to be premature in our studies, as the nature of assortative mating should first be solidly established.

The so-called *equal environments assumption* (EEA) has been subject to heated debates (e.g., Horwitz, Videon, Schmitz & Davis, 2003). It is well known that MZ twins tend to have more similar environments than DZ twins. For instance in childhood, MZ twins are more likely to share friends, share the same room and to dress alike. In adulthood, MZ twins often have a higher contact frequency

than DZ twins (Kendler, 1993). The EEA posits that these environmental differences are not related to the phenotype under study. Otherwise, a higher similarity of MZ twins compared to DZ twins could be due to genetic influences, environmental influences, or both, whereas the classical twin design ascribes a difference in similarity to genetic factors only. The EEA has been shown to be met for a wide range of phenotypes (Kendler, 1993).

One common test of the EEA is based on assessing indices of the amount of shared environment in twin pairs (such as perceived similarity of treatment by others, similarity of appearance or contact frequency) and to test their impact on heritability estimates. The first explicit test of the EEA for doing sports during leisure time and related physical activity phenotypes (Eriksson et al., 2006) investigated the effect of twins' contact frequency on their phenotypic similarity. Heritability was calculated in twins with high versus low contact frequency and higher estimates were found for the former group. This might be interpreted as a violation of the EEA. Eaves, Foley and Silberg (2003) challenge this interpretation, however, by proposing that genetic factors might drive niche selection such as the choice of with whom twins spend their time. Thus, so-called "environmental" differences might actually have a genetic origin and might thus not constitute a violation of the EEA. Based on simulated data, Eriksson et al. (2006) show indeed that the observed patterns could entirely be explained by niche selection.

A different approach to testing the EEA is based on the fact that there are twin pairs that misperceive their zygosity (e.g., Conley, Rauscher, Dawes, Magnusson & Siegal, 2013). Twins might think that they are monozygotic in the first place and then a closer examination (e.g., a DNA test) reveals that they actually are dizygotic or vice versa. Heritability can be estimated based on "real" versus "perceived" zygosity. If heritability estimates are higher in the latter case, this would imply that monozygotic twins are more similar to each other than what would be expected based on the difference in genetic relatedness with DZ twins alone, and that the EEA would thus not be met. To the best of my knowledge, this has not been tested for exercise behavior or other physical activity-related traits.

Thus, the EAA has been shown to be met for a wide range of phenotypes, but explicit tests in the context of physical activity are scarce. As suggested in Chapter 6, it would be particularly interesting to examine whether treatment of twins by others is more similar with a higher resemblance in athletic appearance and how this relates to heritability estimates in traits that are

associated with physical activity. One might argue that MZ twin correlations could be inflated as MZ twins are treated more equally due to how others perceive their athletic capabilities (based on their appearance), independent of their actual physical abilities.

Last but not least, twin research is based on the fundamental assumption that *twins are representative of the general population*. Testing this assumption is not only relevant to researchers, but also to the twins themselves. As twins are born in all strata of society, there is no reason to assume a systematic bias in the first place. Pregnancy and birth outcomes are often less favorable for twins than for singletons, however (Croft, Morgan, Read & Jablensky, 2010). Twins tend to be delivered preterm and they have lower birth weights (Croft et al., 2010; Estourgie-van Burk, Bartels, Boomsma & Delemarre-van de Waal, 2010). In a Dutch sample, Estourgie-van Burk et al. (2010) showed that compared to the general population, twins were smaller and weighted less on average at birth and around their first birthday. At the age of 4 years, they were not significantly different from the general population in height and weight, but they had somewhat smaller BMIs. Any differences in height, weight or BMI had disappeared at the age of 18 years (Estourgie-van Burk et al., 2010). Differences between twins and singletons in exercise behavior have - to the best of my knowledge - not been tested systematically before, although there is no reason

TABLE 1 Percentage of non-exercisers and mean weekly MET hours (variances; N) for multiples versus singletons.

		% non-exercisers*		Weekly MET hours		
Age	**Sex**	**Multiple**	**Singleton**	**Multiple**	**Singleton**	**p****
13	**M**	20.4%	17.6%	28.7 (24.9; 314)	29.7 (24.6; 119)	.77
	F	22.7%	19.1%	20.5 (23.1; 423)	19.2 (18.8; 131)	.82
14	**M**	19.1%	28.3%	29.7 (26.6; 1282)	26.7 (25.8; 60)	.36
	F	21.1%	19.3%	21.4 (22.8; 1688)	23.6 (24.0; 88)	.40
15	**M**	20.1%	22.3%	32.1 (28.5; 528)	27.8 (24.5; 103)	.28
	F	25.2%	23.7%	21.3 (23.3; 686)	24.2 (26.1; 131)	.39
16	**M**	24.5%	24.3%	31.1 (29.3; 912)	30.8 (30.4; 210)	.70
	F	28.3%	26.3%	21.2 (23.9; 1292)	21.3 (25.3; 262)	.84
17	**M**	25.0%	30.1%	31.5 (29.6; 520)	30.6 (29.4; 206)	.75
	F	32.4%	30.4%	19.4 (21.9; 720)	19.6 (23.7; 273)	.91
18	**M**	25.8%	28.8%	31.4 (32.5; 62)	27.0 (29.1; 208)	.44
	F	41.8%	37.6%	16.3 (21.7; 122)	15.6 (19.3; 340)	.79

**<4 weekly MET hours; **p-value of comparing weekly MET hours according to Mann-Whitney U test.*

to assume noteworthy differences in terms of physical capability or exercise motivation. Table 1 depicts a comparison of exercise behavior in multiples versus singletons based on data of the NTR (taken from Chapter 4). First-born multiples and siblings were selected in narrow age ranges (e.g., for “age 13”, they were >=13.0 and <14.0 years old) to compare the percentage of non-exercisers and the means and variances in weekly MET hours between these groups. No systematic differences were apparent.

To sum up, there is no reason to assume that the general assumptions of twin research are not met with regard to exercise behavior.

GENOMICS OF REGULAR VOLUNTARY EXERCISE BEHAVIOR

Chapter 8 contains the largest candidate gene study so far on dopaminergic variants and exercise behavior during leisure time both in terms of the number of genetic variants that were included and in terms of sample size. None of the variants were significantly associated with exercise behavior. Even when looking beyond dopaminergic genes, not a single genetic variant has been shown to affect regular exercise behavior at a level of “proof beyond reasonable doubt” in previous studies. Technological advancements make it possible to not only test a handful of genetic markers for their association with a phenotype, but to test the association with hundreds of thousands of markers simultaneously, covering genetic variation across the whole genome. As theory-based candidate gene studies have not been proven successful, such theory-free genome-wide association studies are the appropriate way forward as they provide the opportunity to discover entirely new pathways (Flint, 2013; Pearson & Manolio, 2008). Due to simultaneous testing of a very large number of genetic markers, many of these markers will be significantly associated with the phenotype merely by chance if not correcting for multiple testing, however (Sullivan, 2007). At the same time, exercise behavior is a quantitative trait that is influenced by many genes with very small effects. With a significance threshold that is corrected for multiple testing (e.g., $\alpha = 5x10^{-8}$), it needs a very large number of individuals to find any genome-wide significant associations at all and to confirm significant hits in independent samples. Although costs of genotyping are lower than ever before, genotyping DNA data of hundreds of thousands of individuals remains a very expensive undertaking, let alone the costs of (both genotypic and phenotypic) data collection and processing. No single research group has the necessary resources to acquire big enough sample sizes to push the field forward from where we are now. It is essential to establish an international consortium that pools data of cohorts with genome-

wide DNA data and corresponding data on exercise behavior (or, more generally, physical activity phenotypes). Such efforts are currently undertaken and the first large-scale GWAS by the GIANT consortium is underway.

MORE OMICS OF REGULAR VOLUNTARY EXERCISE BEHAVIOR

So far, I have investigated and discussed the starting point of the biological paths towards exercise behavior, namely variation in the genome. Although genes are a straightforward beginning to disentangle the biology of behavior, it would clearly be an oversimplification to ignore in this discussion the very complex array of processes from the genetic code to observable behavior that impact upon one another in a dynamic and hardly predictable fashion.

Every cell in our bodies contains the same genetic information. Differences between cells emerge as only genes that are relevant to each specific cell are transcribed into RNA. The basic process leading from genotype to phenotype is well-known. Simply put, DNA is first transcribed to RNA, which in turn is translated to or impacts upon proteins, the basic biological building blocks of phenotypes. Unfortunately, however, there is no one-to-one transformation from genes to proteins, let alone from proteins to behavior. Individual differences in these processes make the interpretation of genetic effects a very complex undertaking. Also, there are feedback mechanisms going from the phenotype back to RNA transcription. Knowing a person's genetic makeup is thus far from knowing the biological origin of that person's behavior. Therefore, epigenomes (Kaminsky et al., 2009; van Dongen et al., 2014), transcriptomes (Jansen et al., 2014; McRae et al., 2007; Tan et al., 2005), proteomes (Altelaar, Munoz & Heck, 2013) and metabolomes (Draisma et al., 2013; Draisma et al., 2015; Dunn et al., 2011; Gieger et al., 2008; Nicholson et al., 2011) are increasingly being studied.

Animal models are a popular means to unravel the underlying "omics" events that connect genes with behavior. Functional annotation studies are often conducted in animals and aim to confirm GWAS-derived results by assessing the effects of genetic variants on, for instance, proteins or metabolites that are relevant to the phenotype under study. Obviously, there are less complex ethical constraints in animal research compared to studies with human subjects, making it possible to study omics in all tissues (including, for instance, brain tissue) and to manipulate the suspected genotypes or the environment in order to elicit changes in the transcriptome, the proteome, the metabolome and/or the ultimate behavior.

Mice are especially suited to study the determinants and consequences of exercise behavior, mainly for three reasons. First, their genome is comparable to that of humans (Paigen, 2003). Second, voluntary wheel running serves as an elegant model for voluntary exercise behavior. And third, rodents have a relatively short lifespan which makes them well suited to study long term health consequences of exercise behavior and/or effects of ageing processes (de Geus et al., 2014). Mice models have confirmed the important role of genes in exercise behavior. In genetically well-characterized inbred strains of mice, larger between-strain differences compared to within-strain differences in running wheel activity became apparent (Lightfoot et al., 2010). In addition, mice can be selectively bred for high voluntary wheel running (Rezende, Gomes, Chappell & Garland, 2009). In line with our hypotheses, various studies in rodents have suggested that ability and motivation underlie differences in voluntary exercise behavior (Garland et al., 2011).

A CALL FOR "DEEP" PHENOTYPING OF EXERCISE BEHAVIOR

To improve the probability of finding significant associations with genetic markers, it is important to have a clearly defined phenotype. We have deliberately focused on regular exercise behavior during leisure time as it can be measured reliably by survey. However, this phenotype might still be defined too broadly as we collapse all kinds of exercise activities during leisure time. It is reasonable to assume, however, that the determinants of doing strength training are qualitatively different from those of playing basketball, for instance, which is likely to be reflected in different biological origins. A better understanding of individual differences as they relate to the choice for a *specific* exercise activity is therefore an important consideration. Relating these narrow and refined phenotypes (e.g., individual sports versus team sports, competitive exercise versus non-competitive exercise) to genes might delineate a more accurate picture than the somewhat broader approach that we have chosen. Lauderdale et al. (1997) have assessed and separately analyzed specific exercise activities (running, bicycling, swimming, racquet sports and other strenuous sports) in twin pairs of the Vietnam Era Twin Registry. Their results indeed show that heritability estimates differ between activity types. They also suggest that more strenuous activities might be more heritable than moderate activities.

Moreover, the motivational precursors of exercise behavior might be fundamentally different across individuals (e.g., performing an activity mainly for social reasons versus for health reasons, reward sensitivity, mood

responses), which again might be reflected in genetic effects on very different biological systems (Bryan et al., 2007). Most importantly, these precursors might be different across age and sex. For instance, competitive exercise ability might be more relevant to adolescents, whereas health benefits might be more relevant to adults. In addition, females might be more inclined to exercise for losing weight, whereas males might want to build up muscles.

PLACING EXERCISE BEHAVIOR IN THE BROADER CONTEXT OF "PHYSICAL ACTIVITY"

This thesis has not examined general physical activity or sedentary behavior ("any waking behavior characterized by an energy expenditure ≤1.5 METs while in a sitting or reclining posture", Sedentary Behaviour Research Network (2012), p.540). This should not detract from the value of overall physical activity to increase health in the general population or from the detrimental effects of sedentary behavior. Exercise interventions are probably the most straightforward and efficient way to increase physical activity levels, but other means, such as increasing active transportation (Mueller et al., 2015) or decreasing sedentary time (van der Ploeg et al., 2012) are also promising intervention targets. Importantly, these behaviors are probably not independent of each other. For instance, Ridgers, Timperio, Cerin and Salmon (2014) showed that higher physical activity levels on one day were compensated with lower physical activity levels on the following day in 8- to 11-year-old children, supporting our hypothesis of a homeostatic need to be active that is rather fixed at a biological set point. The large problem when it comes to twin studies on general physical activity is that very large datasets are needed for genetic analyses, whereas objective assessment of these behaviors on a large scale is very expensive. This leaves subjective report of general physical activity using questionnaires, which may not be very reliable (Adamo et al., 2009; Prince et al., 2008). Other than salient (high intensity) voluntary activities, people are simply not very good at estimating their daily physical activity level.

A number of smaller-scaled twin studies and one larger study have estimated the heritability of physical activity and sedentary time based on accelerometer-derived twin data. Table 2 depicts studies that were (partly) conducted under free-living conditions in same-sex twin pairs aged older than 5 years and the heritability estimates of the accelerometer-derived measures outside the laboratory. Just as for exercise behavior, shared environmental factors seem to play a major role in childhood and genetic components affect physical activity

both in younger and older adults. We have recently collected accelerometer data in MZ twins aged 16-26 years and found a twin correlation of 0.58 for moderate-to-vigorous physical activity and a twin correlation of 0.50 for sedentary behavior (N= 38 pairs). These might be underestimations compared to the general MZ twin population as these twins were selected based on their discordance on exercise behavior. Still, the correlations show that there is clearly a familial factor involved in these behaviors. DZ twin data are now needed to decompose this familial factor into genetic effects and shared environmental effects and a large-scale study investigating the heritability of physical activity and sedentary behavior based on accelerometer data is underway in the NTR.

MAIN IMPLICATIONS AND FINAL CONCLUSION

I hope that this thesis has conveyed the enormous complexity of the factors influencing individual differences in exercise behavior, especially in childhood and youth, where we observed profound changes in the genetic architecture over time. Given these intricacies, it makes a lot of sense that applying “one-size-fits-all” interventions to largely differing subgroups of the population will not bring about satisfactory changes in behavior. Although we have roughly revealed the underlying causes of variation in exercise behavior between the ages of 7 and 18 years by decomposing total variance into genetic, shared environmental and non-shared environmental effects, the specific factors that underlie those variance components have not been identified, which is the most important challenge for future research in order to develop successful “personalized” interventions. The genetic component in regular exercise behavior tells us that it will be harder to engage some people in physical exercise than others - but “harder” does not mean “undoable”. On the contrary, accepting innate human variation can help us increase the net yield of intervention efforts.

Based on the findings of this thesis, I have suggested that we are best served by family-based interventions for children and individual-based interventions for adolescents. Family-based interventions for children 1) should explicitly include parents, 2) should be tailored to different needs of boys versus girls and 3) they may want to primarily target at risk families (e.g., low parental education). Individual-based interventions for adolescents should aim to target genetically caused differences in exercise behavior. If it would appear, as we speculated, that exercise ability strongly impacts on adolescent exercise behavior, sports clubs entrusted with our 12- to 18-year-olds might want to emphasize

TABLE 2 Overview of studies on the heritability of physical activity and sedentary time as assessed by accelerometry in free-living conditions (twins aged >5 years).

Reference	Sample	Phenotype	ACDE (percentages)
Joosen, Gielen, Vlietinck & Westerterp (2005)	N=18 twin pairs (12 MZ, 6 DZ); 2 sibling pairs; Aged 18-39 years	Total PA	A=78, E=22
Fisher et al. (2010)	234 individuals (57 MZ twin pairs, 60 DZ); Aged 9-12 years	(1) Total PA (2) MVPA (3) Sedentary time	(1) C=73, E=27 (2) C=61, E=39 (3) C=55, E=45
den Hoed et al. (2013)	1654 twins (420 MZ pairs, 352 DZ); Mean age 56.3 years; Mostly women	(1) Acceleration of the trunk (2) PA energy expenditure (3) MVPA (4) Sedentary time As derived from a combined heart rate and movement sensor.	(1) A=35, C=2, E=63 (2) A=47, C=2, E=52 (3) A=47, E=53 (4) A=31, C=15, E=55
Gielen et al. (2014)	51 twin pairs, 1 male triplet (29 MZ, 23 DZ); Mean age 22 years	(1) Total PA (2) Low-intensity PA (3) Moderate-intensity PA (4) high-intensity PA	(1) A=57, E=43 (2) A=38, E=62 (males); A=72, E=28 (females) (3) ACE, AE, CE and ADE equally parsimonious; AE: A=55, E=45 (4) AE: A=47, E=53 or ADE: D=55, E=45

PA=physical activity, MVPA=moderate-to-vigorous physical activity, MZ=monozygotic, DZ=dizygotic, A=additive genetic effects, C=shared environmental effects, D=dominant genetic effects, E=non-shared environmental effects.

participation and pleasure rather than performance to increase exercise behavior in this group by 1) not selectively favoring resources (availability of trainers, coaches, fields and equipment) to the best teams/players, 2) de-emphasizing the competitive aspect and 3) offering a larger selection of exercise activities and intensities to suit different levels and forms of physical ability.

For understandable reasons, exercise interventions are often the domain of those who have a strong personal interest in and affiliation with exercise: there are very few exercise interventionists who are not themselves (ardent and proficient) exercisers. Notwithstanding the many benefits of such positive role models ("believers"), there are also dangers in strong personal beliefs and in propagated folk wisdom about exercise benefits. The studies on attitudes and BMI are an illustration of these dangers. Instead of repeating over and over again that exercise behavior will lead to weight loss or "feeling energetic", a more nuanced and realistic message would be wiser. Namely that for some, it will lead to weight loss and for others, it does not, and that for some, it will lead to feeling energetic and for others, it does not. It does not make any sense to convince individuals of the benefits of exercise behavior if these benefits simply do not match their genotypes. In fact, the initial motivation to exercise may be irreparably damaged when these promised effects do not occur. Up until today, exercise intervention programs have been mainly informed by the genome of regular exercisers. To increase their impact on the overall population, however, future intervention programs should also carefully heed the genome of those who are currently non-exercisers.