Regular exercise, subjective wellbeing, and internalizing problems in adolescence: causality or genetic pleiotropy?

Meike Bartels1,2*, Marleen H. M. de Moor1,2, Niels van der Aa1,2, Dorret I. Boomsma1,2 and Eco J. C. de Geus1,2

1 Department of Biological Psychology, VU University Amsterdam, Amsterdam, Netherlands
2 EMGO Institute for Health and Care, VU University Medical Centre, Amsterdam, Netherlands

INTRODUCTION

In the adult population regular exercise has consistently been associated with higher levels of wellbeing such as fewer depressive and anxious symptoms and higher levels of subjective wellbeing (SWB). Longitudinal studies showed that exercise participation at baseline predicted fewer internalizing problems at follow-up (Camacho et al., 1991; Wise et al., 2006; de Moor et al., 2008). In general, the conclusion that has been drawn from these studies is that increased levels of psychological wellbeing found in exercisers are the reflection of causal effects of exercise.

However an alternative hypothesis explaining the prospective associations between exercise behavior and psychological wellbeing may be that underlying factors influencing exercise behavior at baseline also influence psychological wellbeing at follow-up. Genetic variation among individuals may represent such an underlying factor. Under this hypothesis, prospective associations between exercise behavior and psychological wellbeing reflect the effects of common genetic factors influencing a disposition for exercise participation as well as a disposition for psychological wellbeing, a process referred to as genetic pleiotropy (de Geus and de Moor, 2008). When genetic effects on exercise behavior at baseline are associated with genetic effects on psychological wellbeing at a later time point, this leads to a longitudinal correlation between exercise behavior and psychological wellbeing that can mistakenly be taken for a causal effect of exercise.

Evidence for the hypothesis of genetic pleiotropy in exercise behavior and psychological wellbeing was supported in population-based adult twin studies (Stubbe et al., 2007; de Moor et al., 2008). de Moor et al. (2008) demonstrated cross-sectionally and longitudinally that genetic factors influencing a person’s tendency to engage in exercise also influence lower risk for depressive and anxious symptoms, whereas environmental factors affecting exercise behavior do not affect depressive and anxious symptoms. Strong support for a non-causal explanation came from the finding that in genetically identical twin pairs, the twin who exercised more did not have fewer internalizing problems and increased SWB. This was found cross-sectionally and longitudinally. We conclude that exercise behavior is associated with fewer internalizing problems and higher levels of SWB. The association largely reflects the effects of common genetic factors on these traits.

Keywords: regular exercise, subjective wellbeing, internalizing behavior, causality, genetic pleiotropy, twin-sibling
Most research up to date has focused on adults. It is important to further study the association between exercise behavior and psychological wellbeing during adolescence because several studies reported that the prevalence of exercise participation declines during adolescence (van Mechelen et al., 2000; Stubbe et al., 2005; van der Aa et al., 2010a), while the prevalence of internalizing problems shows its most prominent increase (Costello et al., 2005). Furthermore, the genetic architecture of exercise behavior was found to be vastly different across the life span with a peak in heritability during late adolescence (Beunen and Thomis, 1999; Maia et al., 2002; Stubbe et al., 2005; van der Aa et al., 2010a). van der Aa et al. (2010a), for instance, showed that variation in exercise behavior during adolescence was largely accounted for by genetic factors (72–85% in boys and 38–80% in girls), whereas shared environmental factors only accounted for part of the variation in young adolescent girls (46%). Among adolescents, individual differences in internalizing and SWB have also been found to be substantially influenced by genetic factors, whereas shared environmental factors accounted for variation in internalizing only during young adolescence and were absent with regard to SWB (Bartels and Boomsma, 2009; Lamb et al., 2010; Bartels et al., 2011).

In the present study we tested whether exercise behavior causally influences internalizing problems and SWB in a large population-based sample of adolescent twins and their non-twin-siblings. Cross-sectional and longitudinal data were used, providing a strong non-experimental design to test the causal hypothesis. Applying the procedure as described by de Moor et al. (2008), four predictions generated by the causal hypothesis were tested. A graphical representation of the four models used to test causal effects of exercise behavior on internalizing problems and SWB is presented in Figure 1.

A first prediction is that, if exercise causally influences internalizing problems and SWB, genetic and environmental factors influencing individual differences in exercise behavior will also, through the causal chain, influence individual differences in internalizing problems and SWB. This can be tested in a bivariate genetic model by computing genetic and environmental correlations between two traits. If genetic and environmental correlations are both significant this would be in line with the causal hypothesis (see Figure 1A). If only the genetic correlation is significant this is more compatible with the hypothesis of common genetic factors underlying exercise behavior, internalizing problems, and SWB.

A second prediction of the causal hypothesis is that monozygotic within-twin pair differences in exercise behavior should be associated with within-twin pair differences in internalizing problems and SWB. The twin who exercises more should display fewer internalizing problems and higher levels of SWB than the co-twin who exercises less (see Figure 1B). This would be in line with the causal hypothesis and it excludes confounding by genetic factors (the twins are genetically identical), whereas if the association is not significant this would falsify the causal effect of exercise.

A third prediction made by the causal hypothesis is that, if exercise causes fewer internalizing problems and higher levels of SWB, there should be a significant longitudinal association between exercise at baseline and internalizing problems and SWB at follow-up. This should be independent of common genetic factors influencing exercise behavior, internalizing problems, and SWB. This can be tested in a bivariate genetic model by computing genetic and environmental correlations between exercise behavior at baseline and internalizing problems and SWB at a successive time point. If genetic and environmental correlations between exercise behavior at baseline and internalizing problems or SWB at follow-up are both significant this would be in line with the causal hypothesis (see Figure 1C). If only the genetic correlation is significant this is more compatible with the hypothesis of common genetic factors underlying exercise behavior, internalizing problems and SWB.

To extend the procedure by de Moor et al. (2008), a fourth prediction was tested. Under the causal hypothesis, monozygotic intrapair differences in within-individual changes in exercise behavior over time should be associated with parallel intrapair differences in within-individual changes in internalizing problems and SWB over time. In other words, the twin showing the largest increase in frequency and intensity of exercise behavior over time should display larger decreases in internalizing problems and increases in SWB over time than the co-twin showing a lower increase or even a decrease in frequency and intensity of exercise behavior over time (see Figure 1D). This would be compatible with the causal hypothesis and it excludes confounding by genetic factors that would independently cause increases in exercise behavior as well as wellbeing over time.

**MATERIALS AND METHODS**

**SUBJECTS**

Participants were registered with the Netherlands Twin Registry (NTR), established by the Department of Biological Psychology at the VU University in Amsterdam (Boomsma et al., 2006; Bartels et al., 2007). The large majority of twins had been registered with the NTR as newborns and they were followed throughout childhood. At the time twins were 13 years old, their parents were asked for consent to send their children a survey. If their parents consented, twins and their non-twin-siblings received an online or a paper and pencil self-report survey when they were 14, 16, and 18 years. The survey contained items about behavior, exercise, lifestyle, and wellbeing. In this way, longitudinal data on exercise behavior, internalizing problems, and SWB were collected on three different occasions during adolescence. Data collection started in 2005 and currently data have been collected on two occasions (14 and 16 or 16 and 18). Since our data collection is an ongoing process, data on three different occasions during adolescence will become available in the coming years. When twins and siblings did not return the survey on time they were contacted by mail for a first reminder and next they were contacted by phone for a second reminder. For a detailed description of the adolescent survey collection in the NTR see Bartels et al. (2011). A total of 3645 families with twins born between 1986 and 1994 participated in this ongoing study at least once so far. The overall family response rate is 56%. Because the response rate of the survey collection was moderate, non-response analyses were conducted in which responding and non-responding families were compared on a number of different characteristics, such as socioeconomic status (SES), parental educational level, emotional and behavioral problems of the twins during childhood, and educational achievement.
FIGURE 1 | Graphic representation of four models used to test the hypothesis that exercise causes lower levels of internalizing problems. (A) A causal effect of exercise predicts a correlation between genetic, shared environmental, and non-shared environmental factors influencing exercise behavior and internalizing problems. Cross-sectional correlation between exercise and internalizing problems = \( g_{\text{exc}} \times r_{g} \times g_{\text{int}} \) + \( c_{\text{exc}} \times r_{c} \times c_{\text{int}} \) + \( e_{\text{exc}} \times r_{e} \times e_{\text{int}} \). (B) A causal effect of exercise predicts that in genetically identical twins, the twin who exercises more has lower levels of internalizing problems. (C) A causal effect of exercise predicts a correlation between genetic, shared environmental, and non-shared environmental factors influencing exercise behavior at time 1 and internalizing problems at time 2. Longitudinal correlation between exercise at time 1 and internalizing problems at time 2 = \( g_{\text{exc}} \times r_{g1,2} \times g_{\text{int}} \) + \( c_{\text{exc}} \times r_{c1,2} \times c_{\text{int}} \) + \( e_{\text{exc}} \times r_{e1,2} \times e_{\text{int}} \). (D) A causal effect of exercise predicts that in genetically identical twins, the twin who increases exercise intensity over time also shows decreased levels of internalizing problems over time.

of the twins. Observed differences on these characteristics were relatively small, i.e., responding and non-responding families were fairly comparable. In other words, response bias is not likely to constitute a serious threat to the validity of the results in the present study.

 Twins and siblings with an illness or handicap interfering with their daily lives were not selected for the present study. Information on this came from parental reports from questionnaires in previous years, and from the parental consent form on which parents could indicate if a twin or a sibling was not able to participate in questionnaire research. Data from one additional sibling per family were included. From families with more than one additional participating sibling, we selected the sibling closest in age to the twin. This resulted in exclusion of 157 siblings from the analyses. Maximum age difference between twins and siblings was 5.7 years (Mean age difference = 2.4 years, SD = 1.0). The total sample consisted of 6317 twins (44% male) and 1180 non-twin-siblings (44% male), from 3486 families, for whom data were available on at least one time point. For the cross-sectional analyses, data from the measurement occasion at the youngest age were selected. The age of the twins in the cross-sectional dataset ranged between 13.2 and 19.9 years, with a mean age of 15.7 years (SD = 1.4) and the age of the non-twin-siblings ranged between 13.0 and 20.0 with a mean age of 16.9 years (SD = 1.9). The cross-sectional dataset consisted of 528 MZ male, 455 DZ male, 778 MZ female, 559 DZ female, 1085 DZ opposite-sex twin pairs, 519 brothers, and 661 sisters. Of these, there were 2912 (86%) complete and 493 (14%) incomplete twin pairs. There were 81
families for which only non-twin-sibling data were available. For 802 (34.6%) of the same-sex twin pairs zygosity was determined based on blood group or DNA typing. Zygosity for the remaining same-sex twin pairs was determined by longitudinally collected questionnaire items about physical similarities and confusion by family members and strangers. These items allow accurate determination of zygosity in 93% of same-sex twin pairs (Rietveld et al., 2000).

Longitudinal analyses were based on a subsample of 1658 twins (38% male) and 150 non-twin-siblings (39% male), from 979 families for whom data on exercise behavior, internalizing problems, and SWB were available at two time points (at age 14 and 16 or 16 and 18). The longitudinal dataset consisted of 135 MZ male, 91 DZ male, 259 MZ female, 179 DZ female, and 289 DZ opposite-sex twin pairs, 58 brothers, and 92 sisters. Of these, there were 705 (74%) complete and 248 (26%) incomplete twin pairs. There were 26 families for which longitudinal data were available for non-twin-siblings only.

MEASURES
Participants were asked to indicate what type(s) of regular leisure exercise they were involved in at the time of assessment. A list of 21 common individual (including for example fitness, jogging) and team-based exercise activities (soccer, field hockey) was provided plus five open entries for less common activities. For each exercise activity endorsed, the participants further reported on the number of months per year, weekly frequency, and average duration of the activity. Ainsworth’s Compendium of physical activity (Ainsworth et al., 2000) was used to assign a MET score (Metabolic equivalent) to each exercise activity, reflecting its energy expenditure as a multiple of the basal energy expenditure (approximately 1 kcal/kg/h) in an average subject engaged in that activity.

For each participant a total weekly MET score was computed across all exercise activities by summing the products of the number of hours spent weekly on each exercise activity and its MET score. Activities were only scored if that participant had engaged in them for at least 3 months during the past year. Exercise during physical education classes at school was not included in the weekly MET score, since this is compulsory at all schools in the Netherlands. Thus the dependent variable reflects leisure time exercise behavior only. The distribution of the total weekly MET score was skewed (distribution, 2.58) and demonstrated kurtosis (distribution, 16.06). Therefore, for the bivariate genetic models, log-transformation was applied to the weekly MET score which significantly reduced skewness and kurtosis to −0.51 and −1.12, respectively. For computation of the difference scores, we used untransformed weekly MET scores, because the difference scores were normally distributed.

Self-reported internalizing problems were collected with the Youth Self-Report (YSR; Achenbach and Rescorla, 2001). The YSR was developed for adolescents to rate behavioral and emotional problems. The questionnaire consisted of 112 items from the 2001 version supplemented with six items from older versions of the YSR. The items had to be answered on a three-point scale based on the occurrence of the behavior during the past 6 months: 0 if the item was not true, 1 if the item was somewhat or sometimes true, and 2 if the item was often or very true. We used a Dutch translation of the YSR which shows good psychometrical properties (Verhulst et al., 1997). The internalizing scale consists of the subscales Anxious/Depressed, Somatic Complaints, and Withdrawn and has 31 items. Example items are “I am too shy” and “I feel lonely.” Scores on the individual items were summed to get an overall score of internalizing problems which could range from 0 to 62 with high scores indicating high levels of internalizing problems. Internal consistency of the scale was good with a Chronbach’s Alpha of 0.88.

Three measures of self-reported evaluations of SWB were used in the present study: (1) satisfaction with life (SAT) was assessed with the Satisfaction with Life Scale (Diener et al., 1985). The scale consists of five items which had to be answered on a seven-point scale ranging from 1 = “strongly disagree” to 7 = “strongly agree.” Example items are “My life is going more or less as I wished” and “I am satisfied with my life.” Internal consistency of the scale was good with a Chronbach’s Alpha of 0.86. (2) Subjective happiness (HAP) was assessed with the Subjective Happiness Scale (Lyubomirsky and Lepper, 1999). The scale consists of four items which had to be answered on seven-point scale ranging from 1 = “strongly disagree” to 7 = “strong agree.” Example items are “On the whole I am a happy person” and “On the whole, I am not very happy.” Internal consistency of the scale was good with a Chronbach’s Alpha of 0.84. (3) Quality of life in general (QLg) was assessed with the Cantril Ladder (Cantril, 1965). The ladder has 10 steps: the top indicated the best possible life, and the bottom the worst possible life. Participants had to indicate the step of the ladder at which they place their lives in general.

A genetic factor analysis of SAT, HAP, and QLg demonstrated that these measures assess the same genetic construct (Bartels and Boomsma, 2009). Therefore measures of SAT, HAP, and QLg were summarized into a single score for SWB. For all participants, a latent factor score for SWB was estimated in Mplus (Muthén and Muthén, 2010). Phenotypic correlations between SAT, HAP, and QLg ranged between 0.54 and 0.74. Correlations between the resulting latent factor scores and SAT, HAP, and QLg ranged between 0.70 and 0.95, and the factor loadings were estimated at 0.89, 0.84, and 0.65 respectively, indicating that the latent factor score is a good representation of the three observed measures of SWB.

STATISTICAL ANALYSES
Causal effects of exercise on internalizing problems and SWB were tested in four different models. The first two tests were based on cross-sectional data. In the first test it was assessed whether the genetic (rc), shared environmental (rc), or non-shared environmental (rc) correlation between exercise behavior and internalizing problems and SWB were significantly different from zero. If the genetic (rc), shared environmental (rc), or non-shared environmental (rc) correlation between exercise behavior and internalizing problems, and between exercise behavior and SWB is not significantly different from zero, a causal effect is not likely to be the source of the association. It is important to note that the significance of rc can only be tested if individual differences in exercise behavior, internalizing problems, and SWB are accounted for by shared environmental factors. If, for example, variation in SWB is not accounted for by shared environmental
Exercise behavior and internalizing problems and SWB for a 2-year time interval were computed in Mx. If exercise participation causes fewer internalizing problems and increased levels of SWB, there should be a significant longitudinal correlation between exercise behavior at baseline and subsequent internalizing problems and SWB. This significant longitudinal correlation does not necessarily provide evidence for causality if common genetic factors drive the association between exercise behavior at baseline and subsequent internalizing problems and SWB. The genetic information in our sample was again used to more robustly test causality by fitting a longitudinal bivariate genetic model in Mx (see Figure 1C). If the effect of exercise behavior on internalizing problems and SWB is causal, \( r_{g1,2}, r_{c1,2}, \) and \( r_{e1,2} \) should be significant. \( r_{g1,2}, r_{c1,2}, \) and \( r_{e1,2} \) represent the genetic, shared environmental, and non-shared environmental correlation between the measurement of exercise behavior at baseline and successive measurement of internalizing problems and SWB with a 2-year interval.

As with the cross-sectional bivariate genetic model, the genetic architecture of the covariation between exercise behavior at baseline and internalizing problems and SWB at follow-up was specified, and \( r_{g1,2}, r_{c1,2}, \) and \( r_{e1,2} \) were estimated. Sex differences in the magnitude of \( r_{g1,2}, r_{c1,2}, \) and \( r_{e1,2} \) were assessed by testing whether constraining \( r_{g1,2}, r_{c1,2}, \) and \( r_{e1,2} \) for boys and girls to be equal resulted in a significant deterioration of model fit. The hypothesis of a causal effect was assessed by testing whether constraining \( r_{g1,2}, r_{c1,2}, \) and \( r_{e1,2} \) to zero would lead to a significant deterioration of model fit. It is important to note that the significance of \( r_{c1,2} \) can only be tested if individual differences in exercise behavior, internalizing problems, and SWB are accounted for by shared environmental factors. If, for example, variation in SWB is not accounted for by shared environmental factors there will be no \( r_{c1,2} \) between exercise behavior and SWB. In this case, the causal hypothesis will be supported if \( r_{g1,2} \) and \( r_{e1,2} \) are both significant, whereas a causal effect is not likely to be the source of the association if \( r_{g1} \) or \( r_{e1} \) is not significantly different from zero.

In the fourth test of the causal hypothesis we assessed whether MZ twin intrapair differences in individuals’ changes in exercise behavior over time are associated with intrapair differences in individuals’ changes in internalizing problems and SWB over time (see Figure 1D). Significant associations between these intrapair differences would be compatible with the causal hypothesis, whereas non-significant associations would falsify this hypothesis. Within-individual differences between baseline and follow-up scores were computed for the measures of exercise behavior, internalizing problems, and SWB. Subsequently, differences between an MZ twin and his or her co-twin were computed for the within-individual changes in exercise behavior, internalizing problems, and SWB over time. The resulting MZ intrapair difference scores were normally distributed. Using maximum-likelihood estimation in Mplus, the MZ intrapair differences in within-individual changes in internalizing problems and SWB over time were regressed on the MZ intrapair differences in within-individual changes in exercise behavior over time.

**RESULTS**

In Table 1, means and variances of exercise behavior, internalizing problems, and SWB (uncorrected for age) are presented as
a function of sex. Significant effects of age were found on the means of exercise behavior \( [\chi^2 (1) = 43.28, p \leq 0.01]\); standardized regression coefficient \( \beta = -0.08\), internalizing problems \( [\chi^2 (1) = 7.83, p \leq 0.01]\); standardized regression coefficient \( \beta = 0.03\), and SWB \( [\chi^2 (1) = 26.74, p \leq 0.01]\); standardized regression coefficient \( \beta = -0.06\), indicating that mean scores for exercise behavior and SWB were higher for younger participants, whereas the mean score for internalizing problems was higher for older participants. Means of exercise behavior \( [\chi^2 (1) = 149.75, p \leq 0.01]\) and SWB \( [\chi^2 (1) = 355.84, p \leq 0.01]\) were significantly higher for boys, whereas the mean of internalizing problems \( [\chi^2 (1) = 18.69, p \leq 0.01]\) was significantly higher for girls. Variance in exercise behavior \( [\chi^2 (1) = 4.13, p = 0.04]\) was significantly larger for boys, whereas variance in internalizing problems \( [\chi^2 (1) = 251.31, p \leq 0.01]\) and SWB \( [\chi^2 (1) = 23.78, p \leq 0.01]\) were significantly higher for girls.

Twin and twin-sibling correlations for exercise behavior, internalizing problems and SWB are presented in the left part of Table 2. The DZ twin correlations for the traits were equal to twin-sibling correlations (all \( p \)-values > 0.05). The MZ twin correlations were significantly larger than the DZ twin and twin-sibling correlations (all \( p \)-values < 0.01), indicating that genetic influences accounted for individual differences in exercise behavior, internalizing problems and SWB. For girls, DZ twin, and twin-sibling correlations for exercise behavior and internalizing problems were higher than half the MZ twin correlations, suggesting that shared environmental factors play a role in explaining variation in exercise behavior and internalizing problems. For exercise behavior and internalizing problems in boys and SWB in boys and girls, DZ twin, and twin-sibling correlations were about half the MZ twin correlations, suggesting no shared environmental effects.

### CROSS-SECTIONAL ASSOCIATIONS

The phenotypic correlation between exercise behavior and internalizing problems was \(-0.13 [95\% \text{ CI: } -0.14 \text{ to } -0.10]\); \( \chi^2 (1) = 93.81, p \leq 0.01 \) and between exercise behavior and SWB it was \(0.12 [95\% \text{ CI: } 0.09 \text{ to } -0.15]; \chi^2 (1) = 91.90, p \leq 0.01 \). The phenotypic correlations were equal for boys and girls (all \( p \)-values > 0.05). This indicates that increased levels of exercise behavior are cross-sectionally associated with decreased levels of internalizing problems and increased levels of SWB.

The phenotypic correlation between internalizing problems and SWB was \(-0.51 [95\% \text{ CI: } -0.53 \text{ to } -0.49]; \chi^2 (1) = 1766.78, p \leq 0.01 \). Cross-twin cross-trait correlations are presented in the right part of Table 2. The cross-twin cross-trait correlations in MZ and DZ twin pairs, and twin-sibling pairs were significant (all \( p \)-values < 0.05). This indicates that exercise behavior in one twin is predictive of internalizing problems and SWB in his or her co-twin or non-twin-sibling. The DZ cross-twin cross-trait correlations were equal to twin-sibling cross-trait correlations (\( p \)-values exceeded a 5% significance level). The MZ cross-twin cross-trait correlations were significantly larger than the DZ cross-twin cross-trait and the twin-sibling cross-trait correlations (all \( p \)-values < 0.05), indicating that genetic influences partly account for the covariation between exercise behavior and internalizing problems, and between exercise behavior and SWB. MZ and DZ cross-twin cross-trait correlations were equal for boys and girls (all \( p \)-values > 0.05).

### Bivariate genetic model

Bivariate genetic model fitting results are presented in Table 3. In the upper part of the table the results of the bivariate genetic model for exercise behavior and internalizing problems are presented. \( r_c \) was not modeled for boys, because shared environmental influences did not contribute to variation in exercise behavior and internalizing problems in boys (Lamb et al., 2010; van der Aa et al., 2010a; Bartels et al., 2011). In model 2 the genetic \( (r_g) \) and non-shared environmental correlation \( (r_c) \) were constrained to be equal for boys and girls which did not result in a significant deterioration of model. Subsequent models were therefore compared to this model. Models 3–5 tested whether constraining \( r_g \) and \( r_c \) for girls, and \( r_c \) zero would lead to a significant deterioration of model fit. \( r_g \) was statistically significant for boys and, and \( r_c \) was statistically significant for girls, whereas \( r_c \) was not statistically significant. \( r_g \) was \(-0.19 [95\% \text{ CI: } -0.26 \text{ to } -0.15] \) and \( r_c \) for girls

### Table 1

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Var</th>
<th>SWB</th>
<th>Exercise–internalizing</th>
<th>Exercise–SWB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exercise behavior (weekly MET score)</td>
<td>Boys 26.8</td>
<td>664.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internalizing problems (scale score)</td>
<td>Boys 6.8</td>
<td>31.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWB (factor score)</td>
<td>Boys 0.06</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Exercise</th>
<th>Internalizing</th>
<th>SWB</th>
<th>Exercise–internalizing</th>
<th>Exercise–SWB</th>
</tr>
</thead>
<tbody>
<tr>
<td>MZM</td>
<td>0.69 (0.65–0.73)</td>
<td>0.63 (0.56–0.68)</td>
<td>0.41 (0.33–0.48)</td>
<td>(-0.11 \text{ to } -0.06)</td>
</tr>
<tr>
<td>DZM/sibling</td>
<td>0.32 (0.26–0.38)</td>
<td>0.34 (0.25–0.43)</td>
<td>0.11 (0.03–0.19)</td>
<td>(-0.07 \text{ to } -0.01)</td>
</tr>
<tr>
<td>MZF</td>
<td>0.72 (0.69–0.75)</td>
<td>0.50 (0.45–0.55)</td>
<td>0.47 (0.41–0.52)</td>
<td>(-0.12 \text{ to } -0.09)</td>
</tr>
<tr>
<td>MZF/sibling</td>
<td>0.42 (0.36–0.47)</td>
<td>0.30 (0.25–0.35)</td>
<td>0.24 (0.18–0.30)</td>
<td>(-0.08 \text{ to } -0.04)</td>
</tr>
<tr>
<td>DOS/sibling</td>
<td>0.28 (0.24–0.33)</td>
<td>0.24 (0.19–0.28)</td>
<td>0.20 (0.15–0.24)</td>
<td>(-0.05 \text{ to } -0.02)</td>
</tr>
</tbody>
</table>

MZ, monozygotic twin pair; DZ, dizygotic twin pair; M, male; F, female; DOS, opposite-sex twin pair.
was $-0.45$ (95% CI: $-1.0$ to $-0.19$). These results suggest that the association between exercise behavior and internalizing problems in boys seems to be mainly due to genetic factors which are common (i.e., overlapping) for exercise behavior and internalizing problems. For girls, genetic and shared environmental factors both account for the association. Because $r_e$ was not different from zero the causal hypothesis was not supported in either boys or girls.

In the lower part of Table 3 the results of the bivariate genetic model for exercise behavior and SWB are presented. For SWB, $r_e$ was not modeled, because shared environmental influences did not contribute to variation in SWB in boys and girls (Bartels and Boomsma, 2009; van der Aa et al., 2010b). In model 2, $r_g$ and $r_e$ were constrained to be equal for boys and girls which resulted in a significant deterioration of model fit. Therefore, $r_g$ and $r_e$ were allowed to differ between boys and girls. Model 3 and 4 tested whether constraining $r_g$ and $r_e$ to zero would lead to a significant deterioration of model fit. $r_g$ was statistically significant in boys and girls and estimated at 0.22 (95% CI: 0.14 to 0.29) and 0.31 (95% CI: 0.23 to 0.39) for boys and girls respectively. For boys, $r_e$ was not significantly different from zero, $p \leq 0.01$ and estimated at 0.02 (95% CI: $-0.06$ to 0.10). For girls, $r_e$ was significant and close to zero (0.05; 95% CI: 0.01 to $-0.09$). These results suggested that the association between exercise behavior and SWB in girls is mainly due to genetic pleiotropy.

### MZ twin intrapair differences model

In the upper part of Figure 2, cross-sectional associations between MZ within-twin pair difference scores in exercise behavior and internalizing problems (Figure 2A) and SWB (Figure 2B) are presented. It can be seen that there were no significant associations between within-twin pair differences in exercise behavior and internalizing problems and SWB. This indicates that in genetically identical twin pairs, a twin who exercises more does not have fewer internalizing problems or higher levels of SWB than his or her co-twin who exercises less. This is not in line with the causal hypothesis that increased levels of exercise cause decreased levels of internalizing problems and increased levels of SWB.

### Longitudinal associations

Based on the longitudinal subsample, twin and twin-sibling correlations for exercise behavior, internalizing problems and SWB at the first (T1) and second (T2) measurement occasion are presented in the left part of Table 4. The DZ twin correlations for the traits at both measurement occasions were equal to twin-sibling correlations ($p$-values $> 0.05$). Twin and twin-sibling correlations on T1 and T2 were largely similar to those presented in Table 2 which were based on the cross-sectional sample.

The longitudinal phenotypic correlation between exercise behavior and internalizing problems 2 years later was $-0.15$ [95% CI: $-0.21$ to $-0.08$] and between exercise behavior and SWB it was 0.16 [95% CI: 0.11 to 0.21]. These correlations were significant larger than the DZ cross-twin cross-trait and the twin-sibling cross-trait correlations which were based on the cross-sectional sample.

### Table 3 | Bivariate model fitting results for exercise behavior and internalizing problems and SWB.

<table>
<thead>
<tr>
<th>Model</th>
<th>$r_e$, sex differences</th>
<th>$-2LL$</th>
<th>df</th>
<th>$\chi^2$</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. $r_g$, $r_e$ boys; $r_g$, $r_c$, $r_e$ girls</td>
<td>–</td>
<td>58490.629</td>
<td>14427</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2. $r_g$, $r_e$ boys; $r_g$, $r_c$, $r_e$ girls</td>
<td>1</td>
<td>58492.491</td>
<td>14429</td>
<td>1.86</td>
<td>2</td>
<td>0.39</td>
</tr>
<tr>
<td>3. $r_g$, $r_e$ girls</td>
<td>2</td>
<td>58514.756</td>
<td>14430</td>
<td>22.26</td>
<td>1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>4. $r_c$ girls</td>
<td>2</td>
<td>58503.387</td>
<td>14430</td>
<td>10.90</td>
<td>1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>5. $r_e$ girls</td>
<td>2</td>
<td>58493.414</td>
<td>14430</td>
<td>0.92</td>
<td>1</td>
<td>0.34</td>
</tr>
</tbody>
</table>

$r_e$: genetic correlation; $r_c$: shared environmental correlation; $r_c$: non-shared environmental correlation; vs., versus; $-2LL$: −2 log-likelihood; df: degrees of freedom; $\chi^2$: chi-square test statistic; $p$, p-value.
FIGURE 2 | Cross-sectional and longitudinal associations between MZ within-twin pair differences in exercise behavior and MZ within-twin pair differences in internalizing problems and SWB. (A) Corresponding standardized regression coefficient $\beta = -0.03 \left[ \chi^2 (1) = 1.21, p = 0.27 \right]$ for cross-sectional association between MZ within-twin pair difference in exercise behavior and internalizing problems. (B) $\beta = -0.01 \left[ \chi^2 (1) = 0.03, p = 0.85 \right]$ for cross-sectional association between MZ within-twin pair difference in exercise behavior and SWB. (C) $\beta = 0.07 \left[ \chi^2 (1) = 0.915, p = 0.34 \right]$ for longitudinal association between MZ within-twin pair differences in exercise behavior and internalizing problems. (D) $\beta = 0.12 \left[ \chi^2 (1) = 3.70, p = 0.06 \right]$ for longitudinal association between MZ within-twin pair differences in exercise behavior and SWB.

Longitudinal bivariate genetic model

Longitudinal bivariate model fitting results are presented in Table 5. In the upper part of the table results of the longitudinal genetic model for exercise behavior and internalizing problems are presented. Model 2 tested whether constraining the genetic ($r_{g1,2}$) and non-shared environmental correlation ($r_{e1,2}$) to be equal for boys and girls led to a significant deterioration of model fit. $r_{g1,2}$ and $r_{e1,2}$ were not significantly different for boys and girls and therefore constrained to be equal. Models 3–5 tested whether constraining $r_{g1,2}$ and $r_{e1,2}$ for boys and girls, and $r_{c1,2}$ for girls, to zero would lead to a significant deterioration of model fit. $r_{g1,2}$ and $e_{1,2}$ were not significantly different from zero. $r_{c1,2}$ for girls and $r_{e1,2}$ were not different from zero. This suggests that the longitudinal association between exercise behavior and internalizing problems in both sexes is mainly accounted for by genetic factors which are common for exercise behavior and internalizing problems. Because $r_{c}$ for girls were not different from zero the causal hypothesis was not supported.

In the lower part of Table 5 results of the longitudinal genetic model for exercise behavior and SWB are presented. Model 2 tests whether constraining $r_{g1,2}$ and $r_{e1,2}$ to be equal for boys and girls resulted in a significant deterioration of model fit. $r_{g1,2}$ and $r_{e1,2}$ were not significantly different for boys and girls. Models 3 and 4 tested whether constraining $r_{g1,2}$ and $r_{e1,2}$ to zero would lead to a significant deterioration of model fit. $r_{g1,2}$ was statistically significant, whereas $r_{e1,2}$ was not significantly larger than zero. $r_{g1,2}$ was $0.31 \ (95\% \ CI: 0.19 \ to \ -0.43)$. This suggests that the longitudinal association between exercise behavior and SWB is also best explained by genetic factors which are common (i.e., overlapping) for exercise behavior and SWB. Because $r_{e}$ was not different from zero the causal hypothesis was not supported.

Longitudinal MZ twin intrapair differences model

In the lower part of Figure 2, associations between MZ within-twin pair difference scores in individuals’ changes in exercise behavior over time and MZ within-twin pair difference scores in individuals’ changes in internalizing problems (Figure 2C) and SWB
Table 4 | Twin and twin-sibling correlations for exercise behavior, internalizing problems and SWB at the first (T1) and second (T2) measurement occasion, and the longitudinal cross-twin cross-trait and cross-sibling cross-trait correlations for exercise behavior, internalizing problems and subjective wellbeing.

<table>
<thead>
<tr>
<th></th>
<th>Exercise</th>
<th>Internalizing</th>
<th>SWB</th>
<th>Exercise</th>
<th>Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T1</td>
<td>T2</td>
<td>T1</td>
<td>T2</td>
<td>T1–internalizing T2</td>
</tr>
<tr>
<td>MZM</td>
<td>0.74</td>
<td>0.66</td>
<td>0.71</td>
<td>0.40</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>(0.66–0.81)</td>
<td>(0.56–0.74)</td>
<td>(0.60–0.78)</td>
<td>(0.09–0.59)</td>
<td>(0.09–0.44)</td>
</tr>
<tr>
<td>DZM/sibling</td>
<td>0.27</td>
<td>0.21</td>
<td>0.27</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.12–0.40)</td>
<td>(0.02–0.38)</td>
<td>(0.08–0.43)</td>
<td>(−0.22 to 0.25)</td>
<td>(−0.15 to 0.22)</td>
</tr>
<tr>
<td>MZF</td>
<td>0.81</td>
<td>0.75</td>
<td>0.57</td>
<td>0.44</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>(0.76–0.84)</td>
<td>(0.69–0.79)</td>
<td>(0.48–0.65)</td>
<td>(0.32–0.55)</td>
<td>(0.41–0.60)</td>
</tr>
<tr>
<td>DZF/sibling</td>
<td>0.42</td>
<td>0.45</td>
<td>0.32</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.33–0.50)</td>
<td>(0.33–0.55)</td>
<td>(0.23–0.40)</td>
<td>(0.24–0.51)</td>
<td>(0.12–0.30)</td>
</tr>
<tr>
<td>DOS/sibling</td>
<td>0.36</td>
<td>0.24</td>
<td>0.24</td>
<td>0.20</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>(0.28–0.43)</td>
<td>(0.14–0.34)</td>
<td>(0.14–0.33)</td>
<td>(0.00–0.36)</td>
<td>(0.06–0.23)</td>
</tr>
</tbody>
</table>

T1, first measurement occasion; T2, second measurement occasion; MZ, monozygotic twin pair; DZ, dizygotic twin pair; M, male; F, female; DOS, opposite-sex twin pair.

Table 5 | Longitudinal bivariate model fitting results for exercise behavior and internalizing problems and SWB.

<table>
<thead>
<tr>
<th>Model</th>
<th>Exercise vs. Internalizing</th>
<th>−2LL</th>
<th>df</th>
<th>χ²</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EXERCISE BEHAVIOR–INTERNALIZING PROBLEMS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. r_{g1.2}, r_{e1.2} boys; r_{g1.2}, r_{e1.2} girls: sex differences</td>
<td>−</td>
<td>9583.709</td>
<td>2838</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2. r_{g1.2}, r_{e1.2} boys; r_{g1.2}, r_{e1.2} girls: no sex differences</td>
<td>1</td>
<td>9584.841</td>
<td>2840</td>
<td>1.13</td>
<td>2</td>
<td>0.57</td>
</tr>
<tr>
<td>3. r_{e1.2} = 0</td>
<td>2</td>
<td>9588.744</td>
<td>2841</td>
<td>3.90</td>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>4. r_{e1.2} = 0</td>
<td>2</td>
<td>9586.936</td>
<td>2841</td>
<td>2.10</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>5. r_{e1.2} = 0</td>
<td>2</td>
<td>9585.028</td>
<td>2841</td>
<td>0.19</td>
<td>1</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>EXERCISE BEHAVIOR–SWB</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. r_{g1.2}, r_{e1.2}: sex differences</td>
<td>−</td>
<td>7237608</td>
<td>3455</td>
<td>−</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>2. r_{g1.2}, r_{e1.2}: no sex differences</td>
<td>1</td>
<td>7239.709</td>
<td>3457</td>
<td>2.10</td>
<td>2</td>
<td>0.35</td>
</tr>
<tr>
<td>3. r_{e1.2} = 0</td>
<td>2</td>
<td>7265.019</td>
<td>3458</td>
<td>25.31</td>
<td>1</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>4. r_{e1.2} = 0</td>
<td>2</td>
<td>7239.716</td>
<td>3458</td>
<td>0.01</td>
<td>1</td>
<td>0.93</td>
</tr>
<tr>
<td>5. r_{g1.2} = r_{g1.2}exp</td>
<td>2</td>
<td>7245.787</td>
<td>3458</td>
<td>6.08</td>
<td>1</td>
<td>0.01</td>
</tr>
<tr>
<td>6. r_{e1.2} = r_{e1.2}exp</td>
<td>2</td>
<td>7248.324</td>
<td>3458</td>
<td>8.62</td>
<td>1</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

\( r_{g1.2}, \) genetic correlation; \( r_{e1.2}, \) shared environmental correlation; \( r_{e1.2}, \) non-shared environmental correlation; vs., versus; −2LL, −2 log-likelihood; df, degrees of freedom; \( \chi^2, \) chi-square test statistic; df, degrees of freedom of \( \chi^2 \)-test; p, p-value.

(Figure 2D) over time are presented. It can be seen that there were no significant associations between within-twin pair differences in exercise behavior and internalizing problems and SWB. This indicates that in genetically identical twin pairs, a twin showing increased frequency and intensity of exercise behavior over time does not show lower levels of internalizing problems or higher levels of SWB over time than his or her co-twin with decreasing levels of exercise behavior over time. This is again not in line with the causal hypothesis that increased levels of exercise cause decreased levels of internalizing problems and increased levels of SWB.

**DISCUSSION**

The present study investigated, in a large sample of Dutch adolescent twins and their non-twin-siblings, whether exercise behavior causally influences internalizing problems and SWB or whether the association between these traits reflects the effects of underlying genetic factors. In keeping with population-based studies investigating adult samples (Camacho et al., 1991; Norris et al., 1992; de Moor et al., 2006, 2008; Wise et al., 2006; Stubbe et al., 2007), we found that exercise behavior was cross-sectionally and longitudinally associated with fewer internalizing problems and increased levels of SWB across the total sample of adolescents. However, within genetically identical twins, a twin who exercised more did not have fewer internalizing problems or higher levels of SWB than his or her less exercising co-twin and, in addition, a twin showing increased frequency and intensity of exercise behavior over time did not show lower levels of internalizing problems or higher levels of SWB over time than his or her co-twin with...
unchanged or decreased levels of exercise behavior. The cross-sectional and longitudinal associations between exercise behavior and internalizing problems and SWB were entirely accounted for by genetic factors in boys and genetic and shared environmental factors in girls, whereas the unique environmental factors causing them to participate in exercise did not cause lower levels of internalizing problems and increased levels of SWB, currently or 2 years later.

Taken together, these results do not support the hypothesis that increased levels of psychological wellbeing found in adolescent exercisers reflect a causal effect of exercise. Instead, the results are compatible with the alternative hypothesis that the association between exercise behavior and psychological wellbeing in adolescence is mainly due to an underlying factor. Pleiotropic genetic effects seem to be the major component of this underlying factor although shared environmental influences also contribute to both exercise behavior and internalizing behavior in girls. In adults, pleiotropic genetic factors have also been found to explain the association between exercise and depressive and anxious symptoms (de Moor et al., 2008), as well as between exercise and SWB (Stuber et al., 2007).

The lack of support for the hypothesis that increased levels of psychological wellbeing found in exercisers reflect a causal effect of exercise needs to be reconciled with findings from training studies that reported improved mood or reductions in coping behavior, depression, and anxiety after a program of aerobic exercise training in comparison to control manipulations (Steptoe et al., 1989; Brosse et al., 2002; Barbour and Blumenthal, 2005). Although others have failed to replicate these training effects (King et al., 1989; de Geus et al., 1993) and critical reviews on this topic express only cautious optimism about the causal role of exercise (Lawlor and Hopker, 2001; Dunn et al., 2005), recent well-designed studies in depression patients report beneficial psychological effects of exercise that match or even exceed those of pharmacological treatment (Blumenthal et al., 2007). It should be recognized, however, that these studies examined the effects of prescribed and externally monitored exercise treatments in select subgroups and that methodological problems cling to even the most well-designed randomized controlled trial due to the strong societal beliefs about the efficacy of this intervention. Furthermore, there have been no randomized controlled training trials in adolescents that applied the same rigor as the trials in adults. Taken our evidence for genetic pleiotropy and the pitfalls in existing randomized controlled trials for adults, i.e., assuming that regular exercise may prevent depression in a non-clinical adolescent sample, because it has been successfully used as a therapy in subsets of clinically depressed adult patients, is still a big leap of faith.

In spite of the above, we do not rule out the possibility that exercise might have a causal effect on psychological wellbeing in specific subgroups of adolescents. In fact, a model has been proposed that is compatible with genetic pleiotropy and still allows exercise to causally increase wellbeing in specific subgroups. To maintain regular exercise participation, appetitive effects of exercise need to outweigh the aversive effects. Individuals for whom the aversive effects of exercise are stronger than the rewarding effects will eventually cease their exercise participation. In contrast, individuals for whom the appetitive effects are stronger than the aversive effects will maintain exercise participation and become regular exercisers. It has indeed been shown that regular exercisers report greater acute exercise-induced mood enhancement than non-exercisers (Gauvin, 1990; Hoffman and Hoffman, 2008). These positive mood experiences after exercise activities may be a positive reward and therefore an important component of the higher wellbeing reported by regular exercisers. Genetic variants tipping the balance between appetitive and aversive effects may be a major source of individual differences in the effects of exercise on these acute mood responses. These genetic effects on the acute mood response to exercise may not only keep adolescents maintaining regular exercise participation. The repeated exercise-induced increases in mood may influence their overall psychological wellbeing. It is important to note that the causal effect of exercise in this proposed model would be limited to a genetically sensitive subgroup and would not apply to the total adolescent population.

In addition, there is also a social–psychological mechanism through which the association between exercise behavior and psychological wellbeing may be explained by genetic pleiotropy (de Geus and de Moor, 2008). The basic idea is that people will seek out the activities that they are good in. This is particularly true during adolescence, because being good in sports is an important source of self-esteem for adolescents and the athletic role model is continuously reinforced by the media (Field et al., 1999; Pope et al., 2001). Therefore, genes coding for exercise ability (endurance, strength, flexibility, motor coordination) may well become genes for adolescent exercise behavior. These genetic factors may further influence psychological wellbeing because high exercise ability will coincide with the experience of strong feelings of competence and mastery as well as positive feedback from peers, leading to an increased sense of self-esteem and psychological wellbeing. In short, genetic variants influencing exercise ability will act to maintain exercise behavior and to increase psychological wellbeing, but again are limited to a selected group of adolescents, i.e., those with high exercise ability.

Apart from the gene-by-exercise interactive effects sketched above, we cannot rule out the presence of more complex mechanisms, like reverse or reciprocal causality, or combinations of causal, reverse causal, and pleiotropic mechanisms. These complex mechanisms should be assessed in future studies making use of more complex causation models and larger sample sizes. Nonetheless, our results already have implications for the prevention of internalizing problems and decreasing levels of SWB during adolescence and the maintenance of exercise behavior at a sufficient level throughout adolescence and during the transition to adulthood. Current intervention strategies are mainly based on the idea that the associations between exercise and psychological wellbeing reflect causal effects of exercise and that exercise interventions exert beneficial effects on all participants. While it remains possible that exercise behavior can prevent some adolescents from developing internalizing problems, our results show that those with a disposition toward internalizing problems also have a higher chance to be a non-exerciser because of their genetic make-up. Approaching these adolescents at risk for internalizing problems with demanding exercise programs may be doomed to...
fail and may even amplify feelings of low self-esteem and consequent internalizing problems. Instead, behavioral therapy aimed to improve psychological wellbeing may need to coincide with or even precede intervention on exercise behavior.

In addition, we cannot fully rule out the effects of genetic rater bias, i.e., the presence of a polygenic factor that predisposes to reporting higher amounts of regular exercise and reporting higher SWB. However, the findings of our longitudinal analyses show that within-individual change in exercise behavior (or reporting of it) over time did not lead to change in wellbeing (or reporting of it) which argues against a significant effect of rater bias.

It is important to note that DZ twins were as similar as non-twin-siblings with regard to exercise behavior, internalizing problems and SWB. In addition, cross-twin cross-trait correlations were also similar for DZ twins and non-twin-siblings. This implies that the present results do not only apply to twins, but that they can be generalized to the total adolescent population.

In conclusion, among Dutch adolescents, regular exercise was found to be associated with fewer internalizing problems and higher levels of SWB. The association could not be explained by causal effects of exercise behavior. Instead, the association was largely accounted for by genetic factors influencing a tendency toward exercise as well as a disposition toward low levels of internalizing problems and high levels of SWB.

ACKNOWLEDGMENTS

Funding was obtained from the following grants: “Spinozapremie” (NWO/SPIN 56-464-14192); “Twin-family database for behavior genetics and genomics studies” (NWO 480-04-004); “Genetic and Family influences on Adolescent Psychopathology and Wellness” (NWO 463-06-001); “A twin-sib study of adolescent wellness” (NWO-VENI 451-04-034). This study was supported by award number RO1DK092127 from the National Institute of Diabetes and Digestive and Kidney Diseases. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Diabetes and Digestive and Kidney Diseases or the National Institutes of Health. M. Bartels is financially supported by a senior fellowship of the EMGO institute for health and care. Marleen de Moor is financially supported by NWO-VENI 016-115-035.

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.